

D1.1 DATA MANAGEMENT PLAN

Revision: v.1.0

Work package	WP 1
Task	Task1.1
Due date	31/07/2023
Submission date	15/10/2023
Deliverable lead	University of Bologna
Version	1.0
Authors	Carla Amatetti, Alessandro Vanelli-Coralli (UNIBO)
Reviewers	Miguel Ángel Vázquez (CTTC), Maria Chiara Campodonico (D4P), Flavia Maragno (Martel)
Abstract	This deliverable presents the project's approach with respect to collection, treatment, and dissemination of data collected during the testing and piloting activities of the project.
Keywords	Data Management Plan, standards, metadata, policies

www.6g-ntn.eu



Grant Agreement No.: 101096479 Call: HORIZON-JU-SNS-2022 Topic: HORIZON-JU-SNS-2022-STREAM-B-01-03 Type of action: HORIZON-JU-RIA



Document Revision History

Version	Date	Description of change	List of contributor(s)
V0.1	01/08/2023	Creation of the document	UNIBO
V0.2	04/10/2023	Version following reviewers	UNIBO, CTTC, MARTEL
V1.0	15/10/2023	Submission version	UNIBO

DISCLAIMER





O Confederazione Svizzera Confederaziun svizra Swiss Confederation

Project funded by

 Schweizerische Eidgenossenschaft
 Federal Department of Economic Affairs,

 Confédération suisse
 Education and Research EAER
 State Secretariat for Education Research and Innovation SERI

6G-NTN (6G Non-Terrestrial Network) 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 101096479. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union. Neither the European Union nor the granting authority can be held responsible for them. This work has received funding from the Swiss State Secretariat for Education, Research and Innovation (SERI).

COPYRIGHT NOTICE

© 2023 - 2025 6G-NTN Consortium

Project co-funded by the European Commission in the Horizon Europe Programme		
Nature of the deliverable:	Nature of the deliverable: DMP	
Dissemination Level		
PU	Public, fully open, e.g., web (Deliverables flagged as public will be automatically published in CORDIS project's page)	*
SEN	Sensitive, limited under the conditions of the Grant Agreement	
Classified R-UE/ EU-R	EU RESTRICTED under the Commission Decision No2015/444	
Classified C-UE/ EU-C	EU CONFIDENTIAL under the Commission Decision No2015/ 444	
Classified S-UE/ EU-S	EU SECRET under the Commission Decision No2015/ 444	

* R: Document, report (excluding the periodic and final reports)

DEM: Demonstrator, pilot, prototype, plan designs

DEC: Websites, patents filing, press & media actions, videos, etc.

DATA: Data sets, microdata, etc.

DMP: Data management plan

ETHICS: Deliverables related to ethics issues.

SECURITY: Deliverables related to security issues

OTHER: Software, technical diagram, algorithms, models, etc.



EXECUTIVE SUMMARY

This deliverable provides the 6G-NTN Data Management Plan (DMP). Being in a very early stage of the project, this document presents preliminary proposals in terms of sharing, volume, and archiving.

The purpose of the DMP is to contribute to good data handling through indicating what research data the project expects to generate and describe which parts of the data that can be shared with the public. Moreover, it gives instructions on naming conventions, metadata structure, storing of the research data, and how to make public data available.

This plan describes:

- what kind of data will be generated, collected, processed, and shared;
- which standards will be applied during data collection and handling;
- the procedures for sharing and open access to the 6G-NTN data and for curation and preservation of the data;
- the procedures to comply with to the General Data Protection Regulation (GDPR) to ensure the protection of the involved companies' data, information and privacy rights.

The 6G-NTN consortium members consider the DMP as a checklist for the future and as a reference for the resource and budget allocations related to data management. The DMP is a living document that at any time will reflect the current state of the consortium's agreements regarding data management, exploitation, and protection of rights and results.





TABLE OF CONTENTS

1	INTRODUCTION	8
1.1	Purpose of the document	
1.2	Structure of the document	
2	DATA SUMMARY	9
2.1	6G-NTN's mission	9
2.2	dataset description	9
3	FAIR DATA	10
3.1	Types of data/research output	
3.2	Data standards and metadata	
3.3	Access, sharing, and preservation	
3.3.1	Non open research data	
3.3.2	Open research data	
4	ALLOCATION OF THE RESOURCES	16
5	DATA SECURITY	17
6	DATA ETHICAL ASPECTS	
7	CONCLUSION	19
APPE	NDIX	





LIST OF FIGURES

FIGURE 1: DATA MANAGEMENT LIFE CYCLE.	8
FIGURE 2: OPEN ACCESS FLOW	15





LIST OF TABLES

TABLE 1: DATASET DESCRIPTION TEMPLATE.	9
TABLE 2: 6G-NTN'S DATASET.	10





ABBREVIATIONS

6G	6 th Generation
AI	Artificial Intelligence
DMP	Data Management Plan
DOI	Digital Object Identifiers
EC	European Commission
EU	European Union
FAIR	Findable, Accessible, Interoperable, and Re-usable
GDPR	EU General Data Protection Regulation
NTN	Non – Terrestrial Network
RIC	Radio Intelligent Controller
RRM	Radio Resource Management
TN	Terrestrial Network
WP	Work Package





1 INTRODUCTION

1.1 PURPOSE OF THE DOCUMENT

The 6G-NTN Data Management Plan (DMP) is a document that provides details on all the research data collected and generated within the 6G-NTN project. In particular, it explains the way research data are handled, organized, licensed, and made available to the public, and how they will be preserved after the project is completed (as shown in Figure 1). The DMP also provides motivations when versions or parts of the project research data cannot be openly shared.

This DMP reflects the current state of the art of the 6G-NTN project. However, the details and the final number of datasets may vary during the research project. As such, this deliverable can be updated when necessary.



FIGURE 1: DATA MANAGEMENT LIFE CYCLE.

1.2 STRUCTURE OF THE DOCUMENT

The sections of the deliverables are organized in the following way:

- Section 2 summarizes information related to the data to be collected, the template to be used for describing the datasets and the objectives of the project that will be met through these data collection and processing.
- Section 3 describes the processes and the mechanisms that will be applied for making the data Findable, Accessible, Interoperable, and Re-usable (FAIR).
- Section 4 provides information on the financial aspects of making data FAIR.
- Section 5 deals with data security.
- Section 6 provides the summary of ethics and privacy strategy.
- Section 7 concludes the document.





2 DATA SUMMARY

2.1 6G-NTN'S MISSION

The broaden ambition of the 6G-NTN project is the design of a Non-Terrestrial Network (NTN) component fully integrated into future 6G infrastructure which allows to meet the vertical industry's needs and consumer market expectations thanks to its greater coverage, increased resilience, and improved sustainability.

In order to achieve this objective, the 6G-NTN project will research and develop a revolutionary three-dimensional (3D) network infrastructure concept for 6G with the ambition to provide, through unification with the Terrestrial Network (TN) component, a ubiquitous coverage with high data rate communication at quasi ultra-low latency while providing a disruptive concept of high accuracy and reliable location service. This infrastructure will leverage: i) more effective terminals in terms of cost, size, and power consumption, compatible with both TN and NTN access; ii) flexible software defined payloads; iii) flexible waveform using data-driven approaches; iv) Artificial Intelligence (AI)-enhanced Radio Intelligent Controller (RIC) to provide effective Radio Resource Management (RRM) solutions; and v) solutions relying on cloud native architecture and open interfaces mainly focused on securing the communication scheme. Eventually, 6G-NTN will define a roadmap for the development of the necessary technical, regulatory and standardization building blocks to enable integrated NTN service provisioning and disruptive market offer in the 2030-35 timeframe. The outcome of the project will contribute to foster Europe's technological leadership in the NTN sector and empower vertical domains well beyond the current 5G capabilities.

In this framework, in order for 6G-NTN to achieve its mission and to meet its objectives, access, collection, process, and management of the main data types is necessary. Additional data may be also used in case this will be deemed necessary, or as a result of the detailed architectural aspects or channel models that will be delivered at a later stage. These new datasets will be included in a next iteration.

2.2 DATASET DESCRIPTION

This subsection provides a preliminary template (as shown in Table 1) to be used for describing the datasets to be produced or collected in 6G-NTN project, which has been identified at this stage of the project. A preliminary list of the dataset produced during the 6G-NTN project is available in Table 2. For a sake of completeness, Table 2 lists also the mailing list, newsletter, and the website, which will be use to disseminate the results.

Datas numb	set ber project	Dataset title
		Name of the dataset defined
Statu	S	It can be:

TABLE 1: DATASET DESCRIPTION TEMPLATE.





	Not yet available if data collection is planned but has not yet started	
	 In progress, if data are currently being collected 	
	Available, if it is completed and ready to be deposited/has been deposited	
ID [ID type]	Type of ID	
Creator/s	Family name, given name [TEAM];	
Contributor/s	Family name, given name [TEAM];	
Contact Person/s	Family name, given name [TEAM, email]; …	
Contents	Description of the dataset explaining the data provenance, origin, and usefulness. Reference may be made to existing data that could be reused.	
Data format	All the format that defines data	
Data volume	State the expected size of the data	
Accessibility and type of license	 Explanation of the sharing policies related to the dataset between the next options: Open: Open for public disposal Embargo: It will become public when the embargo period applied by the publisher is over. In case it is categorized as embargo the end date of the embargo period must be written in DD/MM/YYYY format. Restricted: Only for project internal use. Each dataset must have its distribution license. Specify the license, e.g., commercial, not commercial. Provide information about personal data and mention if the data is anonymized or not. Tell if the dataset entails personal data and how this issue is taken into account 	
Archiving and preservation	The preservation guarantee and the data storage during and after the project.	
Re-used existing data	Y/N. If yes, state the re-use data and how/from where they were retrieved.	
Data utility	Outline to whom the dataset could be useful	

TABLE 2: 6G-NTN'S DATASET.

Dataset name	Short description
6G-NTN_WP6_T1_mailinglists_v1.0	 Internal mailing lists for project management and communication purposes. The data include: Name, Surname Affiliation Email address The data is collected in plain text.





	Amount: About 200 -300 entries overall, spread and/or duplicated within the different mailing lists (one per WP, general one, etc).
6G-NTN_WP6_T1_newsletter _v1.0	 The 6G-NTN newsletter collects the following data: E-mail address of visitors registering to the newsletter. During the registration for the newsletter, the IP address of the computer system assigned by the Internet Service Provider (ISP) is. A tracking pixel, i.e., a miniature graphic embedded in such e-mails, which are sent in HTML format, to enable log file recording and analysis. The Data collected are stored in a GDPR compliant newsletter service (MailerLite) and are encrypted. Format of the data collected from the self-registration to the 6G-NTN website providing consent are CSV. Amount of Data: between 100 and 200 records.
6G-NTN_WP6_T1_website _v1.0	 The 6G-NTN website collects the following data: User account data: i.e., account of users authorised to publish content on the website. This are in general users' part of D4P. This may include name, relevant titles and email address. Contact data: i.e., data provided by web sites visitors filling in contact forms or sending email to info@6gntn.eu to request information to the 6G-NTN consortium. This may include: Name and relevant titles Email Job title Company name Company address (in very rare cases) Access log data: i.e., data that are collected by servers to control access to the web pages of the web site that may be used in the event of attacks and other cases requested by the law. This may include (1) the browser types and versions used, (2) the operating system used by the accessing system, (3) the website from which an accessing system reaches our website (so-called referrers), (4) the sub-websites, (6) an Internet Protocol address (IP address), (7) the Internet Service Provider of the accessing system, and (8) any other similar data and information that may be used in the event of attacks on our information technology systems. The format of the data collected from the self-registration to the 6G-NTN website providing consent are: CSV, XML, Json.
6G-NTN_ WP6_T1_events _v1.0	For the 6G-NTN event registration 6G-NTN will make use of the Ti.to platform. The data collected within the Ti.to platform (event management and ticketing platform managed by Team





	 Tito Limited). Ti.to online application collects only the following personal data: email addresses, name and surname, affiliation, type of organisation, country, subscription to newsletter (optional) Format of the data collected is .xls
6G-NTN_WP4_T2_vessel_spatio_temporal density _v1.0	Maritime density map dataset Type: Matrix Int/Float. Format: MAT-file. Amount of data: few Mbits.
6G-NTN_WP4_T2_average_connectivity_speed _v1.0	Broadband data speed for both fixed and mobile networks at the municipalities level. Type: Table. Format: Parquet. Amount of data: hundreds of Mbits.
6G-NTN_WP4_T2_Population density_v1.0	Statistical number of people and the corresponding geographical positions. Type: Table. Format: CSV-file. Amount of data: hundreds of Mbits.





3 FAIR DATA

This DMP follows EU guidelines¹ and describes data management procedures according to FAIR² principles. The acronym FAIR stands for "findable, accessible, interoperable and reuseable" and describes data management practices that allow maximum knowledge circulation and return on investment. The FAIR principles were generated to improve the practices for data management and data-curation, and FAIR aims at describing the principles in order to be applied to a wide range of data management purposes, whether it is data collection or data management of larger research projects regardless of scientific disciplines. With the endorsement of the FAIR principles by Horizon Europe, the FAIR principles serve as a template for lifecycle data management and ensure that the most important components for lifecycle are covered. This is intended as an implementation of the FAIR concept rather than a strict technical implementation of the FAIR principles.

3.1 TYPES OF DATA/RESEARCH OUTPUT

This subsection outlines the practices regarding the management of dataset generated and collected by 6G-NTN project.

Each dataset to be easily identifiable and re-usable in the future will be identified with a unique name. All data files produced include the term "6G-NTN" followed by the Work Package (WP) number, the task number, file name which briefly describes its content, followed by a version number and the file extension (if relevant):

6G-NTN_Work_Package_Number_Task_Number_Brief_description_of_file_content_Version_Number.file_extention

A README file to explain all relevant details regarding data collection, processing methodologies, and quality assurance will be deposited along with the datasets in .odt, .rtf or .pdf format.

3.2 DATA STANDARDS AND METADATA

All datasets will be described using standard descriptive metadata, in order to ensure metadata interoperability for indexing and discoverability. For each deposited dataset, relevant documentation explaining data collection procedures and analysis is made available along with the data, in order to guarantee intelligibility, reproducibility, and the validation of the project findings.

In the following, an initial identification list of standards that will be used to classify type, formats, and size of the data generated and collected is provided:

- Text-based documents: .xlm, .json, .ppt, .pptx, .csv, .txt, .parquet, .mat
- Illustrations and graphic design: .jpg, .png, .psd, .tiff, .ai, .vsd

¹ Horizon Europe Programme Guide: <u>https://ec.europa.eu/info/funding-tenders/opportunities/docs/2021-</u> 2027/horizon/guidance/programme-guide horizon en.pdf

² The FAIR data principles (GOFAIR), <u>https://www.go-fair.org/fair-principles/</u>



- Vector graphics: .svg, .pdf
- Audio files: .mp3, .wav
- Video files: .mp4, .mkv

These file formats have been chosen because they are accepted standards and in widespread use. Files will be converted to open file formats where possible for long-term preservation. In this initial version of the DMP the Metadata are foreseen as a text-based document in order to provide a full explanation of the data. Further metadata formats will be considered in the future if needed.

3.3 ACCESS, SHARING, AND PRESERVATION

This section explains how and to whom the data is going to be made available.

The data created by 6G-NTN will be diversely curated depending on the sharing policies attached to it:

- Confidential/Restricted: Only for project internal use
- Open: immediately available to public under a specific license
- Open with Embargo: available to public after a predefined embargo period

For both open and non-open data, the aim is to preserve the data and make it readily available to the interested parties for the whole duration of the project and beyond.

3.3.1 Non open research data

Restrictions to access are applied only in the following cases:

- collected data belong to third party which have denied permission for sharing outside the consortium members on account of confidentiality and proprietary issues;
- data anonymization is not possible;
- data availability would jeopardize the project's main aim and a specific partner future exploitation plan.

The non-open research data will be stored in Microsoft Sharepoint. Sharepoint, a web based collaborative platform, is currently being employed to coordinate the project's activities and to store all the digital material connected to 6G-NTN. If certain datasets cannot be shared or are subject to any restrictions, legal, and contractual reasons will be justified.

However, metadata will be made openly available and licenced under a "No Rights Reserved" CC0 license or equivalent.

3.3.2 Open research data

As a guiding principle, 6G-NTN project seeks to make all research data openly available as soon as possible and ensure open access via the repository in order to allow dissemination, validation, and re-use of research results.

6G-NTN will ensure the open access to all peer-reviewed scientific publications relating to its results and will provide access to the research data needed to validate the results presented in deposited scientific publications.







FIGURE 2: OPEN ACCESS FLOW.

The two main routes to open access are (shown in Figure 2):

- Self-archiving / 'green' open access: the author, or a representative, archives (deposits) the published article or the final peer-reviewed manuscript in an online repository before, at the same time as, or after publication. Some publishers request that open access be granted only after an embargo period has elapsed.
- Open access publishing / 'gold' open access: an article is immediately published in open access mode. In this model, the payment of publication costs is shifted away from subscribing readers.

6G-NTN will strive to make available as open access many of the collected datasets. When any restrictions apply, the confidentiality section for that particular dataset will describe why access has been restricted.

The use of a specific repository will depend primarily on the primary creator of the publication and on the data in question.

Most likely, when depositing peer-reviewed scientific publications and dataset, 6G-NTN partners will make use either a domain specific repository or the EU recommended service OpenAIRE while project research data will be deposited to online data repository such as ZENODO. ZENODO is a general-purpose open-access repository developed under the European Open AIRE program and operated by CERN. The repository services offered by ZENODO enable peers to share and preserve research data and other research outputs in any size and format: datasets, images, presentations, publications and software. The digital data and the associated meta-data are preserved through well-established practices such as mirroring and periodic backups. In particular, Zenodo's metadata is compliant with DataCite's Metadata Schema, which will be attributed to the dataset when deposited. Different types of persistent identifiers will be considered, such as Archival Resource Keys (ARKs), Electronic Identifier Serial Publications (EISPs), Digital Object Identifiers (DOIs), Uniform Resource Names (URNs), Persistent Uniform Resource Locators (PURLs), etc.





4 ALLOCATION OF THE RESOURCES

The primary goal of this section is the estimation and coverage of financial costs of processing and the identification of the responsibilities for the data management in the project.

Under this assumption, the data management in 6G-NTN will be done as part of the WP1 and UNIBO as Project Coordinator, will be responsible for data management in 6G-NTN project according to the FAIR principles. UNIBO (as responsible for the production of DMP deliverables), TAS-F, MAR, CTTC have allocated budget and person months to execute these activities.

In addition, the costs to make the data FAIR will be handled by each partner who will have to generate them following the eligibility criteria for Horizon 2020 grant conditions.

Lastly, resources for long term preservation, associated costs, and potential value will be discussed at due time by the General Assembly.





5 DATA SECURITY

This section is fundamental from the point of view of GDPR-compliance as expressly state that the data are processed according to data integrity and confidentiality principle. The Article 5.1f requires that data shall be "processed in a manner that ensures appropriate security of the personal data, including protection against unauthorised or unlawful processing and against accidental loss, destruction or damage, using appropriate technical or organisational measures".

For the duration of 6G-NTN project, collected and generated data will be stored on the Microsoft Sharepoint platform as previously mentioned. The used platform will guarantee the security principle in order to protect against any type of modification: the user authentication, the user authorization, the accounting, the communication security, the data integrity and the availability.





6 DATA ETHICAL ASPECTS

The research activity envisaged in the 6G-NTN project will be conducted applying fundamental ethical principles and relevant national, EU and international legislation, including the Charter of Fundamental Rights of the European Union, the European Convention on Human Rights, and the GDPR - Regulation EU 2016/679. Fundamental principles underpinning the whole research activity are: the respect for the welfare (health and safety) of the participants to the research; participation only informed consent; personal data protection, anonymity and confidentiality; fairness; equity; justice; social responsibility.

Informed consent for human participation to the research. All potential participants will be informed in a clear, transparent, and appropriate manner on the basis of their characteristics and capabilities about the research and the related processing of personal data. The participants will be asked to read, fill, and sign consent forms (both regarding participation in the research, and the processing of personal data, if identified as its legal basis) using a radio button, declaring they have read and understood the information received, connecting to a secured server. A copy of those electronic documents will be stored on research group storage system.

Personal data protection. The collection and the processing of personal data will be carried out in full compliance with GDPR and the related national implementing regulations, and in accordance with the relevant provisions of the national Supervisor Authorities. All beneficiaries involved in personal data processing during the project, except some that do not need it in compliance with the GDPR" confirm to have appointed a Data Protection Officer (DPO) and their contact details are made available to all data subjects involved in the project through the privacy information sheets. The processing of personal data will be carried out only for the specific purposes of the project and according to the correct legal bases. In compliance with the principle of minimization, only the personal data necessary for the pursuit of the indicated research purposes will be collected during the project and will be kept for the time necessary to achieve them, as will be specified to the data subjects. The most rigorous security measures will be adopted to ensure the integrity, confidentiality and control of them, in accordance with the provisions of GDPR and with rules established by the European Data Protection Supervisor and Art. 29 Working Party. The data will be kept only for the time necessary to carry out the research and subsequent verification activities. The additional keeping of such data for future research activities will take place only if the data subject expressly authorizes in writing such keeping, as it will be indicated in a specific section of the privacy information notice.

Involvement of non-EU countries: 6G-NTN consortium includes Martel, Digital For Planet, and Thales Alenia Space UK LTD. Activities carried out in Switzerland and UK will be conducted applying the national laws and regulations. These activities are allowed in at least one of the Member States involved in the project (Italy, France, Germany, Luxemburg, Spain, Sweden). Anyways, the ethics standards and guidelines of Horizon Europe will be rigorously applied, regardless of the country in which the research takes place. Activities carried out in UK will respect the UK law and are allowed in the EU countries.





7 CONCLUSION

This deliverable provides an overview of the data that the 6G-NTN project will produce together with related data processes and requirements that need to be taken into consideration. Also, it outlines an overview about the dataset types, defines a set of attributes to be used for describing each dataset, and presents the open access aspects to be followed by the consortium.

The datasets will be incrementally enriched along the project lifetime. 6G-NTN decisions with respect to making the data FAIR and the respective mechanisms to support these decisions are described in Section 3 while the allocation of resources, the data security and ethical aspects are presented in Sections 4, 5, and 6 accordingly.





APPENDIX

A "README" file is a document that will be deposited with each data dataset, containing relevant information about dataset authorship, terms of reuse and responsibilities, explaining dataset content and structure, collection procedures and analysis (such as specifics, methodologies, codebook of variables, data sources, and further necessary notes).

In the following, the template file of the README file is provided.

README file

Dataset Title: "[insert title as defined in the DMP]"

Dataset Author/s: Name Surname (Affiliation), ORCID (if available);

Dataset Contributor/s: Name Surname (Affiliation), ORCID (if available);

Dataset Contact Person/s: Name Surname (Affiliation), ORCID (if available), email;

Dataset License: this dataset is distributed under a [insert LICENSE]

Publication Year: [insert YEAR]

Project Info: [insert PROJECT ACRONYM] ([project full title], funded by European Union, Horizon 2020 Programme. Grant Agreement num. [insert grant agreement number]; [insert project website url]

Dataset Contents

The dataset consists of:

EXAMPLE 1

- 1 textual qualitative file saved in .rtf format: "6G-NTN_WP3_T3-2_ItalyInterviews_20161221_v01.rtf"
- 1 README file: "README_6G-NTN _WP3_T3-2 _ItalyInterviews_20161221_v01.rtf"

Dataset Documentation

Abstract:

[Insert dataset abstract]

Content of the files:

- file [Insert filename] contains ...
- file [Insert filename] contains ...

⇒ ...

File specifics





[Please indicate instruction/technical info in order to allow potential users to correctly visualize and reuse your data (e.g. specific software, ...). In case of data converted in open formats it could be useful to provide some further information. For example, if you deposit for long term preservation a .csv file derived from an excel you can describe the conversion. Here is an example of description of conversion using libre office calc software:

To create the .csv files, "LibreOffice Calc" version: 5.1.4.2 (portable) was used, with the following specifics:

- Character set Europa occidentale (Windows-1252/WinLatin1)
- ➡ Field delimiter « , » (comma)
- Text delimiter « " » (quotes)]

Notes

[Related to the whole dataset or to single files of a multi-file dataset (Optional)]

Data sources

[Optional]

Methodologies

[If necessary to understand how to reuse data]

Codebook of variables

[If necessary to understand the meaning of the variables]

