D1.5 Data management plan V2

Dissemination level:	Public (PU)
Work package:	WP1
Task:	T1.3
Deliverable lead:	VICOM
Version:	V1.0
Submission date:	28/02/2023
Due date:	28/02/2023
Partners:	

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

Authors

Authors in alphabetical order		
Name	**Organisation**	**Email**
Alexis Aivaliotis	O7	aaivaliotis@oseven.io
Andrea Suárez	VICOM	asuarez@vicomtech.org
Claudio Russo	BYL	claudio.russo@bylogix.it
Dimitris Klonidis	UBI	dklonidis@ubitech.eu
Edoardo Bonetto	LINKS	edoardo.bonetto@linksfoundation.com
Eirini Liotou	ICCS	eirini.liotou@iccs.gr
Fotis Foukalas	COGN	fotis@cogninn.com
Francesca Moscatelli	NXW	f.moscatelli@nextworks.it
Manuel Fuentes	5COMM	manuel.fuentes@fivecomm.eu
Marios Zinonos	HIT	m.zinonos@hit-innovations.com
Markus Wimmer	NOKIA	markus.wimmer@nokia.com
Nehal Baganal Krishna	UULM	eirini.liotou@iccs.gr
Rudolf Susnik	ININ	rudolf.susnik@iinstitute.eu

Control sheet

Version history			
Version	**Date**	**Modified by**	**Summary of changes**
V0.1	14th December 2022	Andrea Suárez	Initial ToC
V0.2	19th January 2023	Andrea Suárez	Initial content
V0.3	1st February 2023	Andrea Suárez	Inputs from UC Leaders
V0.4	17th February 2023	Andrea Suárez	Minor updates from UC Leaders and NOKIA
V0.5	22nd February 2023	Andrea Suárez	Integrated input from peer review
V1.0	27th February 2023	Eirini Liotou	Final version ready for submission

Peer review		
	Reviewer name	**Date**
Reviewer 1	Dimitris Klonidis (UBI)	20/2/2023
Reviewer 2	Eirini Liotou (ICCS)	27/2/2023

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
EC	European Commission
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
nApp	Network Application

Executive Summary

WP1 of 5G-IANA project targets to ensure an efficient overall management of the project. In particular, Task 1.3 is devoted to data management procedures. This deliverable, D1.5 - Data Management Plan v2 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. The deliverable is based on the first version of the data management plan (D1.4) and has been updated with the following new information:

- Data Summary (Section 3) has been updated, the expected datasets from all UCs have been identified (Section 3.1.2 - Datasets) and the data for third party engagement already produced have been added (Section 3.2.4 - Data for third party engagement).
- A data collection monitoring methodology implemented (Section 4.2 - Making data openly accessible).
- Data already made openly accessible by the project until M21 (Annex I).

The deliverable has been structured according to the EC'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 3** – Description of the types of data to be generated by the project, and statement about their openness.
- **Chapter 4** – FAIR data explains how data will be made findable, accessible, interoperable and re-used.
- **Chapter 5** – Allocation of resources explains how allocation of required resources for data management is implemented.
- **Chapter 6** – Data protection and ethical aspects details the methodology to be followed to reach compliancy with data protection regulation and consideration of ethical aspects.
- **Chapter 7** – Data Security explains the approach towards guaranteeing data security.
- **Chapter 8** – 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 vertical sector will have the opportunity to develop, deploy and test their services. The provided Automotive Open Experimentation Platform (AOEP) is a set of hardware and software resources that provides the computational and communication/transport infrastructure as well as the management and orchestration components, coupled with an enhanced nApp Toolkit tailored to the Automotive sector, for simplifying the design and onboarding of new nApps. 5G-IANA exposes to experimenters secured and standardized Application Programming Interfaces (APIs) for facilitating all the different steps towards the production stage of a new service. 5G-IANA targets different virtualization technologies integrating different Management and Orchestration (MANO) frameworks for enabling the deployment of end-to-end network services across different segments (vehicles, road infrastructure, Multi-access Edge Computing (MEC) nodes and cloud resources). 5G-IANA nApp toolkit is linked with an Automotive Virtual Network Functions (VNFs) Repository including an extensive portfolio of ready-to-use and openly accessible Automotive-related VNFs and nApp templates, that are available for SMEs to use and develop new applications. Finally, 5G-IANA develops a Distributed Machine Learning (DML) framework, that provides functionalities for simplified management and orchestration of collections of Machine Learning (ML) service components and thus, allows ML-based applications to penetrate the Automotive world, due to its inherent privacy-preserving nature. 5G-IANA will be demonstrated through seven Automotive-related use cases in two 5G Stand Alone (SA) testbeds. Moving beyond technological challenges, and exploiting input from the demonstration activities, 5G-IANA will identify and validate market conditions for innovative, yet sustainable business models for the AOEP platform, supporting a long-term roadmap towards the pan-European deployment of 5G as a key advanced Automotive services enabler.

1.2. Purpose of the deliverable

The purpose of this deliverable D1.5 - Data Management Plan v2 is to complete the initial version of the Data Management Plan (D1.4) providing the datasets that the project expects to generate from each of the Use Cases and how these will support the third party experimentation during the project.

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 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 (AOEP).
- **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 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 nApps.

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. Withthis 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 implementationof 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. Examples are provided in the last column of this Table, while more detailed explanation for the UC-driven datasets is available in Section 3.1.2 that follows.

Table 1, Technical data subcategories

Subcategories	Description	Examples & openness level
Sensors' vehicle external data	Data describing the vehicle's environment as captured by on-boarded sensors.	UC1, UC3
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
Vehicle Dynamics and Attributes	Data providing information on the dynamics of the vehicle from intra-vehicular or extra-vehicular sensors.	UC1, UC2, UC5
Vehicle Position	Data providing information on the position of the vehicle using the OBU.	UC6
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.	UC1, UC2, UC3, UC5, UC7
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 supports 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. The datasets per Use Case that have been defined by the project are presented at Tables 2-8. Data are collected from the attached end user equipment (mobile device, OBU, etc.) and the monitoring mechanisms, following the experimentation procedures and according to the testing and experimentation processes that will be defined in D5.2.

3.1.2.1. Use Case 1 - Remote driving

Table 2, Use Case 1 datasets

UC1				
Lead partner	Fivecomm	Fivecomm	Fivecomm	Fivecomm
Data owner	Fivecomm	Fivecomm	Fivecomm	Fivecomm
Dataset description	Recording of the processed video sent to the cockpit after NF/AF processing	Data processed by the AI object detection algorithm	LiDAR information coming from sensors in the vehicle	Vehicle position, speed and angle
Data origin	Video cameras located at the vehicle	Video cameras located at the vehicle	Sensors	Users' cockpit
Data type	Compressed video	Numerical values	Numerical values	Numerical values
Data purpose	Results analysis and showcase	Results analysis and showcase	Results analysis and showcase	Results analysis and showcase
Reusability potential (especially by third parties)	The videos can be used for comparison and KPI evaluation	Data will be available for comparison and KPI evaluation	Comparison with other LiDARs	Map generation, graphs and comparison results
Which partner(s) will use the data?	Fivecomm	Fivecomm (alternatively Vicomtech)	Fivecomm (alternatively Vicomtech)	Fivecomm
Data format	RTP/H.264	JSON	JSON	JSON
Dissemination level: - *Open* - *Embargo* - *Restricted*	Embargo	Open	Open	Open
FAIR: Will the data be shared in publicly open database?	No (because of privacy legal reasons)	Yes	Yes	Yes

Data Storage (Location, Duration)	MEC server of the project's infrastructure Duration: as long as the demo lasts	MEC server of the project's infrastructure Duration: as long as the demo lasts	MEC server of the project's infrastructure Duration: as long as the demo lasts	OBU
Publication date (tentative)	End of workplan cycle A	End of workplan cycle A	End of workplan cycle A	End of workplan cycle A

3.1.2.2. Use Case 2 - Manoeuvres coordination for autonomous driving

Table 3, Use Case 2 datasets

UC2			
Lead partner	BYL	BYL	BYL
Data owner	BYL	BYL	BYL
Dataset description	Data providing information on the dynamics of the vehicle from intra-vehicular or extra-vehicular sensors	Data describing the scene in terms of amount, features and other data from anonymized road participants (vehicles, pedestrians) or obstacles	KPI related experiment measurements collected during the demonstration of the UC to evaluate KPIs relevant to the 5G-IANA project
Data origin	Test vehicles, simulated vehicles	Test vehicles, simulated vehicles	Will be recorded during field trial preparation and/or during the field trials
Data type	GNSS coordinates and confidence, IMU data, CAN data	Occupancy grid data	Numerical recordings of the KPI
Data purpose	Inputs for the manoeuvre coordination service	Actor and obstacle information for the manoeuvre coordination service	Evaluate KPIs relevant to the 5G-IANA project
Reusability potential (especially by third parties)	None, this data is volatile and relates to the specific vehicles and situation involved in UC2. Furthermore, as UC2 is limited to a couple of demonstrations, we cannot provide more than a few hours of data, which is not relevant enough for exploitation.	None, this data is volatile and relates to the specific vehicles and situation involved in UC2. Furthermore, as UC2 is limited to a couple of demonstrations, we cannot provide more than a few hours of data, which is not relevant enough for exploitation.	Analysis and showcase of the potentialities of 5G-IANA, specifically the nApp and functions utilized in UC2. It may also serve as a benchmark for third parties aiming at creating services, functions or nApps

			relevant to what is presented by UC2.
Which partner(s) will use the data?	BYL	BYL	BYL, any partners interested in network service level KPIs for UC2
Data format	Binary data	Binary data	CSV
Dissemination level: - *Open* - *Embargo* - *Restricted*	Restricted	Restricted	Open
FAIR: Will the data be shared in publicly open database?	No, due to its volatility	No, due to its volatility	Yes
Data Storage (Location, Duration)	Will be temporarily buffered in the vehicles/simulator hardware	Will be temporarily buffered in the vehicles/simulator hardware	In Zenodo, as part of a larger DB for all UCs, or a DB which may be expanded by other tests related to UC2
Publication date (tentative)	No publication due to its temporary nature	No publication due to its temporary nature	Between validation and project end

3.1.2.3. Use Case 3 - Virtual bus tour

Table 4, Use Case 3 datasets

UC3					
Lead partner	HIT	HIT	HIT	HIT	HIT
Data owner	HIT	HIT	HIT	HIT	HIT
Dataset description	Anonymized 360º 4k video recorded from a vehicle touring the city of Ulm. Will be aligned to the Position and time recordings dataset also produce in UC3.	Position and time recordings from a vehicle touring the city of Ulm. Will be aligned to the 360º 4k video dataset also produce in UC3.	Anonymous results from questionnaires answered by the UC3 Users and stakeholders concerning the UC3 VR system usability and willingness to pay	KPI related experiment measurements collected during the demonstration of the UCs to evaluate KPIs relevant to the 5G-IANA project	3D models that will be created for the purpose of the use case.

Data origin	Will be recorded during field trial preparation and/or during the field trials.	Will be recorded during field trial preparation and/or during the field trials. Data provided by a service developed in the project.	Questionnaires answered by the Use Case 3 Users and stakeholders	Will be recorded during field trial preparation and/or during the field trials.	Developed for the needs of the project
Data type	Video	Spatial and temporal data	Numerical and ordinal data	Numerical recordings of the KPI	3D models will be saved in fbx format
Data purpose	Create a video recording that can be used by third parties	Create a spatial dataset that can be used by third parties	Evaluate UC3 VR system usability and willingness to pay	Evaluate KPIs relevant to the 5G-IANA project	UC3 preparation
Reusability potential (especially by third parties)	If combined with the spatial and temporal open dataset also provided with UC3, it can be used to train AI models similar to the one used for video slicing in UC3.	If combined with the video recording also provided by UC3, it can be used to train AI models similar to the one used for video slicing in UC3.	N/A	N/A	N/A
Which partner(s) will use the data?	HIT, 3rd parties	HIT, 3rd parties	HIT	HIT	HIT
Data format	.mp4	CSV		CSV	
Dissemination level: - *Open* - *Embargo* - *Restricted*	Open	Open	Restricted, statistics related to the data will be presented in the relevant deliverables	Restricted, statistics related to the data will be presented in the relevant deliverables	Restricted
FAIR: Will the data be shared in	Yes	Yes	No	No	No

publicly open database?					
Data Storage (Location, Duration)	Open database, released for open use	Open database, released for open use	HIT premises, Duration of the project	HIT premises, Duration of the project	HIT premises, infinite
Publication date (tentative)	TBA	TBA	TBA	TBA	TBA

3.1.2.4. Use Case 4 - AR content delivery for vehicular networks

Table 5, Use Case 4 datasets

UC4		
Lead partner	COGN	COGN
Data owner	COGN	COGN
Dataset description	Location of the mobile user and its nearby place data including their type such as restaurants, café, market, banks and gas station.	Mobile users data rate participating in the use case.
Data origin	The dataset will be collected and saved to the MEC server during the preparation for the field trials and during the field trials. The data source will be the mobile user.	The dataset is a form of reports from mobile user to the MEC server.
Data type	Text in Java Script Object Notation (JSON)	Tha dataset is a list of numbers, I.e. the datarate from mobile user to MEC.
Data purpose	To detect nearby places for the corresponding AR content delivery through COGN's AR streaming application.	The data is collected to apply the AR content scheduling at the MEC server.
Reusability potential (especially by third parties)	The datasets can be used to analyse the behaviour of mobile users in certain areas. This could be done through edge AI analytics.	The datasets can be used to analyse the behaviour of mobile users in certain areas. This should be done through edge AI analytics.
Which partner(s) will use the data?	COGN	COGN and NOKIA
Data format	The dataset will be in Json (Java Script Object Notation) format	8-bit number
Dissemination level: - *Open* - *Embargo* - *Restricted*	Open	Open

FAIR: Will the data be shared in publicly open database?	Yes	Yes
Data Storage (Location, Duration)	The dataset will be stored into an open database released for open use (e.g., Zenodo).	The dataset can be stored into an open database released for open use (e.g., Zenodo).
Publication date (tentative)	During Validation Phase of the Project	During Validation Phase of the Project

3.1.2.5. Use Case 5 - High risk driving detection

Table 6, Use Case 5 datasets

UC5			
Lead partner	OSeven	OSeven	OSeven
Data owner	OSeven	OSeven	OSeven
Dataset description	Position and time data indicating the location (latitude, longitude), speed and altitude of the vehicle at a particular time, using the Position and Time Service AF.	Use of mobile phone during driving.	High-risk hotspots indicating risk levels along roads.
Data origin	Will be recorded during the preparation for the field trials and during the field trials. The data source will be the OBU.	Will be recorded during the preparation for the field trials and during the field trials. The data source will be the mobile phone.	Will be recorded during the preparation for the field by the ML model developed during the UC. The data source will be the OBU and the mobile phone.
Data type	Spatial and temporal data	Time series	Spatial and temporal data
Data purpose	To detect hazardous events performed during driving (harsh braking, harsh acceleration, speeding).	To detect mobile phone use during driving.	To provide information based on aggregated data to drivers regarding the average risk level of a road.
Reusability potential (especially by third parties)	If these two datasets are combined, experimenters can try to understand how mobile use affects driving behaviour and possibly correlate them.		Can be used by road operators and public authorities to provide warning signals to drivers when entering high-risk roads.

Which partner(s) will use the data?	OSeven, 3rd parties	OSeven, 3rd parties	OSeven, 3rd parties
Data format	CSV	CSV	CSV
Dissemination level: - *Open* - *Embargo* - *Restricted*	Open (after anonymization)	Open (after anonymization)	Open
FAIR: Will the data be shared in publicly open database?	Yes	No	Yes
Data Storage (Location, Duration)	This dataset will be stored in an open database released for open use.	This dataset will be stored in a restricted database.	This dataset will be stored in an open database released for open use.
Publication date (tentative)	After the first round of experiments (exact date not known yet, although planned in Q4/2023)	After the first round of experiments (exact date not known yet, although planned in Q4/2023)	After the first round of experiments (exact date not known yet, although planned in Q4/2023)

3.1.2.6. Use Case 6 – Network status monitoring

Table 7, Use Case 6 datasets

UC6			
Lead partner	UULM	UULM	UULM
Data owner	UULM	UULM	UULM
Dataset description	The UC6 collected dataset will be a latency and data rate dataset between the Far Edge and the Edge.	Layer 1 Measurements of the 5G Network.	KPI related experiment measurements collected during the demonstration of the UC6 to evaluate KPIs relevant to the 5G-IANA project.
Data origin	The data is collected in Ulm where Nokia's 5G Base station (site DRK) can accessed.	The data are collected in Ulm, where Nokia's 5G Base station (site DRK) can accessed.	Will be recorded during field trial preparation and/or during the field trials.
Data type	The contents of the dataset will be the location of the OBU (latitude and longitudinal coordinates of OBU), the round-trip latency (in msec) and the data rate (in Mbps) between Edge and Far Edge.	The dataset will contain the 5G physical layer metrics such as SINR, Cell ID, and RSRP, along with the GPU coordinates of the Far Edge.	Numerical recordings of the KPI.

Data purpose	The data is collected to train the LSTM model for pQoS.	To better understand the 5G network behaviour.	Evaluate KPIs relevant to the 5G-IANA project.
Reusability potential (especially by third parties)	To help experimenters develop their own UCs.	To help experimenters develop their own UCs.	N/A
Which partner(s) will use the data?	UULM, ICCS	UULM, ICCS	UULM, ICCS
Data format	CSV File	CSV	CSV
Dissemination level: - *Open* - *Embargo* *Restricted*	Open	Open	Restricted, statistics related to the data will be presented in the relevant deliverables
FAIR: Will the data be shared in publicly open database?	Yes	Yes	No
Data Storage (Location, Duration)	The dataset can be stored on Zenodo.	The dataset can be stored on Zenodo.	UULM Premises, Duration of the Project.
Publication date (tentative)	During Validation Phase of the Project.	During Validation Phase of the Project.	TBA

3.1.2.7. Use Case 7 - Situational awareness in cross border road tunnel accidents

Table 8, Use Case 7 datasets

UC7		
Lead partner	ININ	ININ
Data owner	ININ	ININ
Dataset description	Data related to the UC #7 execution, i.e., “situational awareness” application data.	Various 5G network parameters data related to the execution of the UC #7.
Data origin	Synthesized data and measured data originating from TS testbed.	Measured data originating from both testbed networks (Nokia, TS).
Data type	There will be three types of data: (1) environmental (synthetic) data such as temperature, humidity, gas levels, (2) (synthetic) data collected by OBU (e.g., vehicle velocity, etc.), (3) video data (synthetic data and "person acting") – recordings of the road traffic from the (simulated) road-tunnel.	Data will be results of monitoring/measuring various 5G network parameters in 5G network(s): user data rate (DL, UL) RTT, jitter, packet loss rate.

Data purpose	To simulate data necessary for the Use Case implementation and testing.	Validation and verification of the UC#7 related KPIs, further network analysis.
Reusability potential (especially by third parties)	Data can be used by 3rd parties to verify their specific use cases (e.g., upgrades to UC7), for simulating "tunnel" environment and vehicles in it, to understand tunnel environment conditions or to re-create specific environment.	Data can be used for simulating 5G environment and/or comparing them with results collected in other environments, as well as to verify some specific solutions, use cases, etc.
Which partner(s) will use the data?	any 5G-IANA partners interested, including 3rd party SMEs	Any 5G-IANA partners interested, including 3rd party SMEs
Data format	json, xml, mpeg4	SQL
Dissemination level: - *Open* - *Embargo* - *Restricted*	Restricted	Restricted
FAIR: Will the data be shared in publicly open database?	No	No
Data Storage (Location, Duration)	Location: locally at ININ data center (access from public internet on request) Duration: same as project duration	Location: locally at ININ data center (access from public internet on request) Duration: same as project duration
Publication date (tentative)	After the first round of experiments (exact date not known yet, although planned in Q4/2023)	After the first round of experiments (exact date not known yet, although planned in Q4/2023)

3.1.2.8. NOKIA dataset

Lead partner	Nokia
Data owner	Nokia
Dataset description	Drive test results
Data origin	Nokia, Testbed Ulm
Data type	---
Data purpose	Provides information about cell capacities in the Nokia testbed, Ulm; The drive test results may provide e.g. UL & DL cell capacity information, RTT between UE and MEC-Server, and RSRP values
Reusability potential (especially by third parties)	Provides cell capacity information for one antenna site in the Nokia testbed, Ulm; Note: the cell capacity information may change with modified radio configurations and used frequency range;

Which partner(s) will use the data?	The data is available for all partners
Data format	*.kml (G-Nettrack Lite log files) *.log or *.txt (iperf log files) *.txt and *.txt (qMON log files)
Dissemination level: - *Open* - *Embargo* *Restricted*	Open
FAIR: Will the data be shared in publicly open database?	Yes
Data Storage (Location, Duration)	Zenodo
Publication date (tentative)	Whenever new drive test results are available

3.1.2.9. AOEP and nApp toolkit related data

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

- the nApps catalogue entries and the registered nApp, Application and Network Function parameters,
- the integration and performance monitoring data related to the developed AOEP software modules,
- the data from nApp 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: nApps 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.**

3.1.3. 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**.

3.1.4. 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 (e.g., D7.7 – Exploitation plan).

3.1.5. 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 address.

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 temporarilygranted 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 unavailabilityfrom 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 nApp 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.

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. Some of the UCs have included the description of the expected evaluation data already into the datasets identified in Section 3.1.2.

3.2.2. Data from third party experimentation

Third party nApp 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 nApps 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, nApps etc.).

Data ownership in such a case remains with the nApp 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).

3.2.3. 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).

3.2.4. 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, called "Get Involved"):

- A "**Guide for Applicants**" document related to the open calls, i.e., a document that contains all information for the submission of a proposal from third parties. Its first version has been already released together with the launch of the 5G-IANA Open Call #1.
- **Technical manual (user guide):** This manual will be a useful tool to help the experimenters better understand how Use Cases (UCs) and nApps can be implemented, onboarded and run on the AOEP platform (as a guide/example to develop their own UCs), what monitoring KPIs are offered, etc. Some examples of the information to be included 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 nApps and distributed Vertical Services, type of open interfaces offered, manual regarding the available nApp repository, KPI monitoring options, reference to technical support. Moreover, it will include examples of UCs, specifically: a) 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 nApps /AFs/NFs.
- **Manual of the Application Functions (AFs), Network Functions (NFs) and Network Applications provided by the platform**: It will contain the Description, Input required, Output provided, and Examples of communicating AFs/NFs, so that applicants/experimenters can understand how they could combine their proprietary functions with 5G-IANA's AFs/NFs, in order to create tailored nApps. Moreover, these functions could be the basis for inspiration to SMEs for developing new functions and nApps.

It is worth noting that webinars / recorded videos will also be openly available to encourage and support the third party experimentation.

3.2.5. 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 coordinator, the WP leaders and task leaders. The data are stored using a project management tool that requires the authentication of the users (SharePoint). 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 I), through the project's website (data coming from the project outcomes and studies and communication materials mainly), and through GitHub/GitLab. The project GitLab will host mostly the open-source software, so the AOEP components, the AFs/NFs and the nApps 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.

With the aim of coordinating and supervising the collection and publication of the data identified by the project to be provided open during the course of the project, T1.3 has implemented a monitoring methodology. This methodology foresees periodic reviews (every three months) with the Consortium in order to facilitate the exchange of information and the contributions from all partners (with special emphasis to Use Case Leaders), and to support and advise the partners managing this type of data.

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 ofthe research data.

VICOM leads the data management plan tasks and ensures project coordination in termsof 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 intendto 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.

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 toprotect 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 D1.5 Data Management Plan represents the updated release of 5G-IANA Data Management Plan (D1.4). On this document, the project has identified all datasets that are expected to be generated by each Use Case during the project: part of these datasets will be provided open for experimenters to better understand how Use Cases and nApps are implemented and therefore enabling them to run their own Use Cases.

The document also describes the data monitoring methodology that has been defined by T1.3 to coordinate the publication of the data identified by the project to be provided open during the course. This methodology will facilitate the continuous collection of data during the implementation of the project. This information will be reported on the periodic and final reports.

ANNEX I - DATA OPENLY ACCESSIBLE UNTIL M21

Table 9, Data openly accessible until M21

Name	Type of data	Type of publication	Publication date
5G-IANA: A 5G Experimentation Platform for Intelligent Automotive Network Applications	Publication	Open Access	May-21
On the Resource Consumption of Distributed ML	Publication	Open Access	Jul-21
5G-IANA D1.2 Quality management plan	Deliverable	Open Access	Nov-21
5G-IANA D7.1 Brand identity and guidelines	Deliverable	Open Access	Nov-21
A Distributed ML Framework for Service Deployment in the 5G-based Automotive Vertical	Publication	Open Access	Dec-21
5G-IANA #1 Newsletter - January 2022	Communication material	Open Access	Jan-22
5G-IANA D7.2 Communication strategy and plan	Deliverable	Open Access	Feb-22
5G-IANA D7.4 Communication tools	Deliverable	Open Access	Feb-22
5G-IANA Flyer	Communication material	Open Access	Mar-22
5G-IANA Roll-Up Banner	Communication material	Open Access	May-22
5G-IANA Poster	Communication material	Open Access	May-22
Resource Allocation for Distributed Machine Learning at the (Edge-Cloud) Continuum	Publication	Open Access	May-22
Distributed Machine learning for Network Monitoring and predictive QoS in Automotive Applications	Publication	Open Access	Jun-22
5G-IANA Logo	Communication material	Open Access	Jul-22
5G-IANA #2 Newsletter - July 2022	Communication material	Open Access	Jul-22
5G-IANA - D1.1 - Project management plan	Deliverable	Open Access	Oct-22
5G-IANA D1.3 Innovation Management Plan	Deliverable	Open Access	Oct-22
5G-IANA - D1.4 Data Management Plan	Deliverable	Open Access	Oct-22

5G-IANA - D2.1 Specifications of the 5G-IANA architecture	Deliverable	Open Access	Oct-22
5G-IANA - D5.1 - Initial validation KPIs and metrics	Deliverable	Open Access	Oct-22
5G-IANA - D7.6 Dissemination Plan	Deliverable	Open Access	Oct-22
The 5G-IANA platform: Bringing far-edge resources and ML potential to the disposal of automotive third parties	Publication	Open Access	Oct-22
5G-IANA brochure	Communication material	Open Access	Nov-22
5G-IANA #3 Newsletter - November 2022	Communication material	Open Access	Nov-22
Open Call opportunity for SMEs in the automotive sector	Presentation	Open Access	Jan-23
5G-IANA OPEN CALL #1 Guide for Applicants	Open Call conditions	Open Access	Feb-23