



D1.4 Data Management Plan

Dissemination level:	Public (PU)
Work package:	WP1
Task:	T1.3
Deliverable lead:	VICOMTECH
Version:	V1.0
Submission date:	28/02/2022
Due date:	28/02/2022

Partners:



This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101016427

www.5g-iana.eu

Authors

Authors in alphabetical order		
Name	Organisation	Email
Andrea Suárez	Vicomtech	asuarez@vicomtech.org

Control sheet

Version history			
Version	Date	Modified by	Summary of changes
V0.1	29/11/2021	Andrea Suárez	Initial ToC
V0.2	27/01/2022	Andrea Suárez	First Draft
V0.3	28/01/2022	Marcos Nieto	Formatting and revision
V0.4	02/02/2022	Andrea Suárez	Pre-final version for review
V0.5	14/02/2022	Eirini Liotou	Minor inputs
V0.6	14/02/2022	Andrea Suárez	Consolidated version for review
V0.7	24/02/2022	Andrea Suárez, Marcos Nieto	Added modifications from peer-review
V1.0	25/02/2022	Andrea Suárez	Consolidated version for submission

Peer review		
	Reviewer name	Date
Reviewer 1	Eirini Liotou (ICCS)	22/02/2022
Reviewer 2	Dimitris Klonidis (UBI)	23/02/2022

Legal disclaimer

The information and views set out in this deliverable are those of the author(s) and do not necessarily reflect the official opinion of the European Union. The information in this document is provided “as is”, and no guarantee or warranty is given that the information is fit for any specific purpose. Neither the European Union institutions and bodies nor any person acting on their behalf may be held responsible for the use which may be made of the information contained therein. The 5G-IANA Consortium members shall have no liability for damages of any kind including without limitation direct, special, indirect, or consequential damages that may result from the use of these materials subject to any liability which is mandatory due to applicable law.

Copyright © 5G-IANA Consortium, 2021.

TABLE OF CONTENTS

TABLE OF CONTENTS	3
1. INTRODUCTION	7
1.1. 5G-IANA concept and approach.....	7
1.2. Purpose of the deliverable	8
1.3. Intended audience	8
2. DATA SUMMARY.....	9
2.1. Purpose and relation to the objectives	9
2.2. Data categories.....	9
2.2.1. Technical data	10
2.2.2. Evaluation data	11
2.2.3. Open research data.....	11
2.2.4. Internal administrative data	12
2.2.5. Data on project outcomes and studies	12
2.3. Datasets description	12
2.3.1. General data description.....	12
2.3.2. Template used for 5G-IANA dataset description	12
3. FAIR DATA.....	14
3.1. Making data findable, including provisions for metadata	14
3.2. Making data openly accessible	15
3.3. Making data interoperable.....	16
3.4. Increase data re-use.....	17
4. ALLOCATION OF RESOURCES	18
5. DATA PROTECTION AND ETHICAL ASPECTS.....	19
5.1. Compliance with GDPR	19
5.2. Data protection	19
5.3. Ethical aspects	20
6. DATA SECURITY	21
7. CONCLUSION.....	23

List of tables

Table 1, Technical data subcategories.....	10
Table 2, Template 5G-IANA dataset description	13

ABBREVIATIONS

Abbreviation	Definition
5G-IANA	5G for Intelligent Automotive Network Applications
5G-PPP	5G Infrastructure Public Private Partnership
AGV	Automated Ground Vehicle
AI	Artificial Intelligence
AOEP	Automotive Open Experimental Platform
API	Application Programming Interface
ASAM	Association for Standardization of Automation and Measuring Systems
CAN	Controller Area Network
CSV	Comma-Separated Values
DML	Distributed Machine Learning
DMP	Data Management Plan
DMPO	Data Manager and Protection Officer
ETSI	European Telecommunications Standards Institute
EU	European Union
FAIR	Findable, Accessible, Interoperable, Reusable
GDPR	General Data Protection Regulation
GNSS	Global Navigation Satellite System
HTTPS	Hypertext Transfer Protocol Secure
ISO	International Organization for Standardization
JSON	JavaScript Object Notation
KPI	Key Performance Indicator
MANO	Management and Orchestration
MEC	Multi-Access Edge Computing
ML	Machine Learning
MPEG	Moving Picture Experts Group
OBU	On-Board Units
PCD	Point Cloud Data
RSU	Road Side Units
SME	Small Medium Enterprise
VNF	Virtualised Network Function
WP	Work Package
XML	Extensible Markup language

Executive Summary

WP1 of 5G-IANA project targets to ensure an efficient overall management of the 5G-project consortium. In particular, Task 1.3 is devoted to **data management** procedures. This deliverable, D1.4 - Data Management Plan (DMP), aims at providing a detailed description of the data to be generated during the implementation of the project, whether and how it will be exploited or made accessible for verification and re-use, and how it will be curated and preserved. Thus, the objective is to provide guidelines that will define simple and practical pointers during the implementation and validation stages of the project, support compliance on underlying legal obligations and promote the adoption of best data management practices.

The deliverable has been structured according to the European Commission's guidelines¹ as follows:

- **Chapter 1** - Introduction to 5G-IANA project's concept and approach, and to the deliverable.
- **Chapter 2** - Data summary describes the purpose of the DMP, its relation to the project's objectives, and types of data.
- **Chapter 3** - FAIR data explains how data will be made findable, accessible, interoperable and re-used.
- **Chapter 4** - Allocation of resources explains how allocation of required resources for data management is implemented.
- **Chapter 5** - Data protection and ethical aspects details the methodology to be followed to reach compliancy with data protection regulation and consideration of ethical aspects.
- **Chapter 6** - Data Security explains the approach towards guaranteeing data security.
- **Chapter 7** - Conclusion.

¹ https://ec.europa.eu/research/participants/docs/h2020-funding-guide/cross-cutting-issues/open-access-data-management/data-management_en.htm#A1-template

1. INTRODUCTION

1.1. 5G-IANA concept and approach

5G-IANA aims at providing an open 5G experimentation platform, on top of which third party experimenters (i.e., SMEs) in the Automotive-related 5G-PPP vertical will have the opportunity to develop, deploy and test their services. An Automotive Open Experimental Platform (AOEP) will be specified, as the whole set of hardware and software resources that provides the computation and communication/transport infrastructure as well as the management and orchestration components, coupled with an enhanced NetApp Toolkit tailored to the Automotive sector. 5G-IANA will expose to experimenters secured and standardized APIs for facilitating all the different steps towards the production stage of a new service. 5G-IANA will target different virtualization technologies integrating different MANO frameworks for enabling the deployment of the end-to-end network services across different domains (vehicles, road infrastructure, MEC nodes and cloud resources). 5G-IANA NetApp toolkit will be linked with a new Automotive VNFs Repository including an extended list of ready to use open accessible Automotive-related VNFs and NetApp templates, that will form a repository for SMEs to use and develop new applications. Finally, 5G-IANA will develop a distributed AI/ML (DML) framework, that will provide functionalities for simplified management and orchestration of collections of AI/ML service components and will allow ML-based applications to penetrate the Automotive world, due to its inherent privacy preserving nature. 5G-IANA will be demonstrated through 7 Automotive-related use cases in 2 5G SA testbeds. Moving beyond technological challenges, and exploiting input from the demonstration activities, 5G-IANA will perform a multi-stakeholder cost-benefit analysis that will identify and validate market conditions for innovative, yet sustainable business models supporting a long-term roadmap towards the pan-European deployment of 5G as key advanced Automotive services enabler.

1.2. Purpose of the deliverable

The purpose of this deliverable D1.4 - Data Management Plan is to give an insight into data management and the lifecycle of the data that will be generated and used.

The aim of the Data Management Plan is to be a living document which will be updated during the project, when a significant change occurs; an updated version will be issued under D1.5 – Data management plan V2 on month 18.

1.3. Intended audience

The dissemination level of this document is “public” (PU) and is primarily intended to serve as an internal guideline and reference for the appropriate data management of the 5G-IANA project.

2. DATA SUMMARY

2.1. Purpose and relation to the objectives

5G-IANA aims at providing an open 5G experimentation platform to enable the development, deployment and testing of Automotive related 5G applications. Data will be generated for the implementation, operation, analysis and validation of the use cases, and with the primary objective of validating 5G-IANA software outcomes and perform the related KPIs' assessment. In this sense, data gathering is closely related to the following objectives of the project:

- **Objective 1:** Specify and provide an Automotive Open Experimental Platform.
- **Objective 4:** Provide accurate localization and low latency mission-critical applications.
- **Objective 5:** Define, implement and trial Connected and Automated Driving relevant use cases to validate and assess the AOEP suitability and functional improvements.
- **Objective 8:** Increase road safety and reduce automobile carbon footprint by leveraging Connected and Automated Mobility using enhanced network performances.

Data will be also collected as part of T6.1 - Market research and actor-role analysis. It will be used to identify the factors that can affect the market adoption of 5G-IANA. In this sense, data gathering is related to the following objective of the project:

- **Objective 7:** Create new business opportunities and boost market for start-ups and SMEs with Automotive NetApps.

2.2. Data categories

5G-IANA will handle different types of data which can be organised into four categories:

1. **Technical data:** related to the technical development and operation of the use cases.
2. **Evaluation data:** related to testing and evaluation processes.

3. **Open research data:** data and results that might be published by the project to leverage further work in the research community.
4. **Internal administrative data:** data generated/shared internally for administrative and management purposes.
5. **Data on project outcomes and studies:** data generated internally from managerial, technical and scientific activities for reporting project achievements.

More details on the corresponding categories are provided in the upcoming subsections.

2.2.1. Technical data

The technical data category includes the data related to the technical developments and is generated during the development phases. More precisely, in the context of 5G-IANA, it consists of data produced by the vehicles' systems, either from intra-vehicular and extra-vehicular sensors and On-Board Units (OBUs), or logs about internal vehicle information, and also from infrastructure-level equipment, including Road Side Units (RSUs), and Multi-Access Edge Computing (MECs) nodes.

Considering the 5G-IANA use cases, technical data can be divided into subcategories as summarised in Table 1:

Table 1, Technical data subcategories

Subcategories	Description	Examples
Sensors' vehicle external data	Data describing the vehicle's environment as captured by on-boarded sensors.	Video, LIDAR.
Infrastructure data	Data describing the roadside infrastructure including roadside elements characteristics and status.	Network condition models, OBU/RSU bitrate and latencies, number of worker nodes (MEC).
Environmental data	Environmental conditions measured with sensing elements.	Humidity, Road traffic information, level of gases (CO, NOx), smoke level, fire detection, visibility.

Vehicle Position, Dynamics and Attributes	Data providing information on the position and the dynamics of the vehicle from intra-vehicular or extra-vehicular sensors.	CAN data (includes velocity), GNSS, number of occupants, AGV log files, driving behaviour data.
Scene and Traffic-related data	Data describing the scene in terms of amount, features and other data from road participants (vehicles, pedestrians) as observed by infrastructure devices or obtained from communication protocols.	Mobility data, number of vehicles with specified requirements, smartphone data.
Network monitoring data	Data collected through monitoring processes within the platform framework and relate to performance metrics and status of elements (both computing and network elements).	CPU, Memory usage on servers, status of VNFs, node loads, link loads, deployed application status, connectivity status.

Technical data may contain personal data and, in those cases, the GDPR rules apply.

Regarding the sharing of the technical data, it is up to the data generators to decide what they accept to be shared among other partners. However, the data sharing needs to ensure proper development from all partners involved.

2.2.2. Evaluation data

This category shall contain all data necessary to the project's evaluation tasks. More specifically, the evaluation data will support the evaluation of the platform, the use cases and new services enabled by the 5G-IANA platform.

The evaluation data are a subset of the operational data (data produced during the span of the project). Evaluation data are those data relevant for evaluation tasks. These data will be produced in the evaluation tasks defined under WP5. The exact description of the evaluation data will be done under Task 5.1 mainly.

Similar to technical data, evaluation data may contain personal data and the GDPR rules apply in those cases.

2.2.3. Open research data

The open research data category includes data and results that will be made openly available at the end of the project. Further details are provided in section 3.2 of this deliverable.

2.2.4. Internal administrative data

It refers to the data produced by the project management activities such as meeting minutes, recordings, internal reports, for historical purposes and follow-up. This category of data is collected by the management team including the project manager, the WP leaders and task leaders. The data are stored using a project management tool that requires the authentication of the users (Redmine). The internal administrative data are confidential, only for members of the Consortium.

2.2.5. Data on project outcomes and studies

This concept refers to data created during the project and reported as outcomes of managerial, financial, technical, or scientific activities. This category, therefore, includes material like deliverables, scientific and technical papers, or presentations. These data are intended to be published, delivered, or exported to external parties for their inspection and analysis beyond the limits of the consortium.

2.3. Datasets description

2.3.1. General data description

This section provides guidelines on how to describe the different types of datasets to be collected and shared by 5G-IANA. The description of the different datasets, including their reference, file format, standards, methodologies, metadata and repository to be used are given below.

2.3.2. Template used for 5G-IANA dataset description

This section provides the template to be used to describe 5G-IANA dataset.

Table 2, Template 5G-IANA dataset description

Role	Who
Dataset reference	Each dataset will have a reference that will be generated by the combination of the name of the project, the location, datatype and date.
Dataset name	Name of the dataset.
Standards and metadata	<p>The metadata attributes' list to be used to find the dataset. Metadata can be split into 4 categories:</p> <ul style="list-style-type: none"> • design and execution documentation, which corresponds to a high-level description of data collection. • descriptive metadata, which describes each component of the dataset (including origin and quality). • structural metadata, which describes how the data are being organized. • administrative metadata, which set the conditions for how the data can be accessed and how this is being implemented.
File Format	The format of the file that contains the data.
Data sharing	<p>Explanation of the sharing policies related to the dataset between the next options:</p> <ul style="list-style-type: none"> • 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. It should provide information about personal data and mention if the data are anonymized or not. It should tell if the dataset entails personal data and how this issue is considered.
Archiving and Preservation	The preservation guarantee and the data storage during and after the project (for example databases, institutional repositories, public repositories, etc.).

3. FAIR DATA

3.1. Making data findable, including provisions for metadata

Aside from justified exceptions, all the research data produced and/or used in the project will be discoverable with previously agreed metadata tags (name, media presentation description, subtitles, tags, timeline content such as chat messages, multisensorial outputs), identifiable and locatable by means of standard identification mechanisms such as Digital Object Identifiers².

Search keywords will be provided to optimize possibilities for re-use and the following naming convention will be followed:

5G-IANA_[DATA-TYPE] _[DESCRIPTIVE-NAME] _[VERSION-NUMBER]

To facilitate finding but also handling research data produced by the 5G-IANA project, clear version numbers and release notes (when applicable) will be provided.

To the maximum extent possible, metadata standards will be followed. When needed, extensions to the existing standards will be used in the project and proposed as a contribution to the bodies responsible for maintaining those standards.

In a nutshell, in order to be able to retrieve and find data, the requirements have been listed below:

- (meta)data will be assigned a globally unique and persistent identifier
- data will be described with rich metadata
- metadata will clearly and explicitly include the identifier of the data it describes
- (meta)data will be registered or indexed in a searchable resource

² <https://www.doi.org/>

3.2. Making data openly accessible

Most of the data collected during the development of 5G-IANA's project are intended to feed the development (e.g., training Machine Learning models), and testing (e.g., measure network capabilities, latencies) stages of the project; therefore, in general, data will not be made public either because it will not be useful out of the scope of the defined 5G-IANA use cases, because it is synthetic data, or because it may be produced by proprietary testbeds.

Nevertheless, given data will be certainly made publicly available:

- analyses based on the collected data and aimed at computing specific KPIs and evaluating the DML framework will be made public and be presented in project deliverables and related publications.
- partners with research profiles may produce data with wider usage out of 5G-IANA use cases (created data or metadata).

In this context, 5G-IANA will follow an Open-Source policy and has agreed to use Zenodo to store all data that will be openly available, such as project deliverables and related publications (public deliverables will be also available in the project's website). A private GitLab installation will be used to store all data that will not be openly available but will be shared by consortium members. Besides, code produced within the 5G-IANA project will be deposited on GitHub and on secure servers.

The results of the research study will be made available to the public in an aggregated and anonymised form, and open access to research data will be granted. Moreover, open access to publications, 'gold' or 'green' open access³ will be granted.

With regards to internal administrative data, it is considered internal working material, and will therefore be considered confidential for the internal use in reporting, under the 5G-IANA ethics protocols. GDPR needs to

³ https://ec.europa.eu/research/participants/docs/h2020-funding-guide/cross-cutting-issues/open-access-dissemination_en.htm

be applied so as to protect data and privacy, therefore, before making any content available, all data will be anonymised.

3.3. Making data interoperable

Data interoperability is achieved by means of the utilisation of international standards for data formatting (to provide harmonised structure), and via the definition or selection of vocabularies or ontologies (to provide semantics or meaning to data).

Regarding format, international standardisation bodies (e.g., ETSI, ISO, ASAM, etc.), will be examined looking for published standards and recommendations that define data formats for the specific use cases. In case no standard is defined, or the field of application is too specific, data format shall be detailed and defined from project activities. Data files and payloads may in any case follow de-facto industry/academical formats (e.g., JSON and/or XML files for structured textual content, CSV text files for matrix data, H.264/5 or MPEG for video files, PCD for point clouds, etc.), to facilitate the utilisation of data with commonly used programming languages and tools.

Regarding semantics, in some cases, internationally acknowledged vocabularies or ontologies do not exist for a particular data type/use case. In these cases, if it is found unavoidable to use uncommon or to generate project specific ontologies or vocabularies, the first approach will be to extend the existing ones (e.g., more generic vocabularies and ontologies from generic domains). If this option is not feasible, then mappings to more commonly used ontologies will be pursued.

Whatever approach is followed, documentation shall be created which specifies the origin (external or project-specific) of the vocabularies and ontologies used to define the nomenclature required to provided meaning to data.

To be able to exchange data, the following requirements will be met:

- (meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation.
- (meta)data use vocabularies that follow FAIR principles.
- (meta)data include qualified references to other (meta)data.

The data gathered will be analysed using open coding methodologies.

3.4. Increase data re-use

As already stated under section 3.2, in general, data will be created and consumed during the project and will be only relevant in the 5G-IANA context.

However, for the data to be publicly available in Open Access platforms, the licences managed by such platforms (i.e., Common Creative 4.0 for Zenodo) will be adopted.

4. ALLOCATION OF RESOURCES

All research data collected as part of this project is owned by the data producers or partners involved in the use cases. 5G-IANA partners will take the responsibility for the collection, management, and sharing of the research data.

VICOM leads the data management plan tasks and ensures project coordination in terms of the validation data collection, storage and handling. As coordinator of the process, VICOM also ensures that the data handled over the course of the project serves the project's objectives by following up on the procedures to make data FAIR, secure and GDPR compliant.

With regards to Open Access to scientific publications, all 5G-IANA partners that intend to publish scientific papers have a minimum budget of 3,000 EUR for conference and other dissemination costs (on top of travel and expenses). These budget allocations should allow covering (at least partially) of the costs of complying with the open data requirements of the project.

5. DATA PROTECTION AND ETHICAL ASPECTS

5.1. Compliance with GDPR

The approach proposed to handle the GDPR in 5G-IANA project consists of identifying all the concerned parties and the actions they need to take in order to comply with the regulation. This approach will enforce that personal data collected in 5G-IANA shall be (Article 5 of Regulation (EU) 2016/679 (General Data Protection Regulation):

- Processed lawfully, fairly and in a transparent manner in relation to individuals.
 - Collected for specified, explicit and legitimate purposes and further processed for scientific purposes.
 - Adequate, relevant and limited to what is necessary for the purposes for which they are processed.
 - Kept in a form which permits identification of data subjects for no longer than is necessary for the purposes for which the personal data are processed.
 - 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.
- Further refinement of the above-mentioned approach will be provided in next versions of this deliverable.

5.2. Data protection

VICOM also acts as Data Protection Officer for the 5G-IANA project: Andrea Suárez has been appointed as Data Management and Protection Officer (DMPO). This Manager/Officer raises potential issues and proposes solutions for dealing adequately with data privacy and data protection regulations, and will also liaise with the partners who will perform the demonstrations and the testbed members to ensure proper application of the Data Protection policies at the national level.

5.3. Ethical aspects

Further information on ethical issues related to the protection of personal data in 5G-IANA are addressed on deliverable D8.1 H – Requirement No. 1.

6. DATA SECURITY

The data produced during 5G-IANA will be stored per trial site in local servers and in central servers for the whole project. Those data will be processed in compliance with the GDPR. This chapter describes some security principles to be implemented in order to protect against any type of modification. The security principles are listed below:

- **Authentication:** users requiring access to 5G-IANA data servers should be authenticated; also, proper means are used to authenticate the servers.
- **Authorization:** access to 5G-IANA data servers is only available to the authenticated and authorized users. These categories and the rights of those users are defined and enforced. The appropriate access control policies and mechanisms (including physical access control) shall be identified for each trial site and project wide to provide the authorization.
- **Accounting:** any access and modification to a resource by any user is securely logged in order to prevent users from denying that data files were accessed, altered or deleted. Other accounting mechanisms shall be implemented.
- **Confidentiality:** data stored in 5G-IANA servers shall be encrypted during transmission and storage.
- **Communication Security:** access to 5G-IANA servers shall be done through encrypted communication channels such as HTTPS.
- **Availability:** this security principle assures that 5G-IANA servers shall be available for 5G-IANA users during the defined interval of service. Also, regular backups of the data shall be made.

Partners' private data will be stored under partners' premises using existing data storage/recovery/backup plan. No sensitive data will be stored. In order to avoid unauthorised access to data, partners are required to have as many security measures as possible, technical measures (such as firewalls, access controls, access audits, etc.) as well as operational ones (e. g. training, incidence reporting, etc.).

All open data will be safely stored in Zenodo, where a private GitLab installation will be also used to store all data that will not be openly available but will be shared by consortium members. Besides, code produced within the 5G-IANA project will be deposited on GitHub and on secure servers.

Internal administrative data are considered internal working material and will therefore be considered confidential for the internal use in reporting, under the 5G-IANA ethics protocols. GDPR needs to be applied so as to protect data and privacy, therefore, before making any content available, all data will be anonymised.

7. CONCLUSION

This deliverable D1.4 Data Management Plan represents the initial release of 5G-IANA Data Management Plan. Throughout this document we have identified the data that will be generated and gathered, how to make it Findable, Accessible, Interoperable and Re-usable, the allocation of resources needed to maintain the data FAIR, and contemplated the ethical and security aspects of data collection.

This initial version of the Data Management Plan has been completed with rough information and will be further completed as the project progresses. An updated version of the Data Management Plan will be issued under D1.5 – Data management plan V2 on M18, although significant changes will be tracked during the development of the project.