D1.4 Data Management Plan

Dissemination level:	Public (PU)
Work package:	WP1
Task:	T1.5
Deliverable lead:	Vicomtech
Version:	V2.0
Submission date:	28/02/2022 and 19/10/2022
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**
Edoardo Bonetto	LINKS	edoardo.bonetto@linksfoundation.com
Dimitris Klonidis	UBI	dklonidis@ubitech.eu
Eirini Liotou	ICCS	eirini.liotou@iccs.gr
Francesca Moscatelli	NXW	f.moscatelli@nextworks.it
Andrea Suárez	Vicomtech	asuarez@vicomtech.org
Markus Wimmer	NOKIA	markus.wimmer@nokia.com

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
V1.1	20/09/2022	Andrea Suárez, Marcos Nieto	Updates addressing comments raised by the reviewers during First Interim Review
V1.2	30/09/2022	Andrea Suárez	Updated data categories, data openly accessible, data security, and conclusion. Created new section Data Management.
V1.3	5/10/2022	Andrea Suárez	Added inputs from partners
V1.4	7/10/2022	Eirini Liotou	Pre-final version revision
V1.5	10/10/2022	Andrea Suárez, Eirini Liotou	Consolidated version for peer-review
V2.0	14/10/2022	Andrea Suárez	Final version

Peer review		
	Reviewer name	**Date**
Review 1	(VICOM) Marcos Nieto	23/2/2022

Review 2	(ICCS) Eirini Liotou	25/2/2022
Review 3	(ICCS) Eirini Liotou	7/10/2022
Review 4	(UBI) Dimitris Klonidis	10/10/2022

Legal disclaimer

TABLE OF CONTENTS

List of Tables

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
UC	Use Case
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 guidelines as follows:

- **Chapter 1** – Introduction to 5G-IANA project's concept and approach, and to the deliverable.
- **Chapter 2** – Data Management describes the purpose of the DMP, its relation to the project's objectives, and the DMP online tool.
- **Chapter 2** – Description of the types of data to be generated by the project, and statement about their openness.
- **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. 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 and openly accessible Automotive-related VNFs and NetApp templates, that will form a repository for SMEs to reuse, modify 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 21.

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. The public nature of this deliverable facilitates potential third party experimenters (during the project lifetime) to understand what kind of data will be made available to them, either in the form of information (such as datasets, deliverables, videos, etc.) or in the form of technical assets provided to them (such as access to network resources for experimentation). Similarly, the provided information can be useful for any interested parties even after the project lifetime (for further exploitation).

2. DATA MANAGEMENT

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 platform (AOEP) and the testing 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. DMP online tool

The Data Management Plan is intended to be a living document that will be updated when relevant changes arise. Moreover, it will be regularly monitored and consulted. With this aim, the project has decided to use the DMP online tool in order to facilitate the involvement, collaboration and contributions from all partners during the implementation of the project.

3. DATA SUMMARY

5G-IANA will generate and handle different types of data which can be organised in the following main categories:

- **Technical data**: related to the technical development and operation of the AOEP and the deployed use cases.
- **Evaluation data**: related to testing and evaluation processes.
- **Internal administrative data**: data generated/shared internally for administrative and management purposes.
- **Data on project outcomes and studies**: data generated from managerial, technical and scientific activities for reporting project achievements.
- **Data for third party engagement**: data to facilitate the promotion of the AOEP, and to attract third party experimenters through the project's open calls.

These categories are described in the upcoming subsections.

3.1. Technical data

3.1.1. Generic technical data

The generic technical data category includes the data related to the technical developments and are generated during the development, validation and demonstration phases. More precisely, in the context of 5G-IANA, it consists of data produced by the AOEP monitoring mechanisms at the network and application function level and 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 and declared Openness
Sensors' vehicle external data	Data describing the vehicle's environment as captured by on-boarded sensors.	UC1: Video: Open; LIDAR information: Open after anonymization UC3: Anonymized recording of the 360° video stream produced during the Virtual tour Bus: Open
Infrastructure data	Data describing the 5G network infrastructure characteristics and performance.	OBU/RSU bitrate and latencies: Open
Environmental data	Environmental conditions measured with sensing elements.	UC7: Humidity, level of gases (CO, CO2), temperature: Restricted
Vehicle Dynamics and Attributes	Data providing information on the dynamics of the vehicle from intra-vehicular or extra-vehicular sensors.	UC1: Vehicle speed: Open UC2: CAN data, GNSS data, IMU data: Open UC5: Driving behaviour data: Open after pseudonymisation
Vehicle Position	Data providing information on the position of the vehicle using the OBU.	UC6: Vehicle route/coordinates: Open
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	UC1: Information on number of vehicles/persons/traffic signals detected: Open UC2: Occupancy grid data: Open

	obtained from communication protocols.	UC3: Mobility data from the vehicle: Open UC5: Mobility data from vehicle and smartphone, smartphone use data: Restricted[1] UC7: Restricted
Network monitoring data	Data collected through monitoring processes within the platform framework and related 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: Open
Application monitoring data	Data collected at runtime through monitoring processes at the application orchestrator level and related to a set of user-defined operational parameters with respect to the deployed application functions.	CPU, RAM, Disk usage: Open (Application specific data may include confidential information, so they will be restricted)
Local resources monitoring data	Data collected through monitoring the resources and capabilities of OBUs, RSU, edge servers.	Available resources, health and utilization of the resources: Open

3.1.2. Datasets

The project will support data collection and the generation of corresponding datasets from any UCs where this is feasible and meaningful. Datasets will be primarily produced by each UC as well as from the selected third party contributions. Data are collected from the attached end user equipment and the monitoring mechanisms (see Table 1) following the experimentation procedures and according to the testing and experimentation processes that will be defined in D5.2.

[1] **Smartphone users will be notified that some data are retrieved from their smartphone, but that such data are confidential.**

As an example, data will be collected through the UC6-related NetApp (Network Status Monitoring UC), that will be responsible for sensing, pre-processing and aggregation of data. This NetApp will be further also available to complement a DML pipeline or even centralized ML training pipelines potentially proposed by third parties. On top of that, a separate measurement campaign in the City of Ulm will be conducted to further enhance the volume of data collected in the dataset (using vehicles). This specific dataset will hopefully contribute to filling the gap related to the relevant lack of datasets for AI/ML purposes.

The expected datasets from all UCs will be defined in the course of time, once all the details on UC developments have been finalised within WP5, and will be reported in updated versions of this document (D1.5).

All the collected datasets will be made available as open access to the public. A generic dataset template to be used as a guideline for the description of the dataset is provided in Annex I.

3.1.3. AOEP and NetApp toolkit related data

The types of raw log data collected from the AOEP and NetApp toolkit software components include:

- the NetApps catalogue entries and the registered NetApp, Application and Network Function parameters,
- the integration and performance monitoring data related to the developed AOEP software modules,
- the data from NetApp on network monitoring, and
- the data from the application orchestrator monitoring engine related to the monitoring of user defined operational parameters for the deployed application functions at runtime.

These data will serve as the basis for the computation of the KPIs related (but not limited) to: NetApps onboarding time, service creation, provisioning and termination time, identifying the contributions related to the allocation and configuration of resources in single domains. The data will be collected during software integration and testing activities and as part of the trials' execution. The raw logs will be provided as textual files, while the KPIs' statistics related to onboarding and service lifecycle timings will be provided as CSV files. All the data will be made available in the software repository of the relevant components, in a dedicated folder and will be also available for open access.

3.1.4. Local resources monitoring data

In the framework of the 5G-IANA project, third parties' developers will be using the OBUs and RSUs provided by LINKS for deploying their applications. The OBUs and RSU will be accessible through the available interfaces that are provided by the 5G-IANA AOEP. Resources monitoring data, position and other information, that are relevant to the applications deployed, will be collected on the OBUs and on the RSUs. All these collected data will be made available for open access.

3.1.5. Software code

Most of the developed Application and Network Functions (AFs/NFs) will be provided as open-source or open-use, while others might be IPR-protected. Similarly, some of the platform components will be provided as open-source, or at least open-use. The project has already identified which SW AFs/NFs/platform components will be provided under which scheme. The possible categories of openness available are: a) Open-source, b) Open-use, c) Open-use following an NDA, d) Open-use only within the AOEP platform context, e) Open-use only by consortium partners, f) Restricted.

This information will be included in documents related to IPR.

3.1.6. Testbeds access

5G-IANA will provide online access to NOKIA's 5G network to external experimenters (third parties) through the AOEP platform. Specifically:

1) A MEC Server for all UCs in NOKIA OnAir Testbed Ulm will be available, while its capabilities will be open and are described below:

- CPU: AMD Ryzen 9 5900X (12 cores)
- GPU: NVIDIA RTX 3080 TI
- RAM: 128GB
- SSD: 2TB
- HDD: 8TB
- NIC: 2*10G SFP+
- OS: Ubuntu server 22.04 LTS
- Root rights: only NOKIA

The MEC server(s) can only be accessed via project partners from preconfigured static IP addresses.

2) The 5G network capabilities (excluding the MEC Servers) can be outlined in terms of available radio resources, as they are the key limiting factor. The 5G capabilities in NOKIA OnAir Testbed Ulm are summarized below:

- At least 2 antenna sites; radio capabilities for testing purposes are temporarily granted to the test network and may change during the project period:
    - 20 MHz in band b38 (2.6 GHz, TDD) or 40 MHz in band b38 (2.6 GHz, TDD, risk of restricted use).
    - 10 MHz in band b28 (700 MHz, FDD) (restricted use, risk of unavailability from 2023 onwards).

HW used within the 5G network (such as used radio units, used fibre optic multiplexing technology, UPF HW, etc.) is not provided, as the 5G network provides transmission capabilities as a service.

Access to the NOKIA/TS testbeds will be granted on the basis of terms and conditions of the respective testbed owner.

3.2. Evaluation data

3.2.1. Data from UC evaluations, testing and demos

This category shall contain all data deriving from the project's evaluation tasks. The evaluation data will entail the evaluation of the platform, of the use cases and of new services enabled by the 5G-IANA platform. More specifically, data generated from the internal experimentation performed through the project UCs validation/evaluation will be made available together with the related NetApp Packages.

The evaluation data are a subset of the operational data (data produced during the span of the project). These data will be produced in the evaluation tasks defined under WP5. In more detail, data will be collected from attached end user equipment and the monitoring mechanisms following the experimentation procedures to be defined in D5.2. The exact description of the evaluation data will be done under Task 5.1 mainly.

In the next version of this deliverable, we will report also on end-to-end experience-related data for the running service with respect to each UC. The evaluation data can be provided as open access (e.g., KPIs measurements), as long as there is such an interest from third parties or other interested bodies.

3.2.2. Data from third party experimentation

Third party NetApp developers (outside of the 5G-IANA consortium) will also typically create data during the usage of the testbed and the AOEP platform. In other words, it is expected that data will be generated by the NetApps that will be developed and deployed by external SMEs as part of their experimentation in the 5G-IANA platform, following the project's open calls. Such data coming from third parties' experimentation could be a useful source of information for assessing the overall 5G-IANA solution (AOEP, NetApps etc.).

Data ownership in such a case remains with the NetApp developer or experimenter (includes any type of data produced by third party experimenters, e.g., datasets). Note that no personal devices will be used, while application data will be considered like the UCs-specific data (Table 1).

3.3. 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, videos, or presentations. Moreover, it includes outcomes from surveys conducted within the context of WP6, i.e., feedback from third parties or in general any interested stakeholders. These data are intended to be published, delivered, or exported to external parties for their inspection and analysis beyond the limits of the consortium.

Publications, public deliverables, newsletters, communication materials, videos, webinars, surveys, etc. will be **openly** accessible through the 5G-IANA Zenodo Community as well as through the project's website and social media where applicable. Open access to publications, 'gold' or 'green' will be granted.

All project deliverables will be released as **Public**, except for the ones related to exploitation (D7.7 and D7.10) or to techno-economic data (D6.4). An amendment will handle this issue (namely to open the respective deliverables where applicable).

3.4. Data for third party engagement

For the purposes of reaching out to third parties, and making an effort a) to attract their interest and b) to facilitate their experimentation on the AOEP platform, the project will produce the following material (to be provided **openly** online on a specific "For Experimenters" section on the 5G-IANA website):

- A general project presentation (including a simplified version of the AOEP architecture, testbeds description, etc.);
- Rules and procedures for third parties' experimentation (to help ensure that any required conditions are met for the admission of third parties to the project testbeds);
- Guides for applicants of the open calls, eligibility criteria (i.e., a document that contains all information for the submission of a proposal from third parties);
- Technical manual (user guide). This manual will be a useful tool to help the experimenters better understand how UCs and NetApps can be implemented and how they can run on the AOEP platform (as a guide/example to develop their own UCs). Some examples of the information to include are: how third parties can access the platform, technical requirements if any for their applications, options/capabilities that the AOEP provides, manual for onboarding AFs/NFs, creating NetApps and distributed Vertical Services, type of open interfaces offered, manual regarding the available NetApp repository, KPI monitoring options, reference to technical support, etc.;
- Technical support information;
- Open-source SW/AFs/NFs/NetApps git including for instance instructions for uploading NetApps/NFs/AFs in the project's repo for using the whole pre-validation pipeline, instructions for using the pre-validation pipeline to validate packages and on-board them to the AOEP platform, instructions related to the usage of NetApps/AFs/NFs available in the NetApp Toolkit (i.e., licensing), and instructions for reading the content of available NetApps/AFs/NFs packages (e.g., test cases/scripts, results, shared data-set, KPIs).
- Open calls / hackathon / micro-projects related information.
- Examples of use cases, and specifically: a) description of the 5G-IANA UCs as baseline examples of possible services to be realized), b) examples of customizations applicable to the 5G-IANA UCs, and c) examples of additional UCs/services that can be realized using the 5G-IANA NetApps/AFs/NFs;
- FAQ

In more detail, regarding the AFs/NFs provided by the project, a manual with their description will be provided (Description, Input required, Output provided, Examples of communicating AFs/NFs), so that third parties can understand how they could combine their proprietary functions with 5G-IANA's AFs/NFs, in order to form new NetApps.

We highlight that 5G-IANA will provide explicit and detailed information in the form of a user guide for experimenters (including a webinar, that will be also recorded and made available online), explaining how to access and use the platform.

3.5. 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 internal working material and will therefore be considered **confidential** for the internal use in reporting, under the 5G-IANA ethics protocols.

4. FAIR DATA

4.1. 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 Identifiers.

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

4.2. Making data openly accessible

The aim of consortium is to make data generated by the project openly accessible to the fullest extent possible. In this regard, and as described in Section 2 of this document, almost all data generated during the lifetime of the project will be made openly accessible except for confidential deliverables and internal administrative data.

In general, the project will make these data openly accessible through the 5G-IANA Zenodo Community (Annex II), through the project's website and through GitHub/GitLab. The project GitLab will host mostly the open-source software, so the AOEP components, the AFs/NFs and the NetApps that will be released as open-source

by the project partners. 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.

Data collected during validation and experimentation phases can be made accessible along with the produced software. In case of proprietary solutions, the code will reside in the company-specific repositories, while results from the experimentation can be made available and open.

GDPR needs to be applied so as to protect data and privacy, therefore, before making any content available, all data that fall into this category will be always anonymised.

4.3. 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.

4.4. Increase data re-use

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.

In order to also facilitate the reutilisation of the produced data, its nature will be documented (when data themselves are not self-explanatory) and provided along with the respected data, explaining its format, data model, used standards, method of extraction, means of manipulation, etc.

5. ALLOCATION OF RESOURCES

All research data will be collected by the data producers 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 serve 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 budget allocated 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.

6. DATA PROTECTION AND ETHICAL ASPECTS

6.1. 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.

In general, each partner (data generator and processor) will be responsible for complying with the relevant rules. However, central (project-level) monitoring of such compliance will be ensured as well.

6.2. 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.

6.3. Data protection within NOKIA's network

NOKIA is responsible for the "mobile network" component of the 5G-IANA project. The mobile network component does not collect or process any data worthy of protection within the meaning of the EU-GDPR.

The "cellular technology" establishes data connections (IP) for other components of the 5G-IANA project. The data connections are characterized by certain connection qualities (Quality of Service, QoS), for example the latency. Other components of 5G-IANA are connected to the cellular network via UEs (user equipment) or gateways. Users are identified via SIM cards, and their QoS requirements for data transport are stored in NOKIA's prototypical mobile network.

Moreover, the MEC-server HW is the responsibility of NOKIA. The MEC-server represents the platform on which, the central software components of the project run.

The central software components do not lie in the responsibility of NOKIA but lie within the project partner who provides the respective software. NOKIA does not collect any data. If any data are collected or processed worthy of protection within the meaning of the EU-GDPR - in case of UCs where this applies (e.g., UC3: virtual bus tour) -, then the data collection and processing lie in the sole responsibility of the project partners running the respective UC. No data worthy of protection within the meaning of the EU-GDPR shall be stored on the MEC-Servers of NOKIA.

6.4. 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.

7. 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. These 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.

8. 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 by the project, how to make it Findable, Accessible, Interoperable and Re-usable, the allocation of resources needed to maintain the data FAIR, and the ethical and security aspects of data collection.

This initial version of the Data Management Plan summarises the intention of the project in relation to the data to be openly accessible, described in Section 2. This Plan will be further updated as the project progresses.

ANNEX I: GENERIC DATASETS DESCRIPTION

This Annex 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.

The following table provides the generic template to be used to describe a 5G-IANA dataset. The fields will be accustomed based on the specific dataset to be produced, so as to better reflect its objective and value.

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	The metadata attributes' list to be used to find the dataset. Metadata can be split into 4 categories: • 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	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. 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.).