



5GMOBIX

5G for cooperative & connected automated
MOBility on

X-border corridors

D1.4

INITIAL DATA MANAGEMENT PLAN

Dissemination level	Public (PU)
Work package	WP1: Project coordination
Deliverable number	D1.4
Version	V1.1
Submission date	30/04/2019



This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 825496.

www.5g-mobix.com¹

Authors

Authors in alphabetical order		
Name	Organisation	Email
Djibrilla AMADOU KOUNTCHE	AKKA	djibrilla.amadou-kountche@akka.eu
Sadeq ZOUGARI	AKKA	sadeq.zougari@akka.eu

Control sheet

Version history			
Version	Date	Modified by	Summary of changes
0.1	18/02/2019	Sadeq ZOUGARI (AKKA), Djibrilla AMADOU KOUNTCHE (AKKA)	ToC
0.2-0.5	12/04/2019	Sadeq ZOUGARI (AKKA), Djibrilla AMADOU KOUNTCHE (AKKA)	Draft versions
1.0	12/04/2019	Sadeq ZOUGARI (AKKA), Djibrilla AMADOU KOUNTCHE (AKKA)	Pre-final version ready for peer-review
1.1	30/04/2019	Sadeq ZOUGARI (AKKA), Djibrilla AMADOU KOUNTCHE (AKKA)	Pre-final version (revised)

Peer review		
	Reviewer name	Date
Reviewer 1	Vasilis SOURLAS (ICCS)	19/04/2019
Reviewer 2	Sébastien FAYE (LIST)	19/04/2019

Legal disclaimer

The information and views set out in this deliverable are those of the author(s) and do not necessarily reflect the official opinion of the European Union. The information in this document is provided "as is", and no guarantee or warranty is given that the information is fit for any specific purpose. Neither the European Union institutions and bodies nor any person acting on their behalf may be held responsible for the use which may be made of the information contained therein. The 5G-MOBIX Consortium members shall have no liability for damages of any kind including without limitation direct, special, indirect, or consequential damages that may result from the use of these materials subject to any liability which is mandatory due to applicable law. Copyright © 5G-MOBIX Consortium, 2018.

Table of contents

EXECUTIVE SUMMARY	8
1. INTRODUCTION	9
1.1. 5G-MOBIX concept and approach	9
1.2. Purpose of the deliverable	9
1.3. Intended audience.....	10
2. DATA IN 5G-MOBIX: AN OVERVIEW	11
2.1. Datasets Categories	11
2.1.1. Context Data	11
2.1.2. Acquired and Derived Data.....	12
2.1.3. Metadata.....	13
2.2. Data management methodology in 5G-MOBIX.....	14
2.3. Data Flow in 5G-MOBIX	15
2.4. 5G-MOBIX shared dataset description.....	16
2.4.1. General data description	16
2.4.2. Template used for 5G-MOBIX dataset description.....	17
3. PARTICIPATION IN THE OPEN RESEARCH DATA PILOT	18
4. FAIR DATA MANAGEMENT PRINCIPLES	20
5. ALLOCATION OF RESOURCES	22
6. DATA PROTECTION REGULATIONS AND ETHICAL ASPECTS	23
6.1. An approach to comply with GDPR in 5G-MOBIX	23
6.2. Data protection regulation in the Non-EU Countries	24
6.3. Ethical Aspects	24
7. DATA SECURITY.....	25
8. CONCLUSION	26
9. ANNEXES	27
9.1. Annex 1 – Open research data pilot (ORD Pilot)	27
9.2. Annex 2 – Initial GDPR application in 5G-MOBIX	29
9.2.1. Terminology.....	29

9.2.2. Data subject's rights	29
9.2.3. Controller's and Processor's obligations	29
9.2.4. GDPR templates used for 5G-MOBIX data processing	29
9.2.4.1. Data controller's record of processing activities	30
9.2.4.2. Data processor's record of processing activities.....	31

List of figures

Figure 1: Overview of 5G-MOBIX data management	14
Figure 2: 5G-MOBIX CCAM Infrastructure	16
Figure 3: Illustration of the access to 5G-MOBIX repository in ZENODO	19

List of tables

Table 1: Overview of the categories and types of data collected in 5G-MOBIX	12
Table 2: 5G-MOBIX Dataset description template	17

ABBREVIATIONS

Abbreviation	Definition
ADAS	Advanced Driver Assistance Systems
CCAM	Cooperative, Connected and Automated Mobility
DMP	Data Management Plan
DoA	Description of Action
DPIA	Data Protection Impact Assessment
DPO	Data Protection Officer
EC	European Commission
FAIR	Findable, Accessible, Interoperable and Reusable
FESTA	Field operational test support action
FOT	Field Operational Tests
GA	General Assembly
GDPR	General Data Protection Regulation
KPI	Key Performance Indicator
OBU	On-Board-Units
ORDP	Open Research Data Pilot
POPD	protection of personal data
PU	Public
TMT	Technical Management Team
ToC	Table of Content
VRU	Vulnerable Road Users
WP	Work Package

EXECUTIVE SUMMARY

This document is the deliverable D1.4 – Initial Data Management Plan of 5G-MOBIX. The main objective of this deliverable is to describe the processes used to manage the data produced in the 5G-MOBIX project.

The Data Management Plan is an important document for efficient data management. It describes the data management life cycle for the data to be collected, processed and/or generated by a Horizon 2020 project and the process of making research data findable, accessible, interoperable and re-usable (FAIR).

Data in 5G-MOBIX are produced during the trials (WP4) using the specification from WP2 and are used during the evaluation phase (WP5). The exploitation of these data complies with the General Data Protection Regulation (GDPR) and the subset that will be publish for research purposes will be done according to the Open Research Data Pilot (ORDP) approach described in Chapter 3.

The rest of the document is organized as follows:

- Chapter 1 – Introduction briefly presents 5G-MOBIX and describes the purpose of the document and its intended audience.
- Chapter 2 – Data in 5-MOBIX outlines a data overview in the 5G-MOBIX project. It details 5G-MOBIX data categories, the data management methodology and data flow, and provides a template that will be used to describe datasets in future update(s).
- Chapter 3 – Participation in open research data pilot gives insights about the Open Research Data Pilot under H2020 guidelines.
- Chapter 4 describes the FAIR Data Management principle.
- Finally, Chapters 5-7 provide insights about resources allocation, GDPR & Ethical aspects and data security.

1. INTRODUCTION

1.1. 5G-MOBIX concept and approach

5G-MOBIX aims to showcase the added value of 5G technology for advanced Cooperative, Connected and Automated Mobility (CCAM) use cases and validate the viability of the technology to bring automated driving to the next level of vehicle automation (SAE L₄ and above). To do this, 5G-MOBIX will demonstrate the potential of different 5G features on real European roads and highways and create and use sustainable business models to develop 5G corridors. 5G-MOBIX will also utilize and upgrade existing key assets (infrastructure, vehicles, components) and the smooth operation and co-existence of 5G within a heterogeneous environment comprised of multiple incumbent technologies such as ITS-G5 and C-V2X.

5G-MOBIX will execute CCAM trials along cross-border (x-border) and urban corridors using 5G core technological innovations to qualify the 5G infrastructure and evaluate its benefits in the CCAM context. The Project will also define deployment scenarios and identify and respond to standardisation and spectrum gaps.

5G-MOBIX will first define critical scenarios needing advanced connectivity provided by 5G, and the required features to enable some advanced CCAM use cases. The matching of these advanced CCAM use cases and the expected benefits of 5G will be tested during trials on 5G corridors in different EU countries as well as in Turkey, China and Korea.

The trials will also allow 5G-MOBIX to conduct evaluations and impact assessments and to define business impacts and cost/benefit analysis. As a result of these evaluations and international consultations with the public and industry stakeholders, 5G-MOBIX will identify new business opportunities for the 5G enabled CCAM and propose recommendations and options for its deployment.

Through its findings on technical requirements and operational conditions 5G-MOBIX is expected to actively contribute to standardisation and spectrum allocation activities.

1.2. Purpose of the deliverable

This deliverable materializes a first version of the Data Management Plan elaborated of 5G-MOBIX. The purpose of this document is to provide an overview of the dataset types present in the project and to collect the main data management policy adopted by the Consortium.

The data management plan defines how data in general and research data in particular will be handled during the research project and will make suggestions for the after-project time. It describes what data will be collected, processed or generated, what methodologies and standards shall be followed during the collection process, whether and how this data shall be shared and/or made open not only for the evaluation needs but also to comply with the ORDP requirements, and how it shall be curated and preserved.

This deliverable will be updated accordingly to the evolution of the project. Three versions are planned:

- The initial version (**M06**, i.e. April 2019) outlines the data management plan according to the current development of the project. It identifies a first set of data categories, data types and metadata that will be involved in the project and proposes the data management process that will be followed in the next developments. This process includes also how the data owners will contribute to further versions of this deliverable in order to complete their dataset descriptions (definition and purpose, sharing, standard, preservation, etc.).
- The second version (**M18**, i.e. April 2020) will include data flows description, detailed documentation of DPIA and more detailed dataset descriptions according to the process defined in this version. A clear organisation of the different roles of the parties (i.e. data controllers, joint-controllers, data processors) will be implemented. Besides, the description of the categories of processed data and categories of personal data according to Article 30 of the GDPR will be provided.
- The final version (**M34**, i.e. August 2021) will outline the details of all datasets involved in 5G-MOBIX.

1.3. Intended audience

The dissemination level of D1.4 is 'public' (PU) and available to members of the consortium, the Commission (EC) services and those external to the project.

This document is primarily intended to serve as an internal guideline and reference for all 5G-MOBIX beneficiaries, especially the governance bodies such as the General Assembly, the Steering Committee, the Technical Management Team, and the Advisory Board.

2. DATA IN 5G-MOBIX: AN OVERVIEW

The aim of this chapter is to provide a first categorization of the processed data and to identify a list of the data types that will be generated. It also provides recommendations on data collection and sharing processes during the project and beyond.

5G-MOBIX will collect a huge amount of data to measure the benefit of the 5G technologies on the selected CCAM use cases along multiple corridors and trial sites. 5G-MOBIX will consider the FESTA¹ methodology, which provides an extensive set of recommendations for developing an experimental procedure for Field Operational Tests. The data will be defined during a close collaboration, between WP5 (Task 5.2), WP2 (Task 2.5) and WP3 (Task 3.5). These tasks will set the quality requirements for the data to be collected and to be used in the technical evaluation.

From raw data, a large amount of derived data will be produced to address multiple research needs. Derived data will follow a set of transformation: cleaning, verification, conversion, aggregation, summarization or reduction. In any case, data must be well documented in order to facilitate and foster sharing, to enable validity assessments and to enable its usage in an efficient way.

2.1. Datasets Categories

Different categories of datasets will be produced during the project:

- **Context data:** data that describe the context of an experiment on a 5G-MOBIX use case (see D2.1 – “5G-enabled CCAM use cases specifications”);
- **Acquired and derived data:** data that contain all the collected information related to an experiment on a 5G-MOBIX use case (see D2.1).

2.1.1. Context Data

Context data is any information that helps to explain observation during a study. Context data can be collected generated or retrieved from existing data and will provide the environment set up for testing vehicles, devices and users in a test scenario or a test session. The following draft list outlines the kind of information collected as contextual data:

- Type and number of vehicles;
- Type and number of users : Tester, Driver, Developer, Passenger, Pedestrian, Bystander;
- Details about: Road, Weather, Safety limitations, Number of safety interventions, etc.

¹ <http://fot-net.eu/Documents/festa-handbook-version-7>

2.1.2. Acquired and Derived Data

Acquired data is all data collected to be analysed during the trials. Derived data is created by different type of transformation including data fusion, filtering, classification, and reduction. Derived data are easy to use and contains derived measures and performance indicators referring to a time period when specific conditions are met. This category includes measurements from sensors mounted on vehicles, 5G infrastructures and services and subjective data collected from either the users or the environment. A first categorization of these data has been established in Table 1 for 5G-MOBIX and will be refined in the coming versions of this deliverable.

Table 1: Overview of the categories and types of data collected in 5G-MOBIX

Categories and types of data collected in 5G-MOBIX	
In-vehicle measurements are the collected data from vehicles, either using their original in-car sensors or sensors added for 5G-MOBIX purposes. These measures can be divided into different types:	
	<p>Vehicle attributes are measurements that describe the mobility of the vehicle. Measurements can be for example longitudinal speed, longitudinal and lateral acceleration, yaw rate, and slip angle.</p> <p>In-vehicle systems state can be accessed by connecting to the embedded controllers. It includes continuous measures like engine RPM or categorical values like ADAS and active safety systems activation.</p>
	<p>Sensors are the environment data that can be obtained by advanced sensors like RADARs, LIDARs, Cameras, simple optical sensors, etc. These sensors are described and presented in more details in the deliverable D2.4- "5G augmented vehicle specifications".</p> <p>Media: Mostly consist of high definition live video stream data shared with remote site in real-time but also index files used to synchronize the other data categories. They are often collected from the road side units.</p>
	Tele-operation and remote driver actions define the actions and data that enable a remote driver or a V2X application to operate a remote AD vehicle and take over control of the vehicle thanks to localisation, real-time high definition live video stream and 3D bird's eye view of vehicle sensors data.
	Vehicle positioning measurement of geographical location of a vehicle will be collected by different navigation system such as RTK GPS/IMU based positioning, DGPS, etc. See more details in the deliverable D2.4- "5G augmented vehicle specifications".
	<p>On-Board-Units : are measurements collected and stored in the OBU</p> <p>Radio Access Technologies: Consists of data such as the Link Quality Indicator for the RATs used in the OBU for example ITS G5, 3/4G, 5G, WIFI, LTE-V2X.</p> <p>V2X application: Data generated, stored and send through 3GPP V2X communication service (trajectories, sensor data, brake commands, manoeuvre, etc.).</p>
CCAM Infrastructure data are generated by the Cloud and Road Side infrastructures used in 5G-MOBIX and displayed in Figure 2. Some examples are listed below.	
	Edged Road Side Unit and Road-side measurements: are the vehicle's counting speed measurement and positioning, using radar, rangefinders, inductive loops or pressure hose. In 5G-MOBIX systems, it

	may also contain more complex information remotely transferred from vehicles to road-side units (Edged Dynamic Map (EDM), Path planning).
	Collective Perception Messages (CPMs) are sent by a Collective perception environment where vehicles and road side unit (RSU) exchange information in real time to enhance their perception of the environment.
	Sensors (will be further described and presented in the deliverable D2.3- "Specification of the infrastructure for 5G augmented CCAM")
5G infrastructure data are generated by the 5G infrastructure described in deliverable D2.2- "5G architecture and technologies for CCAM specifications".	
Some examples are listed below.	
	V2X scenarios evaluation KPI used to evaluate the uses case categories: <ul style="list-style-type: none"> • Advance driving (Payload (Bytes), Tx rate (Message/ Sec), Max end-to-end latency (ms), Reliability (%), Data rate (Mbps), Min required communication range (meters); • Vehicles Platooning (Payload (Bytes), Tx rate (Message/ Sec), Max end-to-end latency (ms), Reliability (%), Data rate (Mbps), Min required communication range (meters); • Extended Sensors (Payload (Bytes), Tx rate (Message/ Sec), Max end-to-end latency (ms), Reliability (%), Data rate (Mbps), Min required communication range (meters); • Remote Driving (Max end-to-end latency (ms), Reliability (%) Data rate (Mbps)).
	Other data (parameters, KPIs) will defined in the deliverable D2.5 for the evaluation tasks.
5G-MOBIX users' data are all the data generated by the different categories of users targeted by the project such as the Vulnerable Road Users (VRU) equipped with a smartphone.	
Experimental conditions are the external factors which may have an impact on participants' behaviour. They may be directly collected during the experiment, or integrated from external sources. Typical examples are traffic density and weather conditions.	

2.1.3. Metadata

The Data Sharing Framework defined by FOT-NET project² defines meta-data as '**any information that is necessary in order to use or properly interpret data**'. This section reviews the relevant metadata standards developed or used in the previous and ongoing field operational tests (FOT) and naturalistic driving studies (NDS) as a basis for the development of the metadata specifications of the pilot data. Such standards will help the analysis and re-use of the collected data within 5G-MOBIX to determine how the raw data was collected and processed in order to perform data analysis, modelling and interpretation.

These standards serve to establish the 5G-MOBIX's Metadata into four different categories as follows.

² <http://fot-net.eu/wp-content/uploads/sites/7/2017/04/FOT-Net-D3.1-Data-Sharing-Framework-v1.0.pdf>

- **5G-MOBIX pilot design and execution** documentation, which corresponds to a high-level description of a data collection: its initial objectives and how they were met, description of the test site, etc.
- **Descriptive** metadata, which describes precisely each component of the dataset, including information about its origin and quality.
- **Structural** metadata, which describes how the data is being organized.
- **Administrative** metadata, which sets the conditions for how the data can be accessed and how this is being implemented.

Full details of these metadata categories can be found in the Deliverables of the FOT-Net Data project such as D4.1³ Data Catalogue and D4.3⁴ Application of Data Sharing Framework in Selected-Cases which can be found at the project website⁵.

2.2. Data management methodology in 5G-MOBIX

The 5G-MOBIX data collection process and data management is built upon requirements or methodologies coming from several processes defined in WP2 – “Specifications”, WP3 – “Development, integration and roll out” and WP5 – “Evaluation” or guidance rules (ORPD, GDPR, ETHICS):

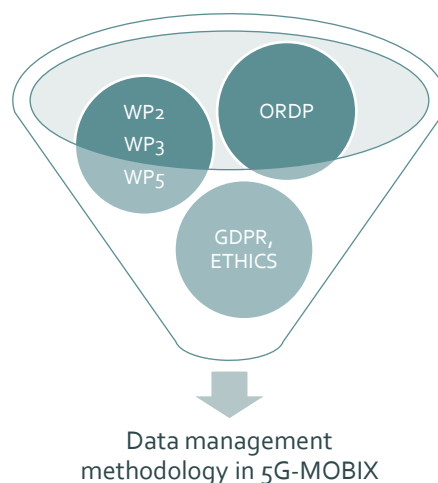


Figure 1: Overview of 5G-MOBIX data management

³ <http://fot-net.eu/wp-content/uploads/sites/7/2017/05/FOT-Net-Data-D4.1-Data-Catalogue-v3.pdf>

⁴ <http://fot-net.eu/wp-content/uploads/sites/7/2017/04/FOT-Net-D4.3-Application-of-Data-Sharing-Framework-in-Selected-Cases-v1.pdf>

⁵ <http://fot-net.eu/Documents/fot-net-data-final-deliverables/>

The evaluation requirement in WP5 – “Evaluation” defines the minimum data that must be collected in order to perform the evaluation process at the end of the project. Deliverable D5.1 – “Evaluation methodology and plan” will describe the evaluation methodology to be used. The evaluation data collection and management task defines the data collection, harmonization, and storage and sharing requirements. Deliverable D3.5 – “Report on the evaluation data management methodology and tools” will describe the evaluation data management methodology to be used. High level requirements defined in the evaluation process will be converted in Task 2.5 – “Initial evaluation KPIs and metrics” into specific and detailed specification of data format, data size, data currency, data units, data file, and storage.

The Open Research Data Pilot (ORDP) defines the requirement related to data sharing of research data. Requirements related to ORDP are defined in this document to guarantee that the collected data will be provided in compliance to European Commission Guidelines on Data Management in Horizon 2020. Those requirements are defined in Section 3 and Annex 9.1.

The processing of these data will comply with the different data protection regulations that apply on 5G-MOBIX such as the GDPR. Those regulations and the ethical aspects are presented in Chapter 6.

All these processes will be defined during the course of the project and must be rigorously followed.

2.3. Data Flow in 5G-MOBIX

Figure 2 below depicts the high-level architecture of CCAM infrastructure that will be instantiated by each use case.

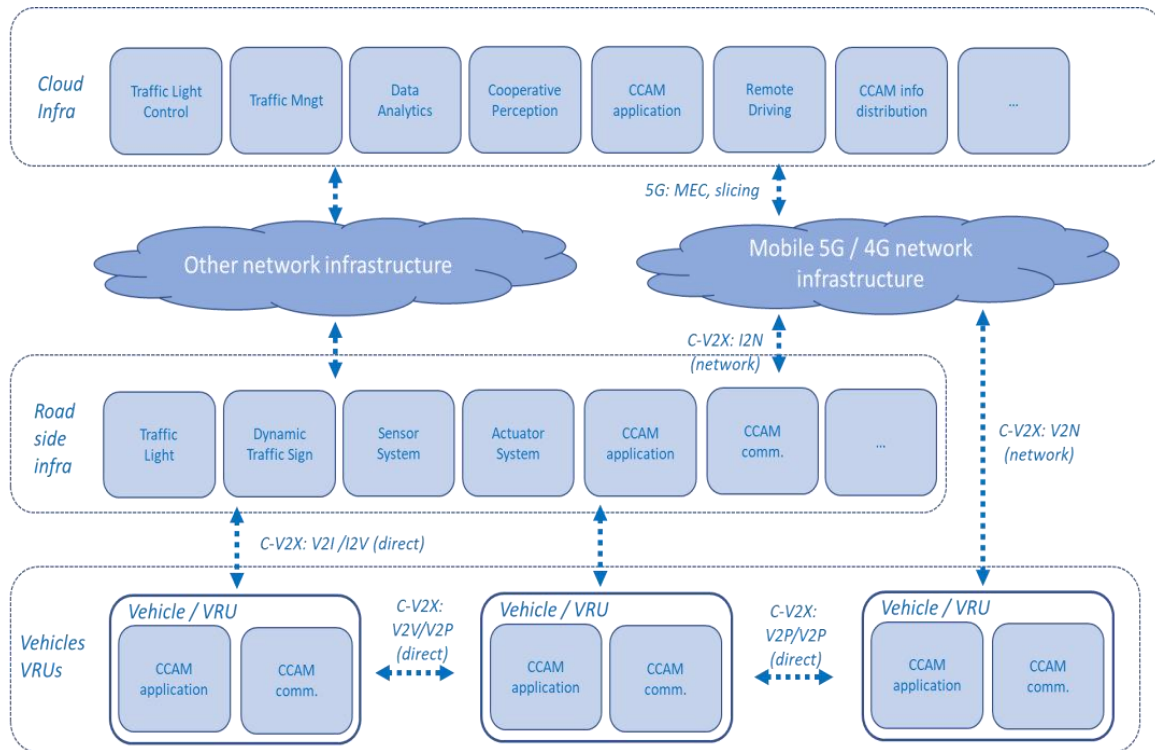


Figure 2: 5G-MOBIX CCAM Infrastructure

The dashed links show the bidirectional flow that link 5G-MOBIX sub/systems. Only part of this data flow will be collected and stored during the trial activities and evaluated by the evaluation team.

Evaluation data will be collected at each trial site. A process of harmonization, quality checking and data formats will be done at the level of each trial site before transferring the data to the centralized server.

In the next version of this DMP a detailed data flows diagram in 5G-MOBIX will be created. They will help to understand the information lifecycle and represent all in- and cross application flows. Besides, for each identified data processing operation the adopted solution to anonymise the different categories of data will be defined. This second version will describe also cross-border data flows towards non-EU countries and carried out in a manner compliant with the GDPR requirements.

2.4. 5G-MOBIX shared dataset description

2.4.1. General data description

This section provides the description guidelines of the different types of datasets to be collected and shared by 5G-MOBIX after the end of the project with respect to ORDP which aims to improve and maximize access to and re-use of research data generated by Horizon 2020 projects.

As the nature and extent of these datasets can evolve during the project, more detailed descriptions will be provided in the next version of the DMP (M18).

The descriptions of the different datasets, including their reference, file format, standards, methodologies and metadata and repository to be used are given below.

2.4.2. Template used for 5G-MOBIX dataset description

This table is a template that shall be used to describe the datasets.

Table 2: 5G-MOBIX Dataset description template

Dataset Reference	Each dataset will have a reference that will be generated by the combination of the name of the project, the trial site, the use case in which it is generated and the datatype: "5G-MOBIX_Trial-Site_UC_Datatype".
Dataset Name	Name of the dataset.
Dataset Description	Each dataset will have a full data description explaining the data provenance, origin and usefulness. Reference may be made to existing data that could be reused.
Standards and metadata	The metadata attributes list to be used to find the dataset.
File format	All the format that defines data.
Data Sharing	<p>Explanation of the sharing policies related to the dataset between the next options:</p> <p>Open: Open for public disposal.</p> <p>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.</p> <p>Restricted: Only for project internal use.</p> <p>Each dataset must have its distribution license.</p> <p>Provide information about personal data and mention if the data is anonymized or not. Tell if the dataset entails personal data and how this issue is taken into account.</p>
Archiving and Preservation	The preservation guarantee and the data storage during and after the project (for example databases, institutional repositories, public repositories, etc.).

3. PARTICIPATION IN THE OPEN RESEARCH DATA PILOT

5G-MOBIX has agreed to participate in the Pilot on Open Research Data in Horizon 2020 and uses the specific Horizon 2020 guidelines associated with 'open' access to ensure that the results of the project provide the greatest impact possible.

5G-MOBIX will ensure the open access⁶ to all peer-reviewed scientific publications relating to its results and will provide access to the research data needed to validate the results presented in deposited scientific publications.

The following lists the minimum fields of metadata that should come with an 5G-MOBIX project-generated scientific publication in a repository:

- The terms: "European Union (EU)", "Horizon 2020"
- Name of the action (Research and Innovation Action)
- Acronym and grant number (5G-MOBIX, 825496)
- Publication date
- Length of embargo period if applicable
- Persistent identifier

When referencing Open access data, 5G-MOBIX will include at a minimum the following statement demonstrating EU support (with relevant information included into the repository metadata):

"This project has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement No 825496".

The 5G-MOBIX consortium will strive to make many of the collected datasets open access. When this is not the case, the data sharing section for that particular dataset will describe why access has been restricted (See. the row Data Sharing in Table 2).

In regards to the specific repositories available to the 5G-MOBIX consortium, numerous project partners maintain institutional repositories that will be listed in the following DMP version, where project scientific

⁶ http://ec.europa.eu/research/participants/docs/h2020-funding-guide/cross-cutting-issues/open-access-data-management/open-access_en.htm

publications and in some instances, research data will be deposited. The use of a specific repository will depend primarily on the primary creator of the publication and on the data in question.

Some other project partners will not operate publically accessible institutional repositories. When depositing scientific publications, they shall use either a domain specific repository or use the EU recommended service OpenAIRE (<http://www.openaire.eu/>) as an initial step to finding resources to determine relevant repositories.

Project research data shall be deposited to the online data repository ZENODO⁷. It is a free service developed by CERN under the EU FP7 project OpenAIRE plus (grant agreement no.283595).

The repository shall also include information regarding the software, tools and instruments that were used by the dataset creator(s) so that secondary data users can access and then validate the results.

The 5G-MOBIX data collections will be accessed in ZENODO repository in a similar address as the following link: <https://zenodo.org/collection/<<5g-mobix>>>

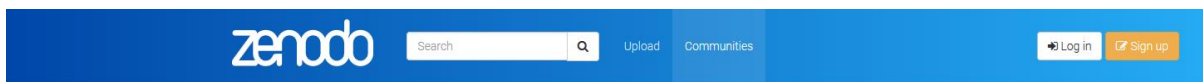


Figure 3: Illustration of the access to 5G-MOBIX repository in ZENODO

In summary, as a baseline 5G-MOBIX partners shall deposit:

- Scientific publications – on their respective institute repositories in addition (when relevant) to the 5G-MOBIX ZENODO repository;
- Research data – to the 5G-MOBIX ZENODO collection (when possible);
- Other project output files – to the 5G-MOBIX ZENODO collection (when relevant).

This version of the DMP does not include the actual metadata about the Research Data being produced in 5G-MOBIX project. Details about technical means and services for building repositories and accessing to this metadata will be provided in the next version of the DMP. A template document is defined in Table 2 and will be used by project partners to provide all requested information.

⁷ <https://zenodo.org/>

4. FAIR DATA MANAGEMENT PRINCIPLES

The data that will be generated during and after the project should be FAIR⁸, that is Findable, Accessible, Interoperable and Reusable. These requirements do not affect implementation choices and don't necessarily suggest any specific technology, standard, or implementation solution.

The FAIR principles were generated to improve the practices for data management and data-curation, and FAIR aims to describe the principles in order to be applied to a wide range of data management purposes, whether it is data collection or data management of larger research projects regardless of scientific disciplines.

With the endorsement of the FAIR principles by H2020 and their implementation in the guidelines for H2020, The FAIR principles serve as a template for lifecycle data management and ensure that the most important components for lifecycle are covered.

This is intended as an implementation of the FAIR concept rather than a strict technical implementation of the FAIR principles. 5G-MOBIX will implement several actions described below to carry on the FAIR principles.

Making data findable, including provisions for metadata:

- The datasets will have very rich metadata to facilitate the findability. Open data format (csv, xml) will be used.
- All the datasets will have a Digital Object Identifiers provided by the public repository (ZENODO).
- The reference used for the dataset will follow a format like the following: "5G-MOBIX_Trial-Site_UC_Datatype_XX" (XX: identifier to be added for similar datasets).
- The standards for metadata will be defined for each dataset as described in Section 2.1.3.

Making data openly accessible:

- The datasets that will be openly available will be described according to Table 2.
- The datasets for evaluation will be accessible via 5G-MOBIX centralized server.
- The datasets will be made available using public repository (e.g. ZENODO) after the project.
- Table 2 will be used to explain the methods or software used to access the data. Basically, no software is needed to access the data
- The data and their associated metadata will be deposited in a public repository or either in an institutional repository.

⁸ http://ec.europa.eu/research/participants/docs/h2020-funding-guide/cross-cutting-issues/open-access-data-management/data-management_en.htm

- The row Data Sharing in the Table 2 will outline the rules to access the data if restrictions exist.

Making data interoperable:

- The metadata vocabularies, standards and methodologies will depend on the public repository and use the recommendations of section 2.1.3.
- The 5G-Mobix will define common data formats. This work was developed in Task 2.5 – “KPI and evaluation data” & Task 5.1 “Evaluation methodology and plans”. The goal is to have the same structure across trial sites and enable evaluators dealing with the same format for all pilot sites.

Increase data re-use (through clarifying licenses):

- All the data producers will license their data to allow the widest reuse possible. More details about license types and rules will be provided in the next version.
- By default, the data will be made available for reuse. If any constraints exist, an embargo period will be mentioned in the row Data Sharing of Table 2 to keep the data for only a period of time.
- The data producers will make their data for third-parties within public repositories. They will be reused for the scientific publications validation purpose.

5. ALLOCATION OF RESOURCES

The costs to make the data FAIR in 5G-MOBIX shall be handled by each partner who will have to generate its data according to WP5 – “Evaluation” and Task 2.5 – “KPI and evaluation data specification” requirements.

In the project, Sadeq ZOUGARI (AKKA) plays the role of Data Manager and liaise with the TMT about the data management issues. The Data Manager leads data management plan tasks and participates in the project coordination in terms of the evaluation data collection, storage and handling, as well as their publication as part of the ORDIP.

All research data collected as part of this project is owned by the data producer or partners involved in trial sites. The partners in 5G-MOBIX will take the responsibility for the collection, management, and sharing of the research data. Quality assessment will be the responsibility of data manager of each trial site.

6. DATA PROTECTION REGULATIONS AND ETHICAL ASPECTS

6.1. An approach to comply with GDPR in 5G-MOBIX

This section uses the terminology described in Annex 9.2 about General Data Protection Regulation (GDPR)⁹ and provides an initial application of those terminologies to 5G-MOBIX project. The approach proposed to handle the GDPR in the 5G-MOBIX project consists of identifying all the concerned parties and identify the actions they need to take in order to comply with the regulation. More specifically, this approach consists of six steps:

1. The first step consists of appointing a Data Protection Officer (DPO) and defining his mission in 5G-MOBIX. According to the grant agreement (5G-MOBIX, 825496), François Fischer acts as *de facto* DPO and the responsibilities of the DPO are described in the Ethics chapter 5 / 5.1 of the grant agreement.
2. During this step, a complete cartography of the personal data processing will be made. A template has been provided in Annex 9.1 which will be exploited to establish such cartography.
3. Based on the results of the previous step, priorities need to be set about the personal data processing which may be riskier.
4. This step consists of conducting a Data Privacy Impact Assessment to evaluate the risks of the processing identified during the previous step.
5. During this step, the procedures are implemented for 5G-MOBIX in order to provide the protection of the personal data according to the GDPR.
6. This final step concerns the provision of a clear documentation of the whole data protection procedure that has been implemented.

Thus, this approach will enforce that personal data collected in 5G-MOBIX shall be (Article 5):

- 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,

⁹ Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016

- 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 including for example a template to be filled by each data processor.

Since the project declare that the research involves personal data collection and/or processing, each data processor under the control of the data controller and the data protection officer will declare how GDPR rules are fulfilled in a separate document.

6.2. Data protection regulation in the Non-EU Countries

Article 3 of the GDPR states that:

- This Regulation applies to the processing of personal data in the context of the activities of an establishment of a controller or a processor in the Union, regardless of whether the processing takes place in the Union or not
- This Regulation applies to the processing of personal data of data subjects who are in the Union by a controller or processor not established in the Union, where the processing activities are related to:
 - The offering of goods or services, irrespective of whether a payment of the data subject is required, to such data subjects in the Union; or
 - The monitoring of their behaviour as far as their behaviour takes place within the Union.
- This Regulation applies to the processing of personal data by a controller not established in the Union, but in a place where Member State law applies by virtue of public international law.

Therefore, 5G-MOBIX, for the trial sites in the three Non-EU countries (China, Korea and Turkey) a clarification will be made about the interactions between the GDPR and the data protection regulations in those countries.

6.3. Ethical Aspects

Ethical issues related to the protection of personal data (POPD) in 5G-MOBIX will be addressed in D8.2 - *POPD – Requirement No. 2.*

7. DATA SECURITY

The data produced during 5G-MOBIX will be stored per trial site in local servers and in central servers for the whole project. According to the previous chapter, those data are made compliant with the GDPR. This chapter describes some security principles that are implemented in order to protect against any type of modification. Also, a more thorough management of the servers could be made using the ISO 27001 standards or the BSI-Standards¹⁰. The security principles are listed below:

- **Authentication:** All the users wanting to get access to the 5G-MOBIX data servers should be authenticated. Also, proper means are used to authenticate the servers. An authentication system could be used to handle the authentication of the users during the course of the project.
- **Authorization:** The access to 5G-MOBIX data servers are 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:** In 5G-MOBIX 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, auditing. Other accounting mechanisms shall be implemented.
- **Confidentiality:** the data stored in 5G-MOBIX servers should be encrypted during transmission and storage.
- **Communication Security:** Access to 5G-MOBIX servers should be done through encrypted communication channels such as HTTPS and IPsec.
- **Data Integrity:** The data collected during 5G-MOBIX should be protected from malicious and accidental modifications by any users during their transmission or their storage. Cryptographic mechanisms such as hash functions and digital signatures shall be used.
- **Availability:** This security principle assures that the 5G-MOBIX servers should be available for 5G-MOBIX users during the defined interval of service. Also, regular backups of the data should be made. Therefore, mechanisms to cope with the charge and DoS attacks should be implemented.

¹⁰ https://www.bsi.bund.de/EN/Publications/BSIStandards/BSIStandards_node.html

8. CONCLUSION

This deliverable (and its scheduled updates) provides an overview of the data that 5G-MOBIX project will produce together with related data processes and requirements that need to be taken into consideration.

Since 5G-MOBIX participates in the Pilot on Open Research Data in Horizon 2020, specific Horizon 2020 guidelines associated with 'open' access are described to ensure that the results of the project results provide the greatest impact possible.

In Chapter 5, the DMP provides a table template for the description of the datasets to be shared after the project. Chapter 5 will be incrementally enriched along the project lifetime. These descriptions include a detailed description, standards, methodologies, sharing and storage methods.

This document highlights the fact that 5G-MOBIX will ensure the open access to all peer-reviewed scientific publications relating to its results and will provide access to the research data needed to validate the results presented in deposited scientific publications.

The document describes the FAIR data management principle and defines several actions to carry on the FAIR principles.

Finally, Data Management Plan is a living document that will be updated from its creation to the end of the 5G-MOBIX project.

9. ANNEXES

9.1. Annex 1 – Open research data pilot (ORD Pilot)

Open access refers to the online provision of scientific information that is free of charge to the end-user and reusable. This scientific information handles the peer-reviewed scientific research articles/publications and the research data underlying publications.

Under the H2020, the project must also aim to deposit the research data needed to validate the results presented in the deposited scientific publications, known as “underlying data”. In order to effectively supply this data, projects need to consider at an early stage how they are going to manage and share the data they create or generate under H2020 guidelines on data management and with respect of 5G-MOBIX grant agreement.

“The Commission is running a flexible pilot under Horizon 2020 called the Open Research Data Pilot (ORD pilot). The ORD pilot aims to improve and maximize access to and re-use of research data generated by Horizon 2020 projects and takes into account the need to balance openness and protection of scientific information, commercialization and Intellectual Property Rights (IPR), privacy concerns, security as well as data management and preservation questions.”

“By extending the pilot, open access becomes the default setting for research data generated in Horizon 2020.

However, not all data can be open. Projects can therefore opt out at any stage (either before or after signing the grant) and so free themselves retroactively from the obligations associated with the conditions – if:

- participation is incompatible with the obligation to protect results that can reasonably be expected to be commercially or industrially exploited
- participation is incompatible with the need for confidentiality in connection with security issues
- participation is incompatible with rules on protecting personal data
- participation would mean that the project's main aim might not be achieved
- The project will not generate / collect any research data or
- There are other legitimate reasons (you can enter these in a free-text box at the proposal stage).”

After depositing publications beneficiaries must ensure open access to those publications via the chosen repository.

“The two main routes to open access are:

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

In the research context, examples of data include statistics, results of experiments, measurements, observations resulting from fieldwork, survey results, interview recordings and images. The focus is on research data that is available in digital format and stored in a public repository. Normally, users can access, mine, exploit, reproduce, and disseminate openly accessible research data free of charge as explained in the following Figure.

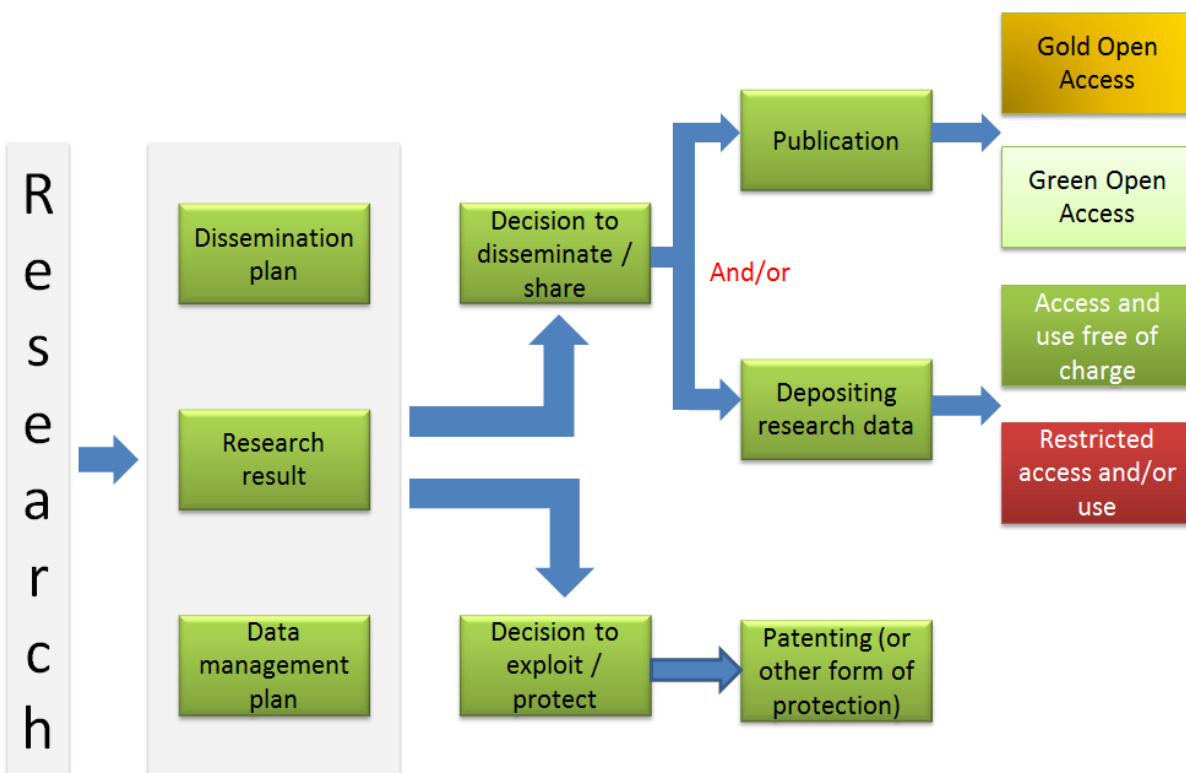


Figure 5: Principles of H2020 open access to research data

9.2. Annex 2 – Initial GDPR application in 5G-MOBIX

9.2.1. Terminology

The following definitions are drawn from the Article 4 of the GDPR.

Personal data means any information related to an identified or identifiable natural person ('data subject'); an identifiable natural person is one who can be identified, directly or indirectly using identifiers, physical, physiological data. An initial identification of such data in 5G-MOBIX is presented in Table 1. In the further versions of this deliverable, the deliverable D2.1 will be used to identify the data subject in 5G-MOBIX.

Processing means any operation or set of operations which is performed on personal data or on sets of personal data. The Article 4 gives an extensive set of such processing. Furthermore, data processing in 5G-MOBIX will exploit the data described in Table 1 and using the KPIs defined in deliverables D2.1 – "5G-enabled CCAM use cases specifications" and D2.5 – "Initial evaluation KPIs and metrics" in order to evaluate the outcomes of the project.

Controller determines the purposes and means of the processing of personal data. An identification of the controllers in 5G-MOBIX will be carried out for the next versions of this deliverable.

Processor means a natural or legal person, public authority, agency or other body which processes personal data on behalf of the controller;

9.2.2. Data subject's rights

According to Article 7, a lawful basis for processing of personal data exists when the data subject has given clear consent to process his/her personal data for a specific purpose. A competent board 5G-MOBIX in order to provide consent.

In 5G-MOBIX mechanisms will be implemented to provide the following rights to the data subject which are described in more details in the chapter III of the GDPR.

9.2.3. Controller's and Processor's obligations

In 5G-MOBIX the identified data controllers and data processor should follow the requirements in the chapter IV of the GDPR.

9.2.4. GDPR templates used for 5G-MOBIX data processing

NOTE: The information requested here is in line with the requirement to maintain data processing records under the GDPR and is specific to personal data. This template comes from previous work (EU/Autopilot project). All data controllers must also keep records of dataset descriptions according to the Data Management Plan

9.2.4.1. Data controller's record of processing activities

1	Contact details of Data Controller (and, if applicable, joint controller, the controller's representative and the data protection officer)
	<i>Name</i>
	<i>Email</i>
	<i>Company address</i>
	<i>Telephone</i>
2	Purpose of processing
3	Description of categories of data subjects and of the categories of personal data
4	Categories of recipients to whom the personal data have been or will be disclosed including recipients in third countries or international organisations
5	Where applicable, transfers of personal data to a third country or an international organisation, including the identification of that third country or international organisation
6	Where possible, the envisaged time limits for erasure of the different categories of data
7	Where possible, a general description of the technical and organisational security measures for
	a <i>the pseudonymisation and encryption of personal data —</i>
	b <i>the ability to ensure the ongoing confidentiality, integrity, availability and resilience of processing systems and services —</i>
	c <i>the ability to restore the availability and access to personal data in a timely manner in the event of a physical or technical incident —</i>
	d <i>a process for regularly testing, assessing and evaluating the effectiveness of technical and organisational measures for ensuring the security of the processing —</i>

9.2.4.2. Data processor's record of processing activities

1	Contact details of Data Processor (and controller on behalf of which the processor is acting, and, where applicable, of the controller's or the processor's representative, and the data protection officer)
	<i>Data Processor(s)</i>
	<i>Email</i>
	<i>Company address</i>
	<i>Telephone</i>
	<i>Data Controller</i>
	<i>Email</i>
	<i>Company address</i>
	<i>Telephone</i>
2	Categories of processing carried out on behalf of the Controller
3	Where applicable, transfers of personal data to a third country or an international organisation, including the identification of that third country or international organisation
4	Where possible, a general description of the technical and organisational security measures for
a	<i>the pseudonymisation and encryption of personal data —</i>
b	<i>the ability to ensure the ongoing confidentiality, integrity, availability and resilience of processing systems and services —</i>
c	<i>the ability to restore the availability and access to personal data in a timely manner in the event of a physical or technical incident —</i>
d	<i>a process for regularly testing, assessing and evaluating the effectiveness of technical and organisational measures for ensuring the security of the processing —</i>