Grant Agreement Number: 731993

Project acronym: AUTOPILOT

Project full title: AUTOmated driving Progressed by Internet Of Things

D. 6.7

DATA MANAGEMENT PLAN

Due delivery date: 30/06/2017

Actual delivery date: 30/06/2017

Organization name of lead participant for this deliverable: AKKA

Dissemination level		
PU	Public	X
PP	Restricted to other programme participants (including the GSA)	
RE	Restricted to a group specified by the consortium (including the GSA)	
CO	Confidential , only for members of the consortium (including the GSA)	

Project funded by the European Union’s Horizon 2020 Research and Innovation Programme (2014 – 2020)

Document Control Sheet

Deliverable number:	D6.7
Deliverable responsible:	AKKA
Workpackage:	6
Editor:	ZOUGARI, Sadeq

Author(s) – in alphabetical order		
Name	**Organisation**	**E-mail**
AIT-AAZIZI Amine	AKKA	Amine.ait-aazizi@akka.eu
FISCHER François	ERTICO	f.fischer@mail.ertico.com
ZOUGARI Sadeq	AKKA	Sadeq.zougari@akka.eu

Document Revision History			
Version	**Date**	**Modifications Introduced**	
		Modification Reason	**Modified by**
V0.1	15/05/2017	First Draft	ZOUGARI, Sadeq
V0.6	09/06/2017	Draft ready for peer review	ZOUGARI, Sadeq
V1.0	June 2017	Peer review	ZOUGARI, Sadeq
V1.1	16/06/2017	Final version submitted to EC	F.Fischer
V1.2	28/06/2017	2nd version for submission	F.Fischer

Abstract
This document identifies and describes all data that will be collected and generated during the AUTOPILOT project and define the data management principles that shall be followed by the Consortium. This includes the elicitation of the different data categories, how data will be created, stored and backed-up, who owns it and who is responsible for the different data, which data will be preserved and shared. This version of the document is the first step. It will be amended regularly and incrementally all along the project lifecycle by taking into account the progress of the work and will ensure the sharing and dissemination of relevant information to all project partners with regards to data management. Two additional versions will document the progress of the work.

Abbreviations and Acronyms

Acronym	Definition
ADAS	Advanced Driver-Assistance Systems
AVP	Automated Valet Parking
CAN	Controller Area Network
DMP	Data Management Plan
EC	European Commission
EC	European Commission
FAIR	Findable – Accessible – Interoperable-Reusable
FESTA	Field opErational teSt supporT Action
GA	Grant Agreement
GPS	Global Positioning System
IOT	Internet Of Things
ITS	Intelligent Transportation Systems
LIDAR	Light Detection And Ranging
ORDP	Open Research Data Pilot
PO	Project officer
PO	Project officer
RADAR	Radio Detection And Ranging
RPM	Revolutions Per Minute
WP	Work Package

1 Table of Contents

List of Figures

List of Tables

Executive Summary

In Horizon 2020 a limited pilot action on open access to research data has been implemented. Participating projects have been required to develop a Data Management Plan (DMP).

This deliverable provides the first version of the DMP elaborated by the AUTOPILOT project. The purpose of this document is to provide an overview of the main elements of the data management policy. It outlines how research data will be handled during the AUTOPILOT project and describes what data will be collected, processed or generated and following what methodology and standards, whether and how this data will be shared and/or made open, and how it will be curated and preserved. Besides, an initial list of data types, metadata and global data collection processes are also provided in this document.

The AUTOPILOT Data Management Plan refers to the latest EC DMP guidelines. This version has explicit recommendations for full life cycle management through the implementation of the FAIR principles, which state that the data produced shall be Findable, Accessible, Interoperable and Reusable (FAIR).

Since the data management plan is expected to evolve during the project while taking into account the progress of the work, forthcoming versions will be produced as additional deliverables at later stages of the project.

1. Introduction

1.1 Objectives of the Project

Automated driving is expected to increase safety, provide more comfort and create many new business opportunities for mobility services. The market size is expected to grow gradually reaching 50% of the market in 2035. The Internet of Things (IoT) is about enabling connections between objects or "things"; it is about connecting anything, anytime, anyplace, using any service over any network.

AUTOmated driving **P**rogressed by **I**nternet **O**f **T**hings" (AUTOPILOT) project will especially focus on utilizing the IoT potential for automated driving.

The overall objective of AUTOPILOT is to bring together relevant knowledge and technology from the automotive and the IoT value chains in order to develop IoT-architectures and platforms, which will bring Automated Driving towards a new dimension. This will be realized through the following main objectives:

- Use, adapt and innovate current and advanced technologies to define and implement an IoT approach for autonomous and connected vehicles
- Deploy, test and demonstrate IoT based automated driving use cases at several permanent pilot sites, in real traffic situations with: Urban driving, Highway pilot, Automated Valet Parking, Platooning
- Create and deploy new business products and services for fully automated driving vehicles, used at the pilot sites: by combining stakeholders' skills and solutions, from the supply and demand side
- Evaluate with the involvement of users, public services and business players at the pilot sites:
    - The suitability of the AUTOPILOT business products and services as well as the ability to create new business opportunities
    - The user acceptance related to using the Internet of Things for highly or fully automated driving
    - The impact on the citizens' quality of life
- Contribute actively to standardization activities as well as to consensus building in the areas of Internet of Things and communication technologies.

Automated vehicles largely rely on on-board sensors (LiDAR, radar, cameras, etc.) to detect the environment and make reliable decisions. However, the possibility of interconnecting surrounding sensors (cameras, traffic light radars, road sensors, etc.) exchanging reliably redundant data may lead to new ways to design automated vehicle systems potentially reducing cost and adding detection robustness.

Indeed, many types of connected objects may act as an additional source of data, which will very likely contribute to improve the efficiency of the automated driving functions, enable new automated driving scenarios as well as increase the automated driving function safety while providing driving data redundancy and reducing implementation costs. These benefits will enable pushing the SAE level of driving automation to the full automation, keeping the driver out of the loop. Furthermore, by making autonomous cars a full entity in the IoT, the AUTOPILOT project enables developers to create IoT/AD services as easy as accessing any entity in the IoT.

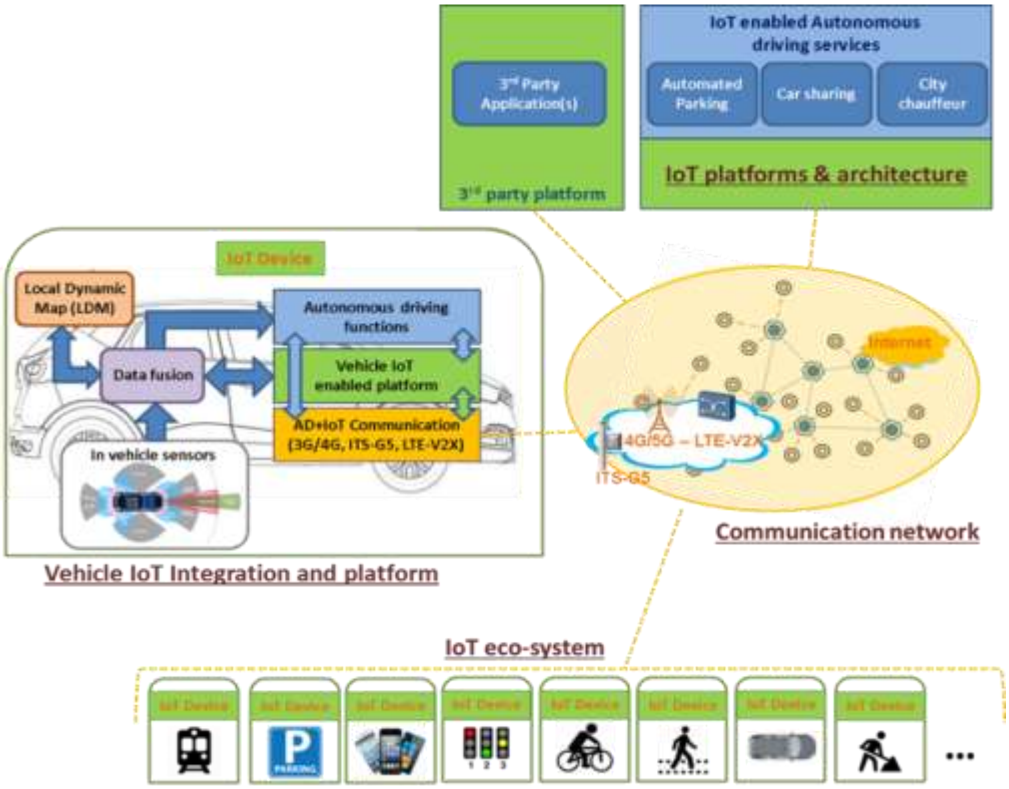


Figure 1 – The AUTOPILOT overall concept

The Figure above depicts the AUTOPILOT overall concept including the different ingredients to apply IoT to autonomous driving:

- The overall IoT platforms and architecture, allowing the use of the IoT capabilities for autonomous driving.
- The Vehicle IoT integration and platform to make the vehicle an IoT device, using and contributing to the IoT.
- The Automated Driving relevant sources of information (pedestrians, traffic lights ...) becoming IoT devices and extending the IoT eco-systems to allow enhanced perception of the driving environment on the vehicle.
- The communication network using appropriate and advanced connectivity technology for the vehicle as well as for the other IoT devices.

1.2 Purpose of the document

This deliverable materializes a first version of the Data Management Plan elaborated for the AUTOPILOT project. The purpose of this document is to provide an overview of the dataset types that will be dealt with in the project and to define the main data management policy adopted by the Consortium.

The data management plan defines how data, in general, and research data, in particular, should be handled during the research project, but also beyond the project lifespan. It describes what data will be collected, processed or generated by the IOT devices and by all the IoT ecosystem, what methodologies and standards shall be followed during the collection process, whether and how these data shall be shared and/or made open not only for evaluation purposes but also to comply with the ORDP requirements, and how it shall be curated and preserved. Besides, the Data management plan identifies the four key process requirements that define the data collection process, and provides first recommendations to be applied.

The AUTOPILOT Data Management Plan will be updated throughout the project lifespan. Three versions are planned:

This **M06 version** (D6.7), which is the initial version, 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 also includes 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 …).
The **M16 version** of the data management plan will include more detailed dataset descriptions according to the process defined in this version. M16 corresponds to the end of WP1 & WP2 and supposes that all the specifications related to the IOT platforms and AD services will be comprehensively defined. The descriptions will be filled following the template provided in section 4.2.
The **M32 version** will outline the details of all datasets involved in the AUTOPILOT project. These datasets include acquired or derived data and aggregated data (IOT data, evaluation data, test data and research data). These dataset types are explained in detail in section 4.3.

Chapter 2 provides an overview of the AUTOPILOT data, with their categories, types and metadata, and the data collection processes to be followed. It also specifies the test data flow and test data architecture environment.
Chapter 3 gives insights into the Open Research Data Pilot under H2020 guidelines.
Chapter 4 provides a template that will be used to describe datasets in the future (M16 and M32 versions).
Chapter 5 gives insights into the FAIR Data Management principle under H2020 guidelines.
Finally, the remaining chapters outline the necessary roles, responsibilities and ethical issues.

1.3 Intended audience

The AUTOPILOT project addresses highly innovative concepts. As such, foreseen intended audience of the project is the scientific community interested in the IOT and/or automotive technologies. In addition, due to the strong expected impact of the project on their respective domains, the other expected audience consists in automotive industrial communities, telecom operator and standardization organizations.

2. Data in AUTOPILOT: an Overview

The aim of this chapter is:

- To provide a first categorization of the data,
- To identify a list of the data types that will be generated,
- To provide a list of metadata that will be used to describe to describe the generated data and enable their reuse
- To provide recommendations on data collection and sharing processes during the project and beyond.

The AUTOPILOT project will collect a large amount of raw data to measure the benefit of IoT for automated driving with multiple automated driving use cases and services, at different pilot locations.

Data from vehicles and sensors will be collected and managed through a hierarchy of IoT platforms as illustrated in the architectural diagram of Figure 2.

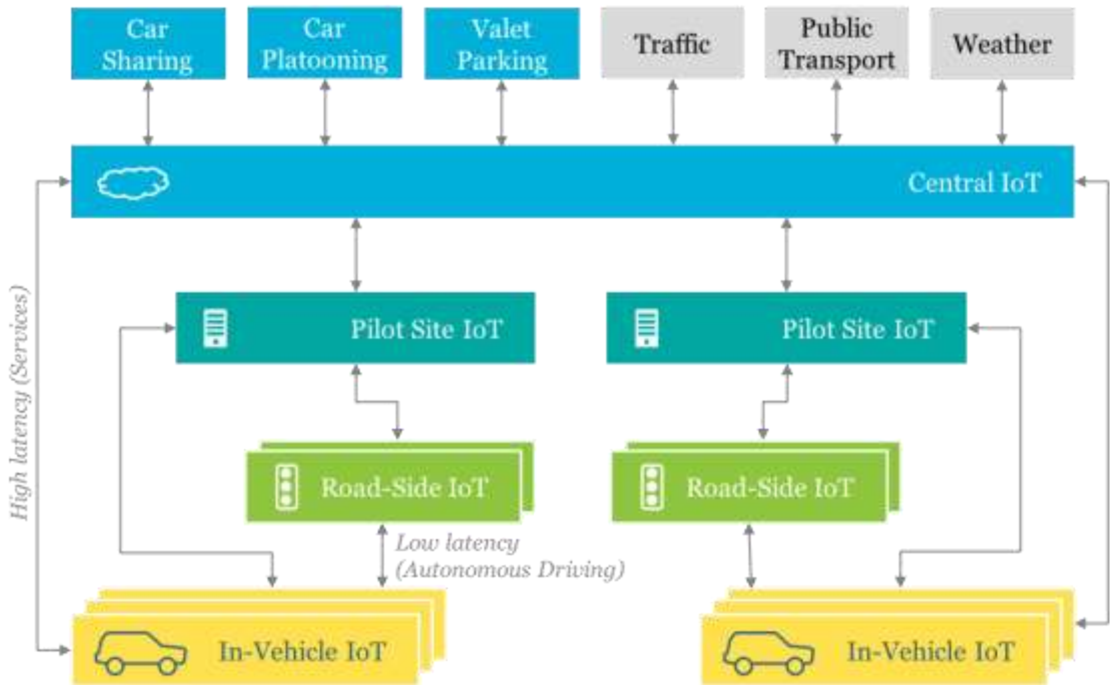


Figure 2 – The AUTOPILOT Hierarchical IoT Architecture

The diagram above shows a federated architecture with the following four layers:

- **In-Car IoT Platforms:** Here is everything that is mounted inside the vehicle, i.e., components responsible for AD, positioning, navigation, real time sensor data analysis, and communication with the outside world. All mission critical autonomous driving functions should typically reside in this layer.
- **Road-Side IoT Platforms:** Road-side and infrastructure devices, such as cameras, traffic light sensors, etc., are integrated and managed as part of road-side IoT PFs covering different road segments and using local low latency communication networks and protocols as required by the devices and their usage.
- **Pilot Site IoT Platforms:** This layer constitutes the first integration level. It is responsible for collecting, processing and managing data at the pilot sites level.

- **Central IoT Platform:** This is a Cloud-based top layer that integrates and aggregates data from the various pilot sites as well as external services (weather, transport, etc.). This is where the common AD services such as car sharing, platooning, etc., will reside. Data, at this level, are standardized using common formats, structures and semantics.

Data will be analyzed according to Field Operational Test studies (FOT[1]) and using FESTA[2] methodology.
From raw data, a large amount of derived data will be produced to address multiple research needs. Derived data will follow a set of transformations: cleansing, 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.
Thus, each dataset must be described using additional information called metadata. The latter must provide information about the data source, the data transformations and the conditions in which the data have been produced.

2.1 Dataset Categories

The AUTOPILOT project will produce different categories of datasets.

- **Context data**: data that describe the context of an experiment.
- **Acquired and derived data**: data that contain all the collected information related to an experiment.
- **Aggregated data**: data summary obtained by reduction of acquired data and generally used for data analysis.

1.1.1 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. For example, it contains information such as vehicle, road or drivers characteristics.

1.1.2 Acquired and Derived Data

Acquired data are all data collected to be analyzed during the course of the study. Derived data are created by different types of transformations including data fusion, filtering, classification, and reduction. Derived data are easy to use, they contain derived measures and performance indicators referring to a time period when specific conditions are met. This category includes measures from sensors coming from vehicles or IOT and subjective data collected from either the users or the environment.

The following list outlines the data types and sources that will be collected:

[1] http://fot-net.eu/
[2] http://wiki.fot-net.eu/index.php/FESTA_Handbook

In-vehicle measures are the collected data from vehicles, either using their original in-car sensors or sensors added for AUTOPILOT purposes. These measures can be divided into different types:	
	Vehicle dynamics 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.
	Driver actions represent the driver actions on the vehicle commands that can be measured; for instance, steering wheel angle, pedal activation or HMI button press variables, face monitoring indicators characterizing the state of the driver, either physical or emotional.
	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.
	Environment detection is the environment data that can be obtained by advanced sensors like RADARs, LIDARs, cameras and computer vision, or more simple optical sensors. For instance, luminosity or presence of rain, but also characteristics and dynamics of the infrastructure (lane width, road curvature) and surrounding objects (type, relative distances and speeds) can be measured from within a vehicle.
	Vehicle positioning the geographical location of a vehicle is determined with satellite navigation systems (e.g. GPS) and the aforementioned advanced sensors.
	Media mostly consist of video. The data consist of media data but also index files used to synchronize the other data categories. They are also often collected from the road side.
Continuous subjective measures Complimentary to sensors and instrumentation, some continuous measures can also be built in a more subjective way, by analysts or annotators, notably using video data.	
Road-side measures are the vehicle's counting speed measurement and positioning, using radar, rangefinders, inductive loops or pressure hose. In ITS systems, it may also contain more complex information remotely transferred from vehicles to road-side units.	
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.	
IOT data are the external sources of data that will be collected from external connected devices:	
	Users Data can be generated by smartphones or wearables. Users may be pedestrians or car drivers. These data help pedestrian detection by vehicle or infrastructure.
	Infrastructure Data are all the data giving additional information about the environment. Typical examples are the traffic status, road works, accidents and road conditions. They can also be directly collected from Road-side cameras or traffic light control units and then transferred to IOT Platforms. For instance, the infrastructure data can transfer hazard warnings or expected occupancy of busses on bus lanes to vehicles using communication networks.
	In-Car data defines the connected devices or sensors in vehicles. Typical examples are navigation status, time distance computations, real-time pick-up / drop-off information for customers, and events detected by car to be communicated to other vehicles or GPS data to be transferred to maps.

1.1.3 Aggregated data

Aggregated data is generally created in order to address a single research question. They are supposed to be verified and cleaned, thus facilitating their usage for analysis purposes.
Aggregated data contains a specific part of the acquired or derived data (raw data). Its smaller size facilitates storage in e.g. database tables and an easy usage suitable for data analysis. To obtain aggregated data, several data reduction processes are performed. The reduction process summarizes the most important aspects in the data into a list of relevant parameters or events, through one or all of the following processes: validation, curation, conversion, annotation.
Besides helping in answering new research questions, aggregated data may be re-used with different statistical algorithms without the need to use raw data. For AUTOPILOT, aggregated data will represent the most important data types that will be shared by the project. It does not allow potentially problematic re-uses because it does not contain instantaneous values that would highlight illegal behaviour of a vehicle, a driver or another subsystem.

2.2 Metadata

This section provides the first recommendations regarding the description of the data provided by AUTOPILOT project. As the project will collect several data categories and several data types, several metadata descriptions must be provided to describe the characteristics of each measure or component and also the way data was produced and collected.
The AUTOPILOT project will follow and adapt the metadata type recommendation provided by the FOT-Net Data project (http://fot-net.eu/). This project identifies in its Data Sharing Framework several metadata types that can be applied to AUTOPILOT. The following list provides a first version of the metadata that will be managed by the project and their content. A more detailed version will be provided in the next version of the DMP at M16.

1.1.4 Metadata attributes of time-history data:

Time-history data corresponds to the history of a measurement over the time. Time-history data can be collected by the vehicle instrument, by an IoT device or by an IoT platform.
Time-history data stores a variation over the time a single or complex physical variable. To enable their re-use, each dataset provides a metadata description that includes the following descriptive attributes:

- Precision (accuracy)
- Unit of the measure
- Sample rate (frequency of the measure)
- Filtering (low-pass, interpolation, etc.)
- Origin (data source CAN, GPS, IoT platform, etc.)
- Type (Integer, Float, String)
- Error codes (full description of error codes)
- Quality (Quality measure related to this measure)
- Enumeration specification (Defines how to convert constant to correct value e.g. : 1 means Left, 2 means Right)

1.1.5 Metadata attributes of aggregated data

As aggregated data varies depending on the purpose of the experiment, it can be described as time history measures or as time segment. Time segment is a sub-set of data parameters (Trip / Platoon session) or measures generated by data summarization or data reduction.

This metadata type should include the following descriptive attributes:

- Description (Purpose of the aggregated data)
- Definition (Algorithm applied on the aggregated measures)
- Origin (Measures used to calculate the aggregated data)
- Unit (Unit of output value)

1.1.6 Metadata attributes of self-reported data

Self-reported data corresponds to interviews, surveys or questionnaires. This metadata type should include the following descriptive attributes:

- Description (Purpose of the questionnaire)
- Instructions (way how the collection process was executed)
- Type (Free text, single or multiple choices, etc.)
- Options (description of possible alternatives)

2.3 Recommendations for Data collection and sharing processes

The AUTOPILOT data collection process is built upon requirements coming from 4 processes:

- **The evaluation** defines the minimum data that must be collected in order to perform the evaluation process at the end of the project
- **The test specification** provides details about the data to be collected on the basis of the evaluation requirements and according to use cases specifications
- **The test data management** defines the data collection, harmonization, storage and sharing requirements using the first 2 processes and the ORDP process
- **The ORDP**[3] defines the requirement related to data sharing of research data

1.1.7 Evaluation process requirements

The evaluation process is defined in task 4.1 which develops the evaluation methodology. Named FESTA, this methodology must be implemented thoroughly and incorporated into the planning to guarantee that all pilots are collecting the required information needed for the evaluation.

The following figure shows a high-level view of the data that will be collected and integrated in the evaluation process. Different types of data (in blue) are collected, stored and analyzed by different processes in a predefined order. The workflow will be defined per pilot site but in a homogeneous way. The data types and requested formats will be defined in the evaluation task deliverable D4.1.

[3] https://www.openaire.eu/opendatapilot

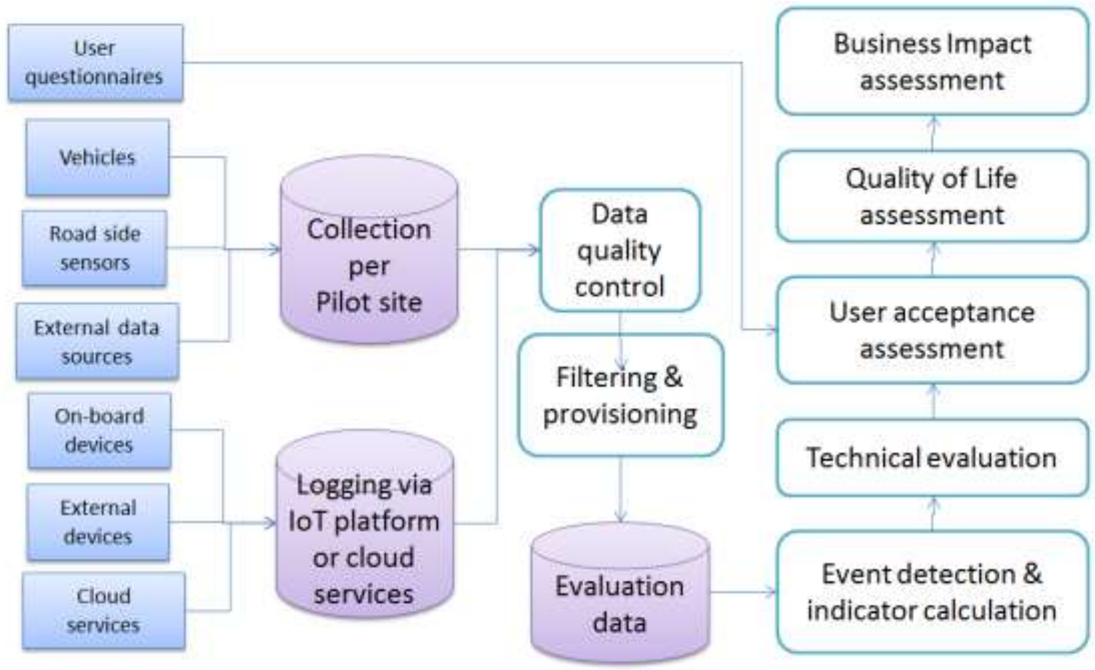


Figure 3 - Evaluation Methodology - Analyzing

To fulfil the project objectives, experiments will be designed during the evaluation task. This design creates a set of requirements that define: the number of scenarios and test cases, the duration of tests and test runs, the number of situations per specific event, the number of test vehicles, the variation in users, the variation in situations (weather, traffic, etc.). Each pilot site must comply with this design and provide sufficient and meaningful data with the required quality level, to enable technical evaluation. Refer to D1.1 for additional information regarding design of experiments and data quality (Time synchronization of devices & logging, Accuracy & frequency of logging, Alternative data sources, cross-checking from automated vehicles, on-board devices, road side detectors, detection of failures in systems and logging).

2.3.1.1 Tests Specification Process Requirements

The pilot tests specification Task T3.1 plays a major role that must be thoroughly implemented. Indeed, this task will convert the high-level requirements defined in the evaluation process into specific and detailed specifications of data formats, data size, data currencies, data units, data files, and storage. The list of requirements will be defined for each of the following items: Pilot sites, Scenarios, Test Cases, Measures, Parameters, Data quality, etc., and will be described in deliverable D3.1. All the development tasks of WP2 must implement completely, if impacted, the requirement defined in D3.1 in order to provide all the data (test data) as expected by the technical evaluation.

1.1.8 Open Research Data Pilot Requirement Process:

Additional requirements related to ORDP are defined in this document to guarantee that the collected data will be provided in compliance to European Commission Guidelines[4] on Data Management in Horizon 2020. Those requirements are defined in paragraph 3, in paragraph

[4] Guidelines on FAIR Data Management in Horizon 2020 - Version 3.0 - 26 July 2016

4 and paragraph 5 and an implementation scheme is provided.

1.1.9 Test Data Management Process requirements:

The test data management process that is part of task T3.4 will define the architecture that enables the implementation of the data collection processes in all the AUTOPILOT sub-systems.
The data collection process spans over all AUTOPILOT sub-systems: Vehicle System, IoT Device, IoT platform, Pilot Sites Test Servers, Centralized Test Servers. The test data management process will focus on services that will be deployed on the Pilot Site Test Servers and on the Centralized Test Server. A specific pilot site test server must collect and store all the data created by a pilot site: contextual data, acquired data and derived data, aggregated data (optional), and metadata. However, the centralized test server will collect for all the sites the aggregated data and their metadata requested for the evaluation.

1.1.10 Recommendation on data management at Pilot site level

We provide a first list of recommendations that shall be followed at the pilot site level. This list will be updated in the following DMP versions:

- Each Pilot site must ensure the data collection according to the evaluation requirement and with the right accuracy.
- Each pilot site must provide, in a timely fashion, the requested data to perform test analysis. Data must be provided with its metadata or a complete description. To enable an accurate usage, data must be generated in compliance with evaluation requirements (formats, measurements, etc.).
- Each pilot site must ensure that IoT data remain in the device until they are stored in the platform.
- Concerning Data formats requirements, Data to be sent to the central IoT platform must follow a predefined schema, non-compliant data being ignored.
- Each pilot site must provide an easy access to collected data and a secure environment for data storage.
- Each pilot site needs to have all the necessary tools to check data quality and to read raw data. Automated scripts may be provided to process large datasets in order to ease and enable post-processing of aggregated data.

2.4 Data Flow in AUTOPILOT Test Environment

The figure shows the data flow in the AUTOPILOT test environment:

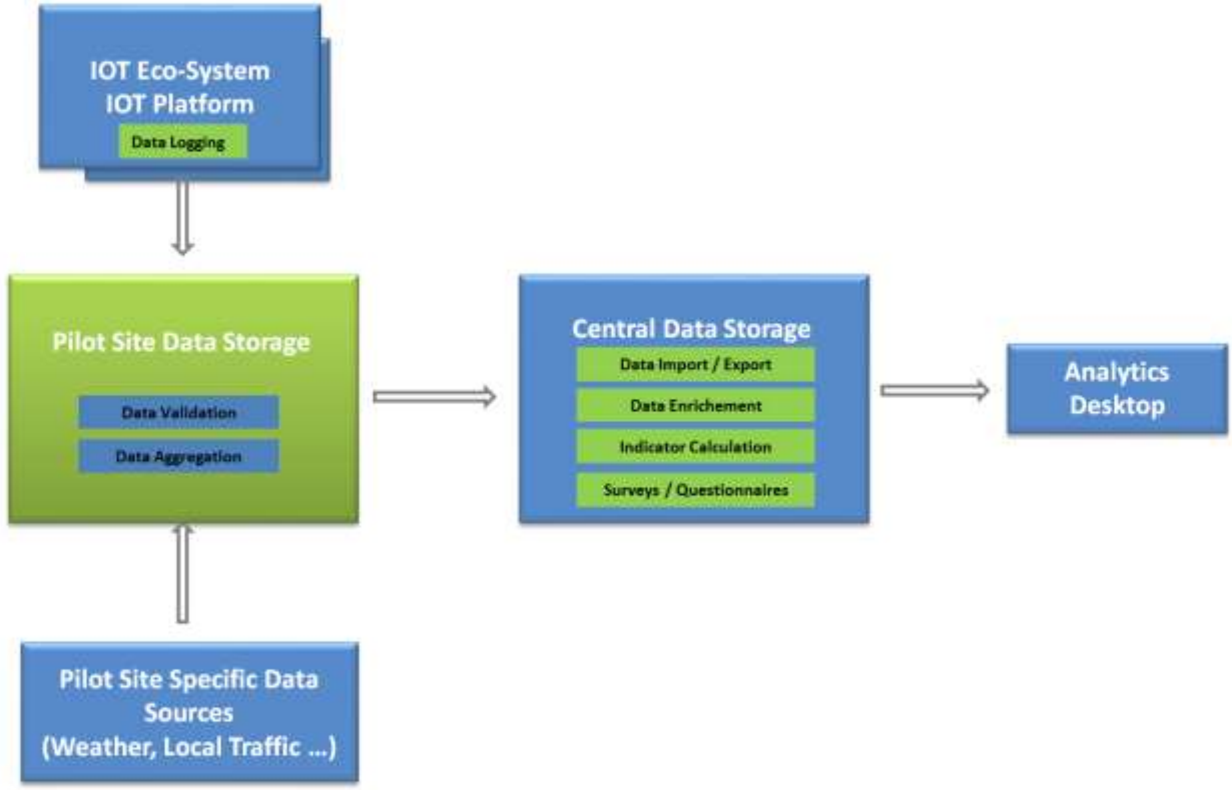


Figure 4 - Data flow in AUTOPILOT test environment

During the pilot site test, different data flows will be exchanged between data sources and platforms. As mentioned above in the test data architecture, IOT data will come from the eco-system and from IOT platforms and from external data sources.

Each Pilot Site Test Server will store IOT data and additional data coming from the specific pilot site (weather, local traffic information …). These data will be validated and aggregated before the central data storage. This last storage will serve for data validation and will enable data querying and enrichment. Finally, Aggregated and valid data are transferred upon request to the data analyst for data analysis.

2.5 AUTOPILOT Test Data environment architecture

The following figure depicts the test data environment that will be deployed for AUTOPILOT project to collect all the data generated by the project.

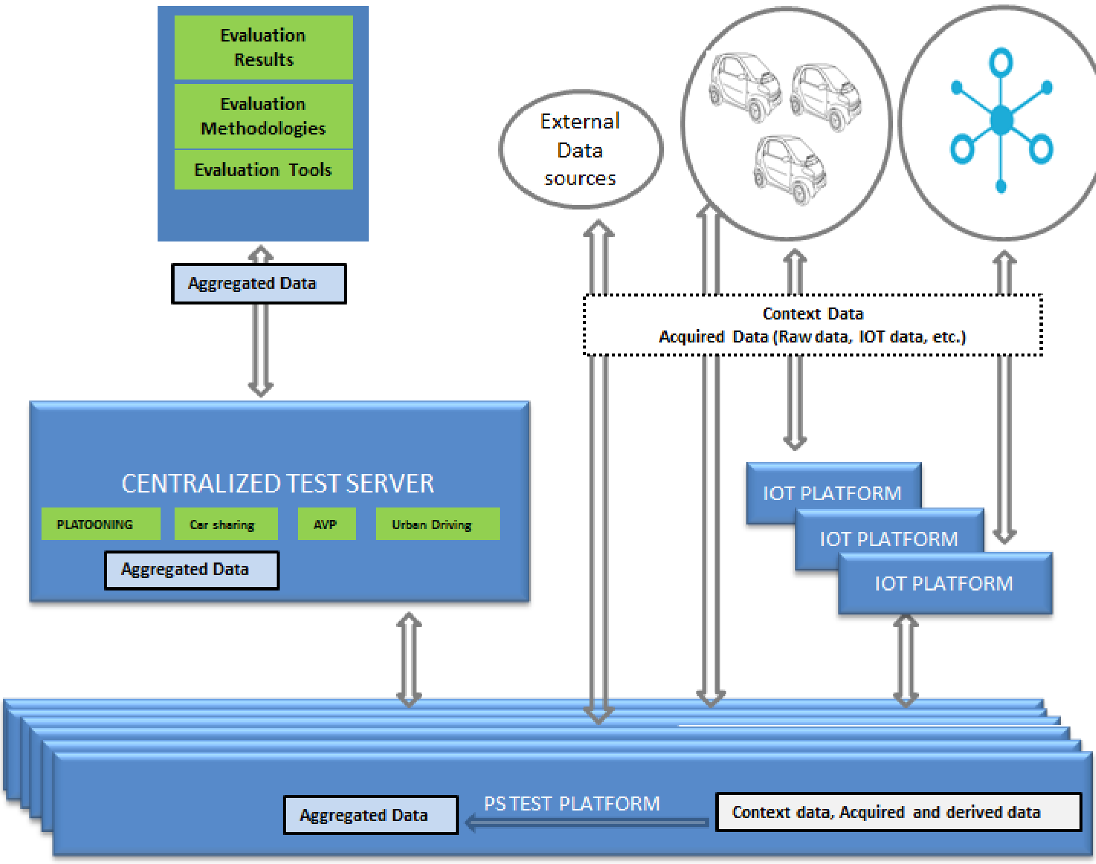


Figure 5 - Pilot Site Test Overview

On each pilot site test, the acquired data (IOT data) coming from the environment and the data coming from vehicles will be generated and transferred through the local IOT Platforms in order to be stored temporarily. The rest of vehicle data will be transferred directly or indirectly to the Pilot Site platform through interfaces (FTP, HTTP, File System) defined in Task 3.4. In addition to the previous data, external sources will be involved in the project and their data will be transferred later to the pilot site test platform.

The pilot site test platform will be used as the distributed storage for each pilot site which aggregates the stored contextual, acquired and derived data.

The stored data will be validated for evaluation purposes and it will be sent to the centralized test server with the appropriate metadata requested for evaluation.

The evaluation will be done following the methodologies and tools defined in WP4, and the KPIs mentioned in the AUTOPILOT Grant Agreement[5].

[5] GRANT AGREEMENT NUMBER — 731993 — AUTOPILOT

3. Participation in the Open Research Data Pilot

The AUTOPILOT project 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 results provide the greatest impact possible.

AUTOPILOT will ensure the open access[6] 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 AUTOPILOT 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 (AUTOPILOT, 731993)
- Publication date
- Length of embargo period if applicable
- Persistent identifier

When referencing Open access data, AUTOPILOT will include at a minimum the following statement demonstrating EU support (with relevant in-formation included into the repository metadata) [3]:
- “This project has received funding from the European Union’s Horizon 2020 research and innovation program under grant agreement No 731993”.

The AUTOPILOT 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. Paragraph 4.2)

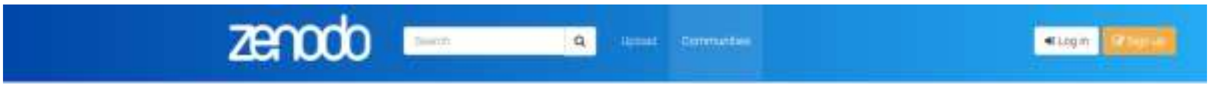
In regards to the specific repositories available to the AUTOPILOT consortium, numerous project partners maintain institutional repositories that will be listed in the following DMP version, where project scientific publications and in some instances, research data will be deposited. The use of a specific 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 ZENODO[7]. It is a free service developed by CERN under the EU FP7 project OpenAIREplus (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.

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

The AUTOPILOT data collection can be accessed in ZENODO repository in a similar address as the following link: https://zenodo.org/collection/<<autopilot>>



In summary, as a baseline AUTOPILOT partners shall deposit:

- Scientific publications – on their respective institute repositories in addition (when relevant) to the AUTOPILOT Zenodo repository
- Research data – to the AUTOPILOT Zenodo collection (when possible)
- Other project output files – to the AUTOPILOT Zenodo collection (when relevant)

This version of the DMP does not include the actual metadata about the Research Data being produced in AUTOPILOT 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 section 4.2 and will be used by project partners to provide all requested information.

4. AUTOPILOT Dataset Description

4.1 General Description

This section provides an explanation of the different types of datasets to be produced or collected in AUTOPILOT project, which have been identified at this stage of the project. As the nature and extent of these datasets can evolve during the project, more detailed descriptions will be provided in the M16 and M32 versions of the DMP.
The descriptions of the different datasets, including their reference, file format, standards, methodologies and metadata and repository to be used are given below.

4.2 Template used in dataset description

This table is a template that shall be used to describe the datasets (see an example in Table 2).

Table 1 – Dataset Description template

Dataset Reference	AUTOPILOT_WPX_TX.X_XX Each dataset will have a reference that will be generated by the combination of the name of the project, the Work Package and Task in which it is generated and (for example: AUTOPILOT_WP3_T3.4_01)
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 The used methodologies
File format	All the format that defines data
Data Sharing	Explanation of the sharing policies related to the dataset between the next options: **Open**: Open for public disposal **Embargo**: It will become public when the embargo period applied by the publisher is over. In case it is categorized as embargo the end date of the embargo period must be written in DD/MM/YYYY format. **Restricted**: Only for project internal use. Each dataset must have its distribution license. Provide information about personal data and mention if the data is anonymized or not. Tell if the dataset entails personal data and how this issue is taken into account.
Archiving and Preservation	The preservation guarantee and the data storage during and after the project (for example : databases, institutional repositories, public repositories …)

4.3 Pro-forma sample description of Infrastructure IOT Data

This pro-forma table is an example of the Infrastructure IOT dataset description to be used in the next updates.

Table 2 - Dataset Description Example

Dataset Reference	AUTOPILOT_WP3_T3.3_01
Dataset Name	Optimization of platooning performance using IoT information from traffic lights
Dataset Description	This dataset was generated during a platooning in urban driving condition in Versailles Pilot Site. It demonstrates the benefits of IoT integration with respect to smooth and efficient moving of the platoon in traffic by timely anticipating traffic lights. Platooning data are aggregated data and contains data coming from connected devices, vehicles, and traffic lights infrastructure. They are mainly measurements that describe the mobility of the vehicles forming the platoon (Speed, acceleration, inter-distance, heading, position status, etc.) and the traffic light status.
Standards and metadata	The metadata is based on ZENODO's metadata, including the title, creator, date, contributor, description, keywords (IOT, Autonomous Driving, Platoon, Urban Driving, etc.), format, resource type, etc. The recommendations mentioned in the section 2.2 will be taken in consideration.
File format	XML, CSV, XLS
Data Sharing	This dataset will be widely open to be used by 3rd party applications and will be deposited in the ZENODO repository. License : Creative Commons License (CC0 : No Rights Reserved)
Archiving and Preservation	During the project, the data will be stored in local pilot site repositories and IoT Platforms. Then, this dataset will be archived and preserved in ZENODO.

5. FAIR Data Management Principles

The data that will be generated during and after the project should be 'FAIR', that is findable, accessible, interoperable and reusable. These requirements don't 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.

Making data findable, including provisions for metadata

- The datasets will have very rich metadata to facilitate the findability
- All the datasets will have a Digital Object Identifiers provided by the public repository (ZENODO)
- The reference used for the dataset will follow this format : AUTOPILOT_WPX_TX.X_XX
- The standards for metadata will be defined in the chapter 4 tables

Making data openly accessible

- All the datasets that are openly available will be described in the chapter 4
- Open access datasets will be made available through a public repository (e.g. ZENODO)
- Data sharing in section 4 explains the methods or software used to access the data. Basically, no software is needed to access the data
- Data and their associated metadata will be deposed in a public repository or either in an institutional repository
- Data sharing in the section 4 will outline the rules to access the data if restrictions exist

Making data interoperable

- Metadata vocabularies, standards and methodologies will depend on the public repository and will be mentioned in the section 0 tables

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 constrains exist, an embargo period will be mentioned in the section 4 tables 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

6. Responsibilities

In order to face data management challenges efficiently, All AUTOPILOT partners have to respect the policies set out in this DMP and datasets have to be created, managed and stored appropriately.

The Data controller role within AUTOPILOT will be undertaken by Francois Fischer (ERTICO) who will directly report to the AUTOPILOT Ethics Board. The Data controller acts as the point of contact for Data Protection issue and will coordinate the actions required to liaise between different beneficiaries and their affiliates, as well as their respective Data Protection agencies, in order to ensure that data collection and processing within the scope of AUTOPILOT, will be carried out according to EU and national legislation. Regarding the ORDP, the data controller must ensure that data are shared and easily available.

Each data producer and WPL is responsible for the integrity and compatibility of its data during the project lifetime. The data producer is responsible for sharing its datasets through open access repositories. He is in charge of providing the latest version.

Regarding ethical issues, the deliverable D7.1 details all the measures that AUTOPILOT will use to comply with the H2020 Ethics requirements.

The Data Manager role within AUTOPILOT will directly report to the Technical Meeting Team (TMT). The Data Manager will coordinate the actions related to data management and in particular the compliance to Open Research Data Pilot guideline. The data manager is responsible for implementing the data management plan and he ensures it is reviewed and revised.

7. Ethical Issues and Legal Compliance

Ethical issues related to the AUTOPILOT project will be addressed in the D7.1.

As explained in Chapter 2, the project will deploy a central IoT platform, which will be a cloud-based platform. It will integrate and aggregate data from the various vehicles and pilot sites. WP1 will provide a deployment view of the architecture. Hosting may be ensured for instance partly or entirely on an IBM infrastructure maintained by IBM IE.

Whatever the deployment choices made by the project, all data transfers to the central IoT Platform are subject to and conditional upon compliance with the following requirements:

- **Prior** to any transfer of data to the central IoT platform, all partners must execute an agreement as provided for in Attachment 6 of Autopilot Collaboration Agreement.
- All the partners must agree to commit not to provide personal data to the central IoT platform and to represent that it secures all necessary authorizations & consents before sharing data or any other type of information ("Background, Results, Confidential Information and/or any data") with other parties.
- Every partner that needs to send and store data in the IBM-hosted central IoT platform has to request access to the servers, and inform IBM IE what type of data they will send.
- IBM IE will review all data sources BEFORE approving them and allowing them into the IBM-hosted central IoT platform, to ensure they are transformed into data that cannot be traced back to personal information.
- No raw videos/images or private information can be sent to the central IoT platform. The partners who will send data to the platform must anonymize data first. Only anonymized information that will be extracted from the raw images/videos (e.g., distance between cars, presence of pedestrians, etc.) will be accepted and stored.
- The central IoT platform will only be made available to the consortium partners, and not to external entities.
- IBM IE reserves the right to suspend partner's access in case of any suspicious activities detected or non-compliant data received on its infrastructure. IBM IE may re-grant access to the platform if a solution demonstrating how to prevent such sharing of personal data and sensitive personal data is reached and implemented.
- IBM IE may implement validation procedures to check that the submitted data structures and types are compliant with what the partners promised to send to IBM-hosted central IoT platform.
- IBM IE will ensure that all the data will be deleted at the end of the project from all its servers of the central IoT platform.

Privacy and security issues related to the AUTOPILOT project will be outlined in the D7.1 and addressed in the WP1 Task 1.5 for Security, Privacy and Data Specification issues.

8. Conclusion

This deliverable provides an overview of the data that AUTOPILOT project will produce together with related data processes and requirements that need to be taken into consideration.

The document outlines an overview about the dataset types with detailed description and explains the processes that will be followed for test sites and evaluation within high level representations.

Chapter 4, which describes the datasets, will be incrementally enriched along the project lifetime. These descriptions include a detailed description, standards, methodologies, sharing and storage methods.

The Data Management Plan is a living document. In order to prepare for the next version, the contribution guidelines will be presented in one or several dedicated workshop(s) that will be organized by AKKA (Data Manager) not later than M12 (at least one, other according to the needs). The workshop(s) will explain the scope of the data management plan and the recommendations to take into account during the data collection/generation of their datasets, how to fill in the template provided in the section 4.2. Face-to-face meetings or webinar tools will be used.

9. Annexes

9.1 Annex 1: Open Research Data 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 AUTOPILOT 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.

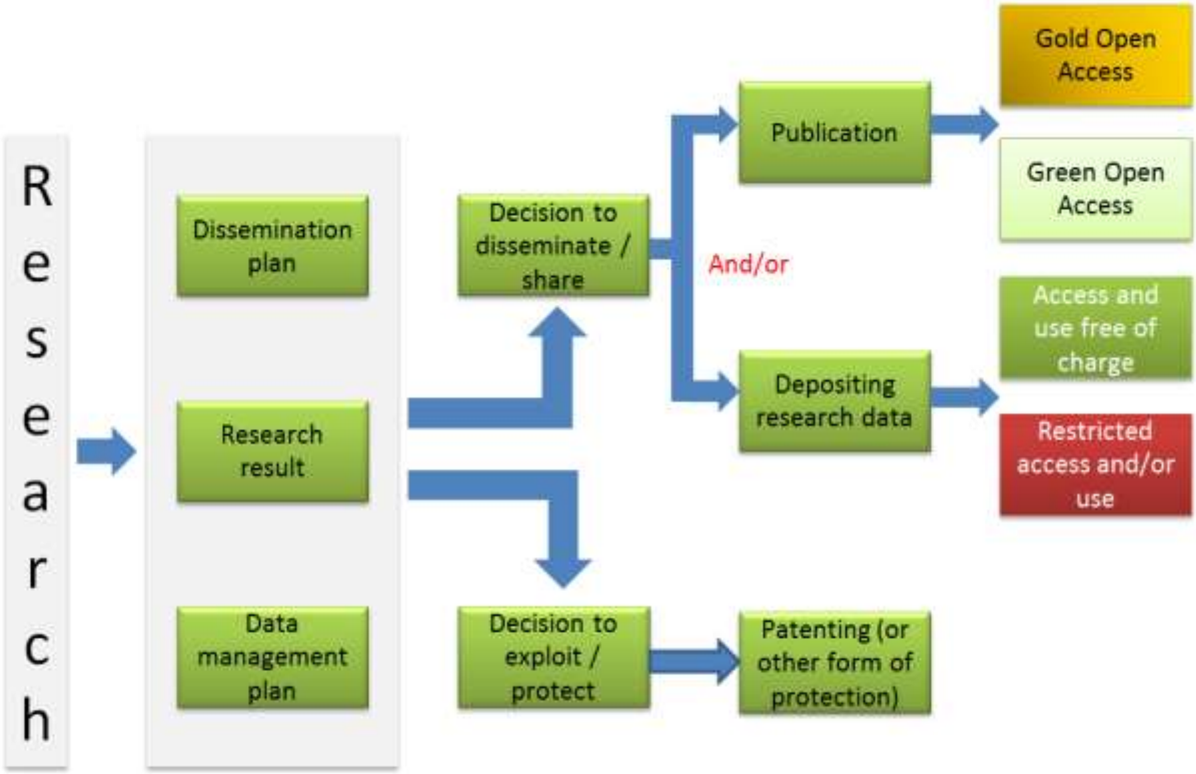


Figure 6 - Principles of H2020 Open Access to Research Data