SHared automation Operating models for Worldwide adoption

SHOW

Grant Agreement Number: 875530

D14.3 Data Management Plan (DMP) – final version

This report is part of a project that has received funding by the European Union's Horizon 2020 research and innovation programme under Grant Agreement number 875530

Legal Disclaimer

Executive Summary

The D14.3 Data Management Plan final version, utilizing the knowledge gained, builds on, integrates, and updates D14.2 Data Management Plan – 1st version of the DMP, elaborated within the SHOW project framework and released in the first period of the project, referring to what data the project will generate, whether and how it will be exploited or made accessible for verification and re-use, and how it will be curated and preserved. The current version of the deliverable also contains information about the Data Privacy Impact Assessments (DPIA), and personal data collected, in accordance with the GDPR and it is complementary to *D3.5: Final SHOW Ethics manual, Data Protection Policy and Data Privacy Impact Assessment* released during this period.

The Data Management Plan (DMP) has been developed as required by the implementation of a Horizon 2020 limited pilot action on open access to research data and refers to the latest EC DMP guidelines.

Data that will be produced within the project lifecycle shall be Findable, Accessible, Interoperable and Reusable (FAIR). It describes what kind of 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.

The current version of the SHOW Data Management Plan is the final official one; however, it will be a living document and it will evolve during the project lifespan according to the progress of project activities, capable to capture and reflect the evolution in the form of dataset updates and/or changes in Consortium policies, complementing *D3.5: Final SHOW Ethics manual, Data Protection Policy and Data Privacy Impact Assessment*.

Document Control Sheet

Start date of project:	01 January 2020
Duration:	48 months
SHOW Del. ID & Title:	D14.3 Data Management Plan (DMP) – final version
Dissemination level:	Public
Relevant Activities:	A14.6: Data Management - ERTICO
Work package:	WP14
Lead authors:	Nikolaos Tsampieris (ERTICO)
Internal Reviewers:	CERTH
External Reviewers:	N/A
Actual submission date:	23/03/2022
Status:	Final
File Name:	SHOW_D14.3_DMP_Final

Document Revision History

Version	Date	Reason	Editor
0.1	01.11.2021	Table of Contents for feedback	Nikolaos Tsampieris (ERTICO)
1.0	28.02.2022	Version sent for internal peer review	Nikolaos Tsampieris (ERTICO)
1.9	10.03.2022	Final version sent for proof-checking	Nikolaos Tsampieris (ERTICO)
2.0	22/03/2022	Final proof-checked version sent for submission.	Maria Gkemou (CERTH)

Table of Contents

List of Tables

List of Figures

Abbreviation List

Abbreviation	Definition
AV	Automated Vehicle
CCAV	Cooperative Connected Automated Vehicle
C-ITS	Collaborative ITS
D	Deliverable
D3WC	International Boundary & Water Commission
DCAT-AP	Data Catalogue Application Profile
DMP	Data Management Plan
DOI	Digital Object Identifier
DPIA	Data Privacy Impact Assessment
DPO	Data Protection Officer
EC	European Commission
EN	European Standard, telecommunications series
ETA	Estimated Time of Arrival
ETSI	European Telecommunications Standards Institute
EU	European Union
FAIR	Findable, Accessible, Interoperable, and Reusable
GDPR	General Data Protection Regulation
GPS	Global Positioning System
IoT	Internet of Things
IPR	Intellectual Property Rights
ISA	Interoperability solutions for public administrations, businesses and citizens
MaaS	Mobility-as-a-Service
OA	Open Access
OEM	Original Equipment Manufacturer
ORDP	Open Research Data Pilot
POPD	Protection Of Personal Data
RT	Reaction Time
SME	Small and Medium Size enterprise
VEC	Vulnerable to Exclusion
W3C	World Wide Web Consortium
WOT	Web of Things
WP	Work Package

1 Introduction

1.1 Purpose of the document

This deliverable - D14.3 - is an update of the *D14.2-Data Management Plan- 1st version* (DMP) using knowledge gained with the purpose toprovide an analysis of the main elements of the data management policy that will be used by the consortium, concerning the datasets that will be generated and/or collected during the project.

SHOW is an H2020 funded project, and within this framework, the DMP must at least cover the specific aspects of the project's datasets, in accordance with the EC DMP guidelines [1], it describes the data management life cycle for all datasetsto be collected, processed, or generated by a research project. It must cover:

- the treatment of research data during the project lifecycle and beyond;
- which data to be collected, processed, and/or generated;
- which methodology and standards will be applied;
- if will data be shared or it will be open access;
- how data will be curated and preserved.

The DMP plays a crucial role in the project's success because it ensures the availability and the quality of the datasets, which will be generated/used within the Project framework, serving the achievement of the project's objectives. The SHOW 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).

1.2 Intended Audience

The SHOW project addresses highly innovative concepts. As such, foreseen intended audience, due to the strong expected impact of the project on their respective domains, is the project authorities (Cities, Municipalities, Ministries) and policymakers, OEM's and transport operators, Tier 1 suppliers, telecom operators, and technology providers, SME's operating in the area, road operators, passengers as well as all road users encompassing VEC (Vulnerable to Exclusion) citizens (through their associations).

Moreover, a further intended audience is the scientific community (research and academia) with an emphasis in the areas of intelligent transport systems, road safety, automotive engineering, and standardisation organisations.

1.3 Interrelations

The DMP is part of WP14 and is closely linked to data collection and processing activities, such as WP5 Big Data collection, processing, and analytics, the pilots, where the primary data sources will come for the pilot related activities (WP9-WP13), taking into consideration the data clusters and specifications that will be defined in WP9, the data relating processes for treating data (collecting, storing, sharing, analysing, and reporting data) in WP5,11,12,13, and 14, but also when only data sharing takes place to perform other activities (e.g. simulations in WP10). Some partners will be only processors (e.g. those involved with the processing of either subjective or performance data for research reasons), so their role is dependent on the decision for data collection taking place by the processors in other activities (e.g. WP1, WP10, WP11, WP12, WP13).

2 Data Summary in SHOW

2.1 Overview

This section aims to:

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

The project involves data collection (in the context of the piloting and validation phases) and a series of SHOW field trials to assess the technologies and effectiveness of the proposed CCAM solutions in real-life conditions. It should be noted here, that due to the fact that the project will collect (partially) personal-related data, the Consortium will comply with any related European and national legislation and Directives relevant to the country where the data collections are taking place; this is dealt in detail in Deliverables D3.2, D3.4 and D3.5.

The SHOW project will collect a large amount of raw data including – among others – real-time GPS data, vehicle sensor data, traffic plans, routes, occupancy levels, RT delays, ETAs, road infrastructure data, etc.

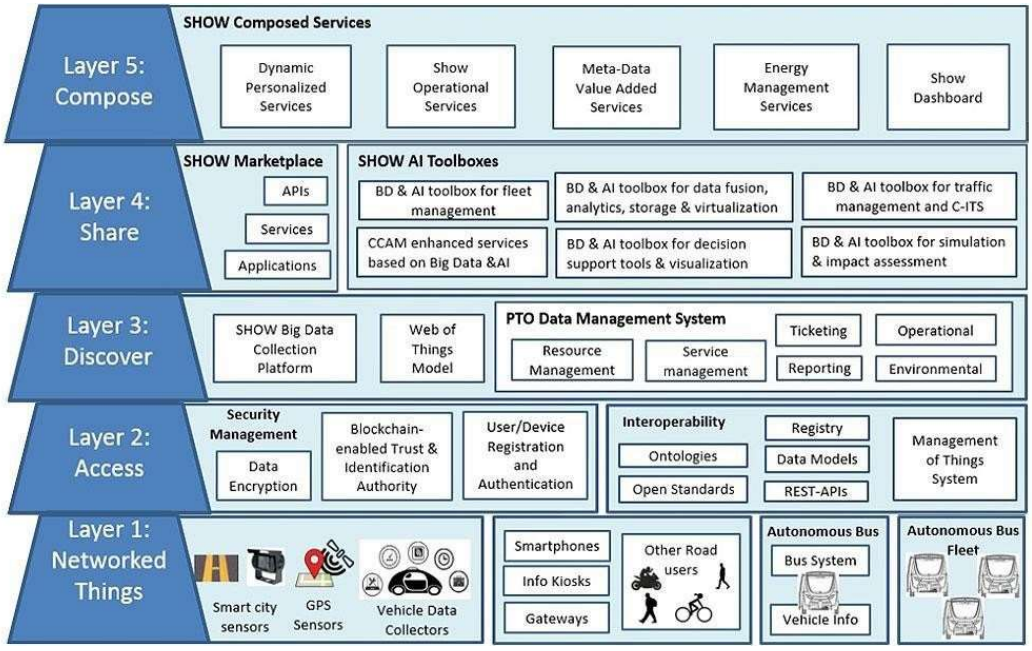

Figure 1: SHOW Architecture as proposed in D4.1 [7]

SHOW's reference architecture for AVs, which is further analyzed in [7], provides an interoperable environment to boost innovative connected vehicle-based business and services to allow communication and cooperation between vehicles, infrastructure, and with

other road users and to enable automated, smart mobility services, innovative fleet management concepts and higher performance of automated vehicle functions. Thus, at the architecture level, SHOW will improve the current conceptualisation of the W3C new incoming global standard Web of Things layer architecture utilizing the best practices of both world (Web and IoT) and providing several functionalities needed for enabling the domain of the connected vehicles by adopting the current W3C WoT layer architecture that includes the following 4 layers: **Access** (provide access to a thing), **Find** (discover things), **Share** (enables things exchange) and **Compose** (enables Things creation from existing ones).

Focusing on the lower Layer 1 Networked Things (Compose), there is a wide collection of heterogeneous sensors and edge IoT devices which include but are not limited to GPS and smart city sensors, data collectors from conventional vehicles, user's mobile devices (and wearables), info kiosks, gateways, autonomous buses Geo-spatial and other fleet/passenger/traffic situation/service real-time information).

Moreover, subjective data collection through all types of subjective surveys, questionnaires, personal and group interviews (WP1 – WP9) takes place in the project. The data may be collected from users/passengers/customers, data providers and operators as well as other types of stakeholders. In all cases, the data is anonymised to ensure privacy and protection of the participants' identity (as defined in D3.2, D3.4 and D3.5). In the case of the users/customers, subjective data mostly deal with travel preferences and needs (and acceptance ratings upon specific solutions/services tried). Additional data sources are used and required for the development of tools and new services, exploiting Big Data and AI, for the data management platform (Table 1).

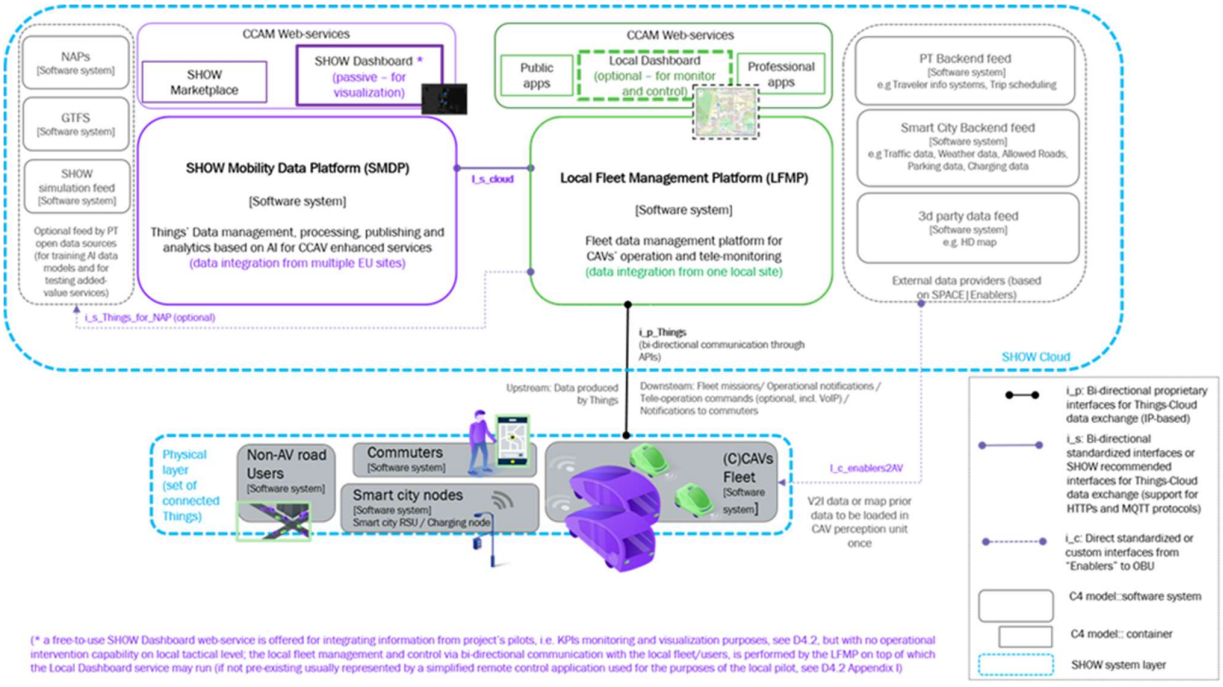

Figure 2: SHOW reference architecture (better viewed in zoom-in mode)

The architecture as seen in **Figure *1*** describes the first proposed version for SHOW project architecture. In **Figure 2**, the updated approach for SHOW project architecture is presented and is further analyzed [7]. Each Pilot site possesses a Local Fleet Management Platform (LFMP),that is connected to the central SHOW Mobility Data Platform (SMDP). Each LFMP

collects the data from the vehicles of each pilot site and visualizes the data to a local dashboard. The data then is forwarded to the central SMDP for processing and is used to develop corresponding KPIs as they are specified in Table 13 of the latest revised D4.1 deliverable [6]. The calculated KPIs are visualized in SHOW Dashboard (https://demo.innovationcloud.ericsson.net/show-project/home). The commuters can access the data that is located in both LFMP and SMDP with corresponding interfaces, according to their role and permissions granted within the SHOW system.

Table 1: Data sources and services used in SHOW.

Data sources	Services
Event data (event type, start time, location and other attributes), mobility data (taxi, buses, autonomous cars location, availability, delays, traffic congestion, etc.)	Demand prediction accounting for the impact of special events for travellers,traffic management centres, public transport service providers, cities authority, event managers, etc.
Bus operation data (location, delays, routes, demand, etc.)	Bus arrival-time / travel-time prediction for travellers, public transport operationcentres, traffic management centres/ operators.
Fleet location data, origin-destination demand data (e.g. passenger request through apps or MaaS platforms.), traffic congestion data (e.g. floating car data)	Predictive routing for a fleet of autonomous vehicles for OEMs, drivers(SAE 3), travellers, navigation service providers, fleet operators, traffic management operators
Telematics, floating car data and external (historical) sources (for example meteorological data)	Fleet management for fleetoperators, freight forwarders and relevant technology providers
Data from conventional equipment such as loops, cameras and radars along with innovative ones such as probe dataand data from connected vehicles. CCAV can be an important source of data, since they can assist in getting a better view of the entire urban environment	Traffic management and C-ITS for traffic managers, traffic control centres and relevant technology providers
Floating car data, Probe data, Social media data, Bluetooth detections	Mobility patterns identification and prediction for mobility service providers,public authorities and public transport operators

Data sources	Services
Traffic flows, vehicle counts, speed measurements, time/speed profiles, mobility patterns, traffic status, etc.	Decision support tools for public administration officials, public transport operators, mobility service providers

The above-mentioned data, collected and managed through the SHOW Data Management Portal (of WP5), where all relevant internal and external data assets will be continually collected and registered, with their appropriate metadata according to the standards that are described in the chapter 3.3 of the current document. described in detail in Deliverable D5.1 Big Data Collection Platform and Data Management Portal.

2.2 Datatypes, Datasets and Dataset Categories

This section presents a description of the data collected, generated and managed in SHOW. More specifically, data is clustered into three different sections: data types, datasets and datasets categories. Data types are related to the source of the data, i.e. vehicle data, traffic data, etc., datasets refer to the file extension of the data and, finally, dataset categories refer to the level of processing that the data has undergone.

The types of data generated, collected and managed within SHOW fall into the following categories:

- Subjective data (user profile/demographics and user request related data)
- Vehicle related data
- Urban environment related data (traffic and city related data)
- Infrastructure related data

As far as datasets are concerned, the following will be handled by SHOW:

- Reports (in the form of word or pdf documents)
- Spreadsheets with raw data, as received/collected by sensors, surveys, etc.
- Video signals
- Databases

The following table is a template used to describe the different datasets, including their reference, file format, standards, methodologies and metadata and repository to be used.

Table 1: Dataset Description Template.

Dataset Reference	SHOW_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 isgenerated
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 and 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 mentionif the data is anonymized or not. Tell if the dataset entails personal data and how this issue is taken intoaccount.
Archiving and Preservation	The preservation guarantee and the data storage during and after the project (for example: databases, institutional repositories, public repositories …)

The SHOW project different categories of datasets are:

- Context data: data that describe the context of a pilot/study.
- Acquired and derived data: data that contain all the collected information related to an pilot/study.
- Subjective data: questionnaires, surveys, personal and group interviews.
- Raw/unprocessed data: data collected directly from the source (either objective or subjective).
- Metadata: descriptions of data that will facilitate the data analysis and data pre-

processing.
- Aggregated data: data summary obtained by reduction of acquired data and generally used for data analysis.
- Consolidated data: data collected across sites and per data type.

As the nature and extent of these datasets can evolve during the project, these can still be referenced in the DMP, according to our understanding that the DMP can be enriched and updated throughout the project lifecycle. A data collection template, with required information per data cluster (e.g. survey, questionnaires, vehicle, automated, purely formative) was circulated in a first attempt to identify the higher-level clusters and the information needed for them. In detail, the template collects information about the following:

- The name of the data
- Whether the data was collected or created
- Data description
- Data category
- Data type
- Data format
- Data size
- Data Ownership
- Privacy level
- Data repository during the project (for private/public access)
- Data sharing
- Back-up frequency
- Status of data at the end of the project (destroyed or not)
- The duration of the data preservation (in years)
- Data repository after the project is complete

Table 2 shows a sample example of the aforementioned data collection template.

Table 2: Data collection template.

Data clustering template														
Collected/ Created	**Name**	**Description**	**Category**	**Type**	**Format**	**Size**	**Owner**	**Privacy level**	**Repository during the project (for private/public access)**	**Data sharing**	**Back-up frequency**	**Destroyed at the end of the project?**	**Duration of preservation (in years)**	**Repository after the project**
Collected/ created	Name of the data/ metadata/ exploitable result	Description of the data/ metadata - Please provide the description of the information you will collect/create (example: is it information about mobility habits from pilot users? Is it information related to vehicles that are used in the pilots? What is the source of the information?	What category does the information match? FW/SW/ Algorithm/Raw data/ Dissemination material/etc.	What is the type of information? Document/video/images /Source code/etc.	File extention/ prototype (docx, .txt., etc)	size in MB/GB	Partner name/ Consortium/ external stakeholder	Public/ consortium/ partner/ etc.	BAL.PM or other Open access repository/ partner storage(private cloud/private drop box)/etc.	**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 data set 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.	How often will you make backups? daily/ monthly / yearly/once	NO (1)/No (2)/NO (3)/ Yes / Unnecessary	number of years	Open access repository/ partner storage (private cloud/private drop box), etc.

Collected/ Created	Name	Description	Category	Format	Size	Owner	Privacy level	Repository during the project (for private/public access)	Data sharing	Back-up frequency	Destroyed at the end of the project?	Duration of preservation (in years)	Repository after the project
Collected		Vehicle-kilometres driven	Raw Data	File extension		Navya	partner	partner storage	Restricted	Daily	No	Undefined	Undefined
Collected		vehicle-kilometres automatic driven	Raw Data	File extension		Navya	partner	partner storage	Restricted	Daily	No	Undefined	Undefined
Collected		locational losses signal numbers	Raw Data	File extension		Navya	partner	partner storage	Restricted	Daily	No	Undefined	Undefined
Collected		trajectory's mistakes	Raw Data	File extension		Navya	partner	partner storage	Restricted	Daily	No	Undefined	Undefined

The information enables us to identify the types of data and determine how to manage it within the Data Management Plan and as instructed by the General Data Protection Regulation (GDPR) principles.

Personal-related data are centrally stored in an anonymised and secure standards- abiding way and in accordance with the GDPR [Regulation (EU) 2016/679 of the European Parliament. Also, data will be scrambled where possible and abstracted in a way that will not affect the final project outcome. Hence, it becomes apparent that part of the research data generated by the project will be made open & will be offered to the Open Research Data Pilot, in which SHOW has declared its intention to participate (Section 3). Also, the allocation of specific roles in the Consortium in compliance with the GDPR requirements, briefly touched upon for completeness in Section 5, are being treated in detail in Deliverables D3.2, D3.4 and D3.5 The type and format of data upon the FAIR principles (Section 3) their way/means of collection (against enhanced GDPR compliant templates) and the open access layer in each category has been also addressed in the Deliverable D3.2 where the Data Protection Policy is included, and in D3.4 and D3.5 where additionally, the Data Privacy Policy and the Data Privacy Impact Assessment is included.

2.3 Data Processing – Record keeping

Data processed within SHOW fall into one of the following categories:

- Personal Data collected in the context of participation in a research study,
- Contact information such as name, address, telephone number and email address,
- Banking and other financial information for payment or invoicing purposes,
- Information about how one uses websites, for the purpose of making them more user-friendly, for example via cookies and
- Information about participation in conferences or courses

Personal data included in this scope (shown also in the data list in Appendix IV, is mainly data considering users and passengers), are also the following:

- Dynamic data (“wheelchair on board”, “passenger with special needs”),
- Booking-ride data (passenger location, destination and timestamp, desired pickup and drop off location and time, planned pickup and drop off location and time, actual pickup and drop off location and time, planned and actual booking route, direct ride distance and duration, actual ride distance and duration and trip reason),
- Infrastructure and other data (internal and external video cameras, traffic in vehicle’s route and vehicle traffic cameras, etc.).

This kind of data, as specified above, must be handled in an exhaustive manner as GDPR regulations dictate, since personal information about a user can be extracted in manipulative ways, while external access to this kind of data may act as a threat; it should be ensured that processing of these types of data is adequate, relevant and limited to what is necessary.

Data processing should be accurate, ensuring that Data are correct and kept updated; a

storage limitation time should be observed so that Personal Data are not kept more than needed for the set purposes. Proper security measures - organisational and technical- e.g. (pseudo)anonymization & encryption should be in place to safeguard the data security, integrity and confidentiality (e.g. access to data is restricted and password protected).

Data Processing in SHOW, is being performed in accordance with the GDPR key principles of a) Lawfulness, fairness and transparency, b) Purpose limitation, c) Data minimization, d) Accuracy, e)Storage limitation, f) Integrity and confidentiality (security) and g) Accountability Moreover, SHOW ensures that the rights for individuals, under the GDPR, such as a) the right to be informed, b) the right of access, c) the right to rectification, d) the right to erasure, e) the right to restrict Processing, f) the right to data portability, g) the right to object and h) rights in relation to automated decision making and profiling are preserved.

A Data Privacy Impact Assessment (DPIA), a process to help the identification & minimization of the data protection risks in a project, it is mandatory in SHOW for all test sites to consider, and perform one, if deemed necessary. In such a case, where the Processing poses a high risk to the individuals involved, the DPIA is conducted at each pilot site in collaboration with the data collectors and processors (respective templates have been annexed in Appendix III), the project DPO and the Data Manager. (crf D14.2, D3.4 and D3.5). In SHOW it is mandatory for all test sites to consider if a DPIA is needed and if yes perform one; the notion and the necessity of the DPIA's has been treated in D3.4 and the results of DPIA's performed in the various project sites have been presented in D3.5.

As part of this process, information specific to personal data processing has been requested from the pilot sites, in line with the requirement to maintain data processing records under the GDPR rules (see also Appendix III), and the data collected have been compiled in the following table.

Table 3: Categories of Data collected in the Test Sites.

Test Site	Categories of Data	Yes	No
Carinthia	[1]Personal Data		x
Graz	Personal Data		x
Gothenburg	Personal Data (i		x
Linköping	Personal Data/banking/financial data	x	
Madrid	Images (external cameras)	x	
Karlsruhe	Personal Data (interviews)	x	

[1] Subjective surveys are planned to be conducted in SHOW through Netigate. No personal data are stored. Still, during the conduct of interviews with stakeholders, it is not impossible that personal data is collected locally. Still, they will not be shared centrally as explained in D3.5.

Test Site	Categories of Data	Yes	No
Rouen	Personal Data Images(internal/external cameras)	x	
Brainport	Personal Data		x
Brno	Personal Data		x
Tampere	Personal Data		x
Trikala	Personal Data/ AV booking)	x	
Turin	Personal Data/ on-demand ride booking	x	
Salzburg	Personal Data		x

In the case of the test sites collecting personal data, as dictated by the project data collection requirements, the collected data will be anonymous and with no association enabler in order to retrieve them.

In some sites, e.g. the Linköping site, the collected banking and other financial information, needed for payment/invoicing services, are stored locally by the respective handling entities and are not shared with any other entity.

In the sites where subjective data are collected during the field trials and in which personal data may be contained within, these will be stored anonymously, i.e. they will not be associated with contact details or any other information that may infer associations revealing traveller identities. The Netigate tool is a dedicated framework for the capturing of the personalized data. This data is going to be processed and to be stored within the SHOW DMP. As a next step, the respective KPIs that are relevant to the personalized data are going to be calculated following the predefined procedure for the rest of KPIs. The Netigate tool guarantees anonymization which is mandatory in the whole SHOW ecosystem. Encryption and authentication will also be implemented for privileged access.

Regarding images/video streaming data, from vehicle and infrastructure cameras/surveillance systems, these data are recorded according to the national law and for traffic safety analyses and risk assessment scope exclusively and in some cases are deleted after some hours. Acknowledgment of video surveillance data is carried in the vehicles to be visible by passengers, safety drivers, etc. In no case, any type of this data is shared with SHOW. Last but not least, apart from the activities held on test sites, personal data (name, contact details, company/organization, emails) may be generated in other types of activities in the project; examples of such cases are mentioned below:

- Surveys e.g., the A3.3 survey on local regulation of shared CCAM
- Local focus groups and dissemination events that have or will be convened in the course of the project

- Subscription to the project newsletter and the project Stakeholders Forum

The collected personal data is only used in each case to get in touch with the respondents in order to serve the specific scope in each case and is not shared with any other than the controlling - in each case - SHOW entity.

Finally, in all cases, participants will be informed that their data will be kept entirely confidential and that their anonymity will be protected and their rights as individuals will be preserved, in accordance with the project data protection policy as described in D3.5 (and reflected also in the project web site at: https://show-project.eu/disclaimer-policy/) and the GDPR rules

3 FAIR Data Principles

The data generated within the project framework should be Findable, Accessible, Interoperable and Reusable (FAIR). These qualities neither affect implementation choices nor impose specific technologies, standards, or implementations. FAIR data principles, therefore, should not be viewed as a standard but rather as a framework to follow, when designing a Data Management Plan, since they ensure that the most important components for lifecycle data management are covered.

3.1 Making Data Findable

Data to be made openly available in SHOW, according to the principle of making data findable, special care should be applied to prevent the storage of data for open access, unless appropriate metadata will be provided.

Internal data can be considered as the data that is generated from the sensors. External data can be thought as the data that are shared in SHOW from third parties such as weather description, traffic and maps. Both of these categories of data will be continually collected and registered, with their appropriate metadata according to the relevant EC standards, mentioned in section 3.3, thus making data visible in a searchable context based on the semantics of data.

3.2 Making data openly Accessible

Part of data in SHOW will be made available with open access. This will not apply to the datasets that will be deemed by the Consortium as IPR protected such as proprietary data owned by consortium beneficiaries or third legal parties, as outlined in the DoA, and also in Deliverables D3.2 and D3.5 - Ethics manual & Data Protection Policy and Data Privacy Impact Assessment (DPIA) and the project Consortium Agreement.

Especially for generated data or for data declared as background assets, access will be granted only to authorised users (Appendix I – ref Article 29.3).

The SHOW's data portal, where all relevant internal and external data assets will be continually collected and registered, will be the common means for exchanging data either among the project partners or between the consortium and third parties.

All datasets will be available for sharing and re-use via the Data Management Portal of the project. Restricted and privileged access management policies have been taken in place in the system. Each entity that is relevant to the impact assessment or the implementation of services could obtain access to the specific SHOW resources. The administrators of SHOW DMP are the only responsible party to give the proper access.

All project datasets that will be agreed to be shared will be stored in a private cloud-based repository, ZENODO (https://zenodo.org/) a free service developed by CERN under the EU FP7 project OpenAIREplus (grant agreement no.283595), in a similar address as the following link: https://zenodo.org/communities/show-h2020/.

In summary, as a baseline, SHOW partners shall deposit:

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

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 datasets in the cloud repository will be linked to the management/exploitation portals of the project. They will be assigned to DOIs in order for third parties to be ableto access them. Through the use of the above repository (or similar, e.g. own-Cloud),we will ensure that the most up to date security features will be applied out of the box,e.g., firewall, password protection, encryption, etc. If any IPR issues exist on sharing, they will be handled accordingly. Moreover, all necessary material that supplements each dataset (software for parsing the datasets, standards documents, etc.) will be provided by the consortium via the data management portal

3.3 Making data Interoperable

Data assets collected in the Data management portal, are accompanied with their appropriate metadata according to relevant EC standards, such as DCAT-AP (D3WC initiative) and any extensions that may be done to it in the context of the ISA2 programme. This application profile is a specification for metadata records to meet thespecific application needs of data portals in Europe while providing semantic interoperability with other applications on the basis of reuse of established controlled vocabularies (e.g. EuroVoc) and mappings to existing metadata vocabularies (e.g. Dublin Core, etc.).

In the procedure of data model's definition, different protocols have been studied and taken into account. Some of them are:

- ETSI EN 302 637-2 [2]
- ETSI EN 302 637-3 [3]
- Transmodel
- NeTEx
- SIRI
- DATEX II, TPEG2 [4,5]

As a result, the SHOW Data Registry has been created in order all the required data attributes to be included and described totally. The adopted definition of the data attributes ensures the interoperability among the SHOW partners and the accurate and common calculation of KPIs.

The SHOW data management portal is based on existing opensource technology, so as to maximize interoperability with open data federators and to facilitate deployment and will become a relevant asset inside the project as well as to those outsider institutions that are willing to reuse the data assets gathered and produced by the project.

The provision of the data could be distinguished into two discrete categories; the real-time one that refers mainly to the data captured by the vehicles' data and the historical one. The real-time connection is established with the usage of the MQTT broker and is utilized simple JSON format messages. The aforementioned messages, except for the main information, also includes the metadata. The historical provision of the data takes place through the open-source CKAN platform. Dedicated formats have been circulated in order for the common terminology to be adopted by all the parts of the SHOW project.

3.4 Increase data Re-use

SHOW participates in the Pilot on Open Research Data launched by the European Commission along with the Horizon2020 Programme. As such, part of the data produced by the project is to be published with open access. Particularly, access will be given through Creative Commons CC0 license for all datasets at this project stage, unless itis defined otherwise.

All datasets will be maintained for the entire duration of the project After the project ends,

all datasets will be stored in a centralized facility to minimize maintenance costs. Datasets with acknowledged long-term value may be kept for a longer period. The long-term value of a dataset will be decided according to the exploitation plan as well as its relation to a scientific publication.

For the period in which the data will be available for open access, no restrictions will be imposed on their access. In case IPR protected or proprietary data will be generatedor made available as background assets in the context of the project by one or more beneficiaries, access to this data will be treated in accordance with the rules and regulations foreseen in the project Consortium Agreement. For all this data, access willbe granted only to authorised users.

All data to be available as open data will undergo a quality assurance process.

At the first level, each partner in charge follows specific procedures to assure the qualityof the data and conformance to standards as referred to in each dataset's description in Section 2. Those procedures regard calibration of the measurements, as well as, internal post-measurement review. On a higher level, the quality will be also monitored periodically on the different versions of the dataset as conservation experts of the consortium will review each update. Finally, additional peer-reviews will take place in case of publication in Journals.

4 SHOW Open Strategy & Participation in the Open Research Data Pilot

The SHOW 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. SHOW will ensure 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 following lists the minimum fields of metadata that should come with a SHOW project-generated scientific publication in the Information HUB section of the project website https://show-project.eu/

- The terms: "European Union (EU)", "Horizon 2020"
- Name of the action (Research and Innovation Action)
- Acronym and grant number (SHOW, 875530)
- Publication date
- Publication type
- Publication place
- Publication name
- Publication authors
- Length of embargo period if applicable
- Persistent identifier

Apart from the scientific publications index, SHOW will publish all its public Deliverables and dissemination material on its website (under "Library"). When referencing Open access data, SHOW 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 researchand innovation program under grant agreement No 875530".

Finally, SHOW will target **Open Access journals**, when applicable, whereas it will alsotarget Gold OA publications and Green OA, wherever Gold OA is not possible. The target is to maximize the impact on scientific excellence through result publication in open access yet highly appreciated journals, without releasing any confidential information that could potentially violate the security nature of the project.

The SHOW 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 also section 3.2).

5 Allocation of Resources

SHOW partners have to observe the policies set out in this DMP and datasets have to be created, managed and stored appropriately, to respond to the data management challenges in an efficient manner. The Data controller/processor responsibilities, alsoin reference to the Data Protection Officer (DPO), have been described in Deliverable *D3.5 – Final Ethics manual & Data Protection Policy and Data Privacy Impact Assessment (DPIA).*

DPIA's are conducted at each pilot site in collaboration with the data collectors and processors (respective templates have been annexed in Appendix III), the project DPO and the Data Manager. Any ethical treatments and data protection mechanisms regarding the main data clusters on pilot and project level are included in the aforementioned Deliverables as required in POPD – Requirement POPD – H – Requirement No. 1 (D18.1) and (POPD – Requirement No. 3 D18.2 and the Data Management Plan (D14.2 & D14.3).

It should be noted that regarding the ORDP, the data controller must ensure that data are shared and are easily available. Each data producer and WP leader is responsible for the integrity and compatibility of its data during the project's lifetime. The data produceris responsible for sharing its datasets through open access repositories. He is in charge of providing the latest version.

SHOW has endorsed an Open Access Strategy, and selected datasets will be shared through the Zenodo account that has been created for the project for this purpose (https://zenodo.org/communities/show-h2020/). As SHOW open data are hosted by an open free of charge platform (e.g. ZENODO), no additional costs are required for hosting the data.

The allocation of specific roles in the Consortium (Data Manager, Data Controller, Data Processor, Data subjects, Data Protection Officer) has been defined whenever and to the level required in compliance with the GDPR requirements. The Data Manager, Dr. Nikolaos Tsampieris (ERTICO), coordinates the actions related to data management, being responsible for the actual implementation of the Data Management Plan as laid done in D14.2 and D14.3 successive versions and the compliance to Open Research Data Pilot guidelines, and working closely with the Project Ethics Board (PEB), the Project Quality Board (PQB) and the Technical & Innovation Manager (CERTH/HIT).

6 Conclusions

Data management within the SHOW project is addressed in A14.6: Data Management on project level and on the basis of the key principles defined in A3.2: Ethical and privacy issues.

The current version D14.3 updates the initial version of the Data Management Plan which aimed at presenting the generalprinciples of DMP and a first data management analysis in anticipation of the respective procedures and infrastructures to be put in place in SHOW, to efficiently manage the generated and/or collected data.

Still, a more elaborate presentation of the data clusters of the project, recognising the personal data and also performing on this basis the project Data Privacy Impact Assessment (DPIA) can be found in D3.4 and *D3.5: SHOW updated & final Ethics manual & Data Protection Policy and Data Privacy Impact Assessment.*

The formal (GDPR compliant templates for data controllers' and processors' records) but also technical measures (cybersecurity mechanisms in WP4 and WP5) have been already put in force to ensure data protection in all aspects. The test sites data registry process that has been launched in WP5 will align with those principles and the data protection policy of the project (the latest version can be found in D3.5), as it has been described in *D5.1: SHOW Big Data Collection Platform and Data Management Portal.*

One of the aspects that have been clarified and reported in D3.4 and D3.5 is also the GDPR obligations and assigned contacts in each test site and whichever other entity is involved in the collection of data. All the respective roles have been identified and mapped to entities in D3.5.

Finally, the project has endorsed an Open Access Strategy, that will be enabled through publications in Open Access journals but also sharing of both publications but also, later in the project, selected data sets through the Zenodo account that has been created for the project for this purpose (https://zenodo.org/communities/show-h2020/).

The information from Data Privacy Impact Assessments (DPIA) which have been completed by the test sites, in response to the DPIA template circulated, the test sites feedback on the DPIA's as reported in D3.5 and the corresponding GDPR forms are part of the broader final data management policy of the project.

As Data Management is an effort spanning the entire project lifecycle, there will be continuous assessment of the data privacy status across the different activities of the project and across the different test sites for the field-specific operations.

References

[1] EUROPEAN COMMISSION Directorate-General for Research & Innovation H2020 Programme Guidelines on FAIR Data Management in Horizon 2020- Version 3.0 26 July 2016 http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf

[2] ETSI EN 302 637-2 V1.3.1 2014-09 - Intelligent Transport Systems (ITS); Vehicular Communications; Basic Set of Applications; Part2: Specification of Cooperative Awareness Basic Service

[3] ETSI EN 302 637-3 V1.2.1 2014-09 - Intelligent Transport Systems (ITS); Vehicular Communications; Basic Set of Applications; Part3: Specification of Decentralized Environmental Notification Basic Service

[4] DATEX II extension: Safety Related Messages ITSTF13004v3. 2013-12-13)

[5] Safety-related message sets – Selection of DATEX II Codes, TPEG2-TEC-Causes

[6] SHOW (2021). D4.1: Open modular system architecture and tools (November 2021 revision). Deliverable of the Horizon-2020 SHOW project, Grant Agreement No. 875530.

[7] SHOW (2021). D4.3 Open modular system architecture - second version: Deliverable of the Horizon-2020 SHOW project, Grant Agreement No. 875530.

Appendix I - Data Management Plan Context

Reference to the Grant Agreement 875530 SHOW

a. Article 29.3

29.3 Open access to research data

*Regarding the digital research data generated in the action ('**data**'), the beneficiaries must:*

(a) deposit in a research data repository and take measures to make it possible for third parties to access, mine, exploit, reproduce and disseminate — free of charge for any user — the following:

(i) the data, including associated metadata, needed to validate the results presented in scientific publications as soon as possible;

*(ii) other data, including associated metadata, as specified and within the deadlines laid down in the **'data management plan'** (see Annex 1);*

(b) provide information — via the repository — about tools and instruments at the disposal of the beneficiaries and necessary for validating the results (and — where possible — provide the tools and instruments themselves).

This does not change the obligation to protect results in Article 27, the confidentiality obligations in Article 36, the security obligations in Article 37 or the obligations to protect personal data in Article 39, all of which still apply.

As an exception, the beneficiaries do not have to ensure open access to specific parts of their research data if the achievement of the action's main objective, as described in Annex 1, would be jeopardised by making those specific parts of the research data openly accessible. In this case, the data management plan must contain the reasons for not giving access.

Appendix II - SHOW relevant standards

Standard to abide	For which work item and how we comply	Anticipated contribution (Yes/No)	How (if Yes)
ETSI EN 302 637-2 Intelligent Transport Systems (ITS); Veh+B43icular Communications; Basic Set of Applications;Part 2: Specification of Cooperative Awareness Basic Service	A4.2	Yes	Inputs for new messages and service standards
ETSI EN 302 637-3 Intelligent Transport Systems (ITS);Vehicular Communications; Basic Set of Applications;Part 3: Specifications of Decentralized Environmental Notification Basic Service	A4.2	Yes	Inputs for new messages and service standards
ETSI TS 103 301 Intelligent Transport Systems (ITS);Vehicular Communications; Basic Set of Applications;Facilities layer protocols and communication requirementsfor infrastructure services	A4.2	Yes	Inputs for new messages and service standards
ETSI TS 102 731 Security Architecture	A4.4 (Cyber Security view)	Tbd in the project Tbd in the project	
ETSI TS 102 940 ITS communication securityarchitecture and security management	A4.4 (Cyber Security view)		
ETSI TS 102 941 Trust and Privacy Management	A4.4 (Cyber Security view)		
ETSI TS 102 942 Access Control	A4.4 (Cyber Security view)		
ETSI TS 102 943 Confidentiality Services	A4.4 (Cyber Security view)		
ETSI TS 103 097 Security header and certificate format	A4.4 (Cyber Security view)		
SAE J2735 Dedicated Short Range Communications(DSRC) Message Set Dictionary	A4.4 (Cyber Security view)		
IEEE 1609.2 Security for C-ITS	A4.4 (Cyber Security view)		
CEN TS13149 : onboard data communication	A4.2	Yes	Proposal of new services
CEN EN 12896 : Transmodel - Reference Data Modelfor Public Transport	A4.2	Yes	Proposal of data model update
CEN EN/TS 15531: SIRI - Service Interface for Realtime Information	A4.2	No	
CEN TS Exchange 16614: Ne-TEx Network Timetable Exchange	A4.2	No	
ETSI EN 302 895 - Local Dynamic Map (LDM)	A4.2	No	

Appendix III - Data Controller & Data Processor Record of processing activities

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. All data controllers and processors must also keep records of data set descriptions according to the latest Data Management Plan and DPIA. Where applicable, this information must be verified by the organizational Data Protection Officer.

I. Data controller's record of processing activities

1	**Contact details of Data Controller**
Email	
Company address	
Telephone	
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**

Where possible, the envisaged time limits for erasure of the different categories of data
Where possible, a general description of the technical and organisational security measures for
the pseudonymisation and encryption of personal data
the ability to ensure the ongoing confidentiality, integrity, availability andresilience of processing systems and services;
the ability to restore the availability and access to personal data in a timelymanner in the event of a physical or technical incident
a process for regularly testing, assessing and evaluating the effectiveness oftechnical and organisational measures for ensuring the security of the processing;

II. Data processor's record of processing activities

Contact details of Data Processor	
Email	
Company address	
Telephone	
Categories of processing carried out on behalf of the Controller	
Where applicable, transfers of personal data to a third country or an international organisation, including the identification of that third country or international organisation	

Where possible, a general description of the technical and organisational security measures for
the pseudonymisation and encryption of personal data;
the ability to ensure the ongoing confidentiality, integrity, availability andresilience of processing systems and services;
the ability to restore the availability and access to personal data in a timelymanner in the event of a physical or technical incident
a process for regularly testing, assessing and evaluating the effectiveness oftechnical and organisational measures for ensuring the security of the processing;

Appendix IV. Excerpt of (current) Data list for SHOW relating to Personal Data

Data Type	Attribute Name	Attribute Value Type	Attribute Description	Standard	Personal Data
Dynamic	Wheelchair on board	Integer	Current number of wheelchairs in the vehicle	SIRI	Yes
Dynamic	Passengers with special needs	Integer	Current number of passengers with special needs in the vehicle	SIRI	Yes
Booking-ride	Desired pickup location	Float	Initial location of a passenger	SIRI	Yes
Booking-ride	Desired pickup time	Time	Correspond time	SIRI	Yes
Booking-ride	Desired drop off location	Float	Destination of a passenger	SIRI	Yes
Booking-ride	Desired drop off time	Time	Correspond time	SIRI	Yes
Booking-ride	Planned pickup location	Float	The location in which the passenger could board calculated by the system	SIRI	Yes
Booking-ride	Planned pickup time	Time	Correspond time	SIRI	Yes
Booking-ride	Planned drop off location	Float	The location in which the passenger could debark calculated by the system	SIRI	Yes
Booking-ride	Planned drop off time	Time	Correspond time	SIRI	Yes
Booking-ride	Actual pickup location	Float	The location in which the passenger actually boards.	SIRI	Yes
Booking-ride	Actual pickup time	Time	Correspond time	SIRI	Yes
Booking-ride	Actual drop off location	Float	The location in which the passenger actually debarks.	SIRI	Yes
Booking-ride	Actual drop off time	Time	Correspond time	SIRI	Yes
Booking-ride	Planned booking route	GeoJSON	Planned vehicle route between pickup and drop off location.	SIRI	Yes

Data Type	Attribute Name	Attribute Value Type	Attribute Description	Standard	Personal Data
Booking-ride	Actual booking route	GeoJSON	Actual vehicle route between pickup and drop off location.	SIRI	Yes
Booking-ride	Direct-ride distance	Float	Length of the fastest direct route between pickup and drop off location	SIRI	Yes
Booking-ride	Direct-ride duration	Float	Duration of the fastest direct route between pickup and drop off location.	SIRI	Yes
Booking-ride	Actual ride distance	Float	Length of the actual route between the actual pickup location and the actual drop off location	SIRI	Yes
Booking-ride	Actual ride duration	Float	Duration between the actual pickup time and the actual drop off time	SIRI	Yes
Booking-ride	Trip reason	String	Work/shopping/ school and so on	SIRI	Yes
Infra-structure	Video-internal cameras		Cameras installed in the vehicle		Yes
Infra-structure	Video-external cameras		Cameras installed out of the vehicle		Yes
Infra-structure	Magnetic loops		It calculates the inferences regarding the traffic condition		Yes
Infra-structure	Sensors for capturing wireless internet traffic		Metadata showing the traffic between the vehicle's sensors and the cloud servers		Yes
Infra-structure	Camera installed on traffic lights or bridge		External camera on the vehicle for counting outside vehicles		Yes

Data Type	Attribute Name	Attribute Value Type	Attribute Description	Standard	Personal Data
Infra-structure	Traffic in Vehi-cle's route		Camera in the Vehicle for traffic situation		Yes
Infra-structure	RadioFrequen-cy sensor		Bluetooth detections		Yes
Infra-structure	Vehicle traffic camera		Camera installed on the vehicle		Yes
Other	Bluetooth Sensor		Probe data, Social media data	SIRI	Yes
Other	Network traffic metadata				Yes