Needs, wants and behaviour of "Drivers" and automated vehicles users today and into the future

Contract No: 815001

D9.4: DATA MANAGEMENT PLAN

Version 1.0

Work package	WP9: Project Management
Activity	A9.3
Deliverable	
Authors	Maria Panou, Evangelia Gaitanidou, Evangelos Bekiaris (CERTH)
Status	Final
Version	1.0
Dissemination Level	ORDP
Document date	24/10/19
Delivery due date	31/10/19
Actual delivery date	1/11/19
Reviewers	DEUSTO/ NTUA

This project has received funding from the European Union's Horizon 2020 Research and Innovation Programme under grant agreement no 815001.

Version History

Document history			
Version	**Date**	**Modified by**	**Comments**
0.1	27/09/19	M. Panou	Table of Contents was created.
0.2	10/10/19	K. Touliou	DPIA templates results consolidated in one table.
0.3	11/10/19	E. Gaitanidou	Chapters 1, 2 and 3 were added.
0.4	21/10/19	M. Panou	Chapters 4, 5 and 6 and annexes were added.
1.0	24/10/19	K. Touliou	Submitted for internal peer review.
Final	1/11/19	E. Gaitanidou	Incorporated feedback from reviewers.

Legal Disclaimer

This document reflects only the views of the author(s). Neither the Innovation and Networks Executive Agency (INEA) nor the European Commission is in any way responsible for any use that may be made of the information it contains.

Table of Contents

List of Tables

Abbreviations List

Abbreviation	Definition
Art.	Article
AV	Autonomous Vehicle
CEA	Cost-Effective Analysis
CSV	Comma-Separated Values
D	Deliverable
DMP	Data Management Plan
DOI	Digital Object Identifier
DPA	Data Protection Authority
DPIA	Data Privacy Impact Assessment
DPO	Data Protection Officer
FAIR	Findable, Accessible, Interoperable and Reusable
GDPR	General Data Privacy Regulation
GPS	Global Positioning System
HMI	Human Machine Interface
h/w	Hardware
ID	Identifier
IDF	International DOI Foundation
M	Month
MCA	Multi – Criteria Analysis
MOB	Mobility Observation Box
ORDP	Open Research Data Pilot
PD	Private Data
PM	Person-Month
s/w	Software
TBD	To Be Decided
TLA	Traffic Lights Assistant
URL	Uniform Resource Locator
VR	Virtual Reality
WP	Work Package
3D	Three-Dimensional

Executive Summary

This report describes the Initial Data Management Plan (DMP) for the Drive2theFuture project. The purpose of the DMP is to set out the main dimensions and characteristics of the project's data management policy for the datasets generated by the project following the regulations of the Pilot action on Open Access to Research Data of Horizon 2020. As D10.2 'H-Requirement no. 4' is the Ethics policy, this Deliverable comprises the Data Privacy Policy of the project.

The first step was to identify the data clusters and specifications that will be collected during the lifetime of the project across the project workplan where data collection or processing is anticipated; namely, these are WP1, WP2, WP5, WP6 and WP8. Identification of data entails elaboration on data and metadata types, descriptions, standards, file formats, programmes used to create these files, embargos and temporal restrictions as well as any other potential confidentiality restriction). The primary sources of data are related to WP5 pilots conducted in 12 pilot sites. As Pilot plans (D5.1) and other related Deliverables from the other WPs are still being developed the time this Deliverable is submitted, only high-level descriptions of these data types (**Annex 2**) are available. Certain data characteristics will be collected and finalized in the next and final version of the Data Management Plan (D9.7; M18).

These categories will serve as the basis for creating the Drive2theFuture data management repositories and the metadata descriptions to accompany any dataset to be shared through or linked to any platform and/ or systems implemented in the project.

In this version the following elements of the Data Management Plan framework are set:

- Purpose of document, intended audience and interrelations **(Chapter 1)**
- Data clusters, specifications, FAIR and Open Data principles **(Chapter 2)**
- Data access, storing, archiving, ownership, sharing, re-using, and security **(Chapter 3)**
- GDPR roles **(Chapter 4)**
- GDPR obligations **(Chapter 5)**
- Data Privacy Impact Assessment (DPIA) process and documentation **(Chapter 6)**
- Conclusion and next steps **(Chapter 7)**

This Deliverable includes 5 Annexes. Other relevant EC guidelines and legislation have been added in **Annex 1**. Guidelines and information on data storage, back, documentation and naming can be found in **Annex 2**, while the initial data clusters and sources are presented in a consolidated table in **Annex 3**. **Annex 4** presents the initial Privacy Disclaimer of the Drive2theFuture accessed through the project's website. Last, a template for controllers and processors is included in **Annex 5**, which each identified Partners who has such a role in the project is required to fill in.

Post-processed datasets, free from any private/personal and identifiable information, will reside in the Drive2theFuture data repositories which will be described in the next revised version of this deliverable. It will also include analytic descriptions of the complete and (non) shareable datasets that will be created during the pilots (WP5), the analyses to follow (WP5 and WP6) and other WPs that require post-processing (e.g. for the preparation of the Use Cases in WP1 and the simulation modelling conducted in WP2). In addition, the final Data Management Plan will contain the complete structure of the database, descriptions of the metadata files to enable self-explainable (re)use of datasets by external parties. Long-term re-usability of these data is of substantial importance, especially in the field of automated driving experience. Embargos (if any) for parts/segments of data, models, evaluation dimensions (e.g. acceptance, trust), the surrogate and horizontal impact and metadata indicators/estimators will be set by the Partners who own these in collaboration with the Data

Manager of the project. Additionally, the final location and format of open Drive2theFuture datasets will be defined. Finally, the final DPIA report will be annexed in the final version Data Management Plan Deliverable (D9.7).

In conclusion, this Deliverable is a living document and it will be regularly updated during the lifetime of the project. If official updates are necessary, then they will be included in the final version of the Project Management Plan (D9.8; M35).

1 Introduction

1.1 Purpose of the Document

Drive2theFuture aims to prepare "drivers", travellers and vehicle operators of the future to accept and use connected, cooperative and automated transport modes and the industry of these technologies to understand and meet their needs and wants.

1.2 Intended audience

Communication of research outputs and metadata increases the potentials for inter-disciplinary collaboration among researchers in the automated driving area. Data sharing enables researchers to run synthetic, comparative and validation studies and, therefore, further research. Current research advocates that when data are shared, research productivity and the number of publications increases compared to when the data are not shared within a community or with others [1]. The data and metadata gathered may be useful to researchers, representatives of the industry, who might not be directly involved in research but are interesting in AV acceptance and training; where they can search and investigate indicators to assist their empirical studies and inferences.

This deliverable aims to present the data sources and types to be evaluated and tested during data collection activities along with the standards and guidelines followed in order to store and communicate the findings (a preliminary compilation can be found in Annex 3). As such, Deliverable D9.4 encompasses the dimensions of the Drive2theFuture DMP and the methods followed to address them within the project. It additionally lays the foundations for data categorization, specification and descriptions. Large data sets will be gathered and stored according to national and European legislation frameworks and standards. The guiding principle remains the new Horizon 2020 effort to create re-usable datasets for furthering sustainable, comparable, and growingly valid and reliable research outcomes.

Data management within the project is:

- **manual** (entering formative and subjective data gathered from questionnaires, interviews, surveys): these data will be stored locally, and their collection is based on the data templates and evaluation material created for data collection across pilot sites and partners (mostly relevant to WP5, WP6 and WP8 activities).
- **automatic** (through the Drive2theFuture systems, sensors, social media, online surveys and polls, etc.): pre-processing and user clustering is occurring at the database and can be used to create the metadata required for the creation and training of the algorithms and the compilation of all pre-defined and agreed indicators per scenario / task, vehicle and Use Case (mostly relevant to WP1, WP2 and WP5 activities).

There are **three treatment layers** of data management:

- **Locally (*Raw*)**- data stored at each pilot site which may (or may not) be shared with other sites; these data are used for recruitment and identification of participants (i.e., prerequisite only for arranging follow-up sessions). These data are stored strictly at pilot site location and they are additionally stored securely. They are separated from any data that will be aggregated or consolidated for drawing inferences and used for reporting results in related WP5 and WP6 Deliverables. At this point, it is not certain if pilot data collection will be pseudonymised or even if it will be a necessity for all sites (i.e. certain data sets might be inherently anonymous and pseudonymisation will not be required). Data will most probably be anonymously collected but as such decision has not been made yet, they are assumed to be at least pseudonymized.

- **Data control and management of collected and aggregated data (and later metadata; *Processed*)** - used for common data types' analyses that are categorised, anonymised, harmonised and coherent. High level data are entered only in English to ensure comprehension. At this early stage, data from pilots will be labelled as pseudonymized to ensure that all possibilities are addressed and as such to be treated by responsible partners.
- **Drive2theFuture big data analysis including sentiments from social media (WP2; *Raw, processed, annotated and potentially other treatments*)** - data collected only from the project and may relate to any type of addressed user, the centre mechanisms (i.e. metrics of use), user clustering, user profiling as well as analytics. This conceptual framework is still under discussion. The conceptual description of the repository will be included at the next version of the deliverable (M18), after the finalisation of the specifications and mechanisms are in place (WP2).

It is evident that Drive2theFuture will collect very diverse data and there will be a fair representation of these types across the WPs and pilot sites (e.g. sensor data, notification to various user groups, videos capturing interactions between users and in-vehicle HMI and systems, completion of digitised questionnaires). Consequently, not following a traditional centralised data management model will not be efficient and applicable. The plethora in data sources shows that at this stage we can only propose data to be collected (Annex 3). Later these data will be specified and then we will be able to cluster them in an upper level and create metadata to support work to be conducted primarily within WP5 and WP6. We imagine the higher-level aggregation can more effectively support efforts to share and communicate results to policy makers and other related projects and organizations.

The final update will include a refined and elaborate account of data gathered throughout the lifetime of the project and their "integration metadata" that might result from the integration of different systems to one platform as well as final indicators and labels fed to behavioural modelling, if we deem important to do so. Furthermore, a final version of this report will include the characteristics of the data and surrogate variables, their storage properties and the parts that can be communicated to public and shared with other research communities.

1.3 Interrelations

During the lifetime of the project, data will be collected through online surveys to shape the Use Cases (WP1), data will be used to create the behavioural models in combination with technology acceptance models (WP2), data will be collected in iterative pilot and demonstration activities from drivers, passengers, operators, and experts (WP5, WP6) as well as during users' training through various tools (e.g. 3D automated scenarios for VR-goggles, web applications and social media platforms) (WP4, WP5, WP6). Finally, data will be collected for CEA and MCA (WP8) purposes and potentially through the project's website and other dissemination events.

2. Drive2theFuture data

For Drive2theFuture to achieve its mission and to meet its objectives, a series of data, including personal data, is required to be collected, processed, used and managed. Data collection and processing in Drive2theFuture adheres to the respective European regulations, encompassing General Data Privacy Regulation (GDPR) [2] (other relevant regulations can be found in Annex 1).

2.1. Clusters of data in Drive2theFuture

The key data clusters are as follows:

1. ***Survey data collected online***
   1a. Survey data about user needs and requirements to satisfy the requirements of WP1 in order to prepare the Use Cases. Additionally, survey data will be fed to the WP2 acceptance models;
   1b. Potential workshop and events data collected (for example, through Mentimeter) to satisfy dissemination, demonstration and/ or training activities.
2. ***Sensor and system data***
   - *Infrastructure sensors' data;*
   - *Vehicle data;*
   - *Driving performance data;*
   - *Observation data (based on video recordings);*
   - *Wearable data;*
   - *Sentiment analysis data from social media (e.g. Twitter account).*
3. ***Qualitative data (questionnaire, focus groups, workshops, events, etc.; offline collection)***
   - *Collected during face-to-face pilot session activities;*
   - *Collected during focus groups, CEA and or MCA activities.*

2.2. Data specifications template

This chapter provides a template to be used for describing the datasets to be produced or collected in Drive2theFuture project. As the nature and extent of the datasets can evolve during the project, changes in the template may occur. For this reason, the current completed template (Annex 3) is slightly different from the final one (Table 1) because the data descriptions are still in progress. These fields will be additionally addressed in the final Data Privacy Impact Assessment (DPIA). Still, those will be revised, and the missing fields will be completed until the end of the project and before the sharing of the data in the context of the Open Research Data Pilot (ORDP). In particular, the fields that are currently missing are namely the Standards and metadata, the Data Sharing, the Re-used existing data and the Data Utility fields.

Table 1. Data specifications template

Data characteristic	Description
Dataset Reference	Drive2theFuture_WPX_AX.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 Activity in which it is generated and its version (for example: Drive2theFuture_WP5_A5.1_01).
Dataset Name	Name of the dataset.

Data characteristic	Description
Dataset Description	Each dataset will have a full data description explaining the data type, 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 Origin	Specify the origin of the data.
Data Size	State the expected size of the data.
Data Sharing	Explanation of the sharing policies related to the dataset between the next options: • **Open**: Open for public use. • **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. Inform if the dataset entails personal data and how this issue is considered.
Archiving and Preservation	The preservation guarantee and the data storage during and after the project (for example: databases, institutional repositories, public repositories, etc.)
Re-used existing data	Y/N. If Yes, state the re-used data and how/from where they were retrieved.
Data Utility	Outline to whom the dataset could be useful – potential secondary users.
Link to Dataset	URL link to actual dataset with the same filename (if **Open**)

2.3. FAIR data

Drive2theFuture project will in principle participate in the Open Research Data Pilot (ORDP) but data marked as “restricted” or under an “embargo” period (see the dataset description above) will be excluded. To this end, the data that will be generated during the project and will be included in ORDP should be ‘FAIR’, that is findable, accessible, interoperable and reusable. These requirements do not affect implementation choices or 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 Drive2theFuture public repository (ZENODO; please see below).
- The reference used for the dataset will follow this format: Drive2theFuture_WPX_AX.X_XX, including clear indication of the related WP, activity and version of the dataset.
- The standards for metadata will be defined in the "Standards and metadata" section of the dataset description table (see the current version of the template in the previous section).

Making data openly accessible

- Datasets openly available are marked as "Open" in the "Data Sharing" section of the dataset description table (see Table 1).
- The repository that each dataset is stored, including Open access datasets, is mentioned in the "Archiving and Preservation" section of the dataset description table (see Table 1). ZENODO will be one of the considered options.
- "Data sharing" section of the dataset description table (see Table 1) will also include information with respect to the methods or software used to access the data of each dataset.
- Data and their associated metadata will be deposed either in a public repository or in an institutional repository.
- "Data sharing" section of the dataset description table (see Table 1) will outline the rules to access the data if restrictions exist.

Making data interoperable

- Metadata vocabularies, standards and methodologies will depend on the repository to be hosted (incl. public, institutional, etc.) and will be provided in the "Standards and metadata" section of the dataset description table (see Table 1).

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 final version of the DMP.
- "Data Sharing" section of the dataset description table (see Table 1) is the field where the data sharing policy of each dataset is defined. Data available for use are subsequently available for future re-use. If any constrains exist, an "embargo period" will be set or "restricting flag" will be explicitly raised in this section of Table 1.
- The data producers will make their data available for third parties within public repositories only for scientific publications validation purposes.

2.4. Open Access principle

Drive2theFuture Consortium has agreed to follow an "open access" approach (as much as possible depending on the specific data type) following the respective Horizon 2020 guidelines [3] to ensure that the results of the project provide the greatest possible impact. Drive2theFuture will ensure the open access[1] to all peer-reviewed scientific publications and Deliverables relating to its results and will provide access to the research data needed to validate the results presented in deposited scientific publications. Publications and research data made available to third parties will not contain any personal information.

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

The following lists the minimum metadata fields that should label any Drive2theFuture 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 (Drive2theFuture, 815001)
- Publication date
- Length of embargo period if applicable
- Persistent identifier

When referencing Open access data, Drive2theFuture will include as a minimum the following statement demonstrating EU support (with relevant information included into the repository metadata):

“Drive2theFuture is funded by the European Union within Horizon 2020 research and innovation programme under grant agreement No 815001”.

The Drive2theFuture Consortium will strive to open as many datasets as it is feasible and agreed by its Consortium members and data owners. When this is not the case, it will be justified with the note in the data sharing field of this dataset stating the reasons for the applied restrictions.

Drive2theFuture Consortium will make public Deliverables and publications available with Open access in ZENODO[2], which is a free service developed by CERN under the EU FP7 project OpenAIREplus (Grant Agreement no.283595), under a dedicated account for Drive2theFuture. Under the same account, all the research derived datasets that will emerge in the project and will be decided to be **Open** for sharing by the Consortium will be shared. By the end of the project, this process will be completed. If data are shared, they will be in an easily accessible and re-usable format (e.g. csv), wherever this is feasible and possible to ensure that the data are accessed with open-access tools and s/w.

The specific repositories where Drive2theFuture datasets will be held during and after the project, they will be noted in the “Archiving and Preservation” field of the dataset. In cases where the project partners maintain additional institutional repositories, these will be also listed in the final DMP version.

In summary, as a baseline Drive2theFuture partners shall deposit:

- Scientific publications – in ZENODO dedicated repository of the project, on their respective institute repositories (when relevant) as well as in the Library of the project web site.
- Research data – in ZENODO dedicated repository of the project.
- Other public project output files (i.e. Deliverables) – in ZENODO dedicated repository of the project and the project web site.

[2] *https://zenodo.org/*

3. Data sharing and access

At each pilot site a nominated person will be responsible for overseeing that data are safe and secure. For organizations that are not designated pilot sites, a person will be nominated as well. Any other relevant EC guidelines can be found in Annex 1. Further information about relevant legislation and guidelines can be found also in D10.2 [3]. In addition, guidelines on storing, backing up and naming data files can be found in Annex 2. These sections accompanied by the guidelines in Annex 2, act as an extension on the technical and organization measures that are listed in D10.2 'H - Requirement No. 4; Section 5.3 and further elaborated in this Chapter.

3.1. Data access

One person will have **access** to full datasets (i.e. higher authorisation level) and the rest of the data team will have medium or lower level of authorisation. Authorisation is granted based on existing internal protocols. Data will be stored in secure areas (physical, network, cloud-based). Higher level of authorisation is granted only for sensitive and personal data. Data to be shared for analysis will not include any personal or identification data. These data, of course, cannot be shared with external databases for further (re-)use. For example, Dropbox cannot be used for this purpose.

Data collection, storing, accessing, and sharing abide to the international legislation (Annex 1) and guidelines (see D10.2 "H-Ethics Requirement No. 4" for an in-depth account).

Different levels of authorisation will exist also for remotely accessing data. High level access to data will not be possible outside the work premises, as they are defined at each pilot site.

Use of cloud store data will be available for medium and lower level of access. Not all individuals will have the same access privileges in order to avoid data corruption, loss and damage. Dataset owners will have full access (read, write, update, delete), however, individuals who want to use/reuse the dataset will be able to read and download but not make any changes or modifications to the specific dataset. Of course, all datasets will be password-protected. In some cases, encryption will be necessary.

The main restrictions with regards to confidentiality are the following:

- *Name*
- *GPS coordinates (only metadata or surrogates)*
- *Raw video and audio recordings*
- *Gender and age (in relation to any of the above)*

These data are identified based on the initial data pools set by partners responsible for data-oriented or data-based project activities and systems. Other data restrictions might arise during the project.

3.2. Data ownership

Any data gathered during the lifetime of the project are under the ownership of the beneficiary or the beneficiaries (joint ownership) that produce them according to subsection 3, Art. 26 of the signed Grant Agreement. The beneficiaries have the intellectual property rights of the data they collect and re-use of data is defined by the limitation they might set in how they will make data available. This means partners decide if they make their data freely and openly available to the research community (no additional restrictions on access to data or publications) or there is an embargo period, whereby permission for accessing the data is given after a certain period of time. As datasets have not been formed yet as well as indicators per pilot type and site, therefore this information will be available in the updated version of this deliverable.

3.3. Data Sharing and Re-use

Under Horizon 2020, all publication resulting by work performed within a project must be in open-access journals. Participating in an open scholar community can assist in making the work of partners, and the project, more visible to researcher working in similar disciplines and research areas. Specifically, for Drive2theFuture, publishing in open-access journals is sought. Relevant dissemination activities target, organize and manage publishing efforts (WP8).

Data re-use by external researchers and other stakeholder groups will be feasible for selected datasets. The embargo period will be at least the duration of the project, as partners would like to easily manage the data whilst collection, analysis and reporting is ongoing. Sharing and-reuse will be applied in the central database according to data depositor's preferences and suggestions.

3.4. Data/meta-data repository (Drive2theFuture databases)

Each partner creating the dataset will also be responsible for their upload to the central database to allocated space. An agreement will be reached between the administrator and the data responsible about the level of sharing between the partner (depositor) and the administrator (and the rest of the Consortium), defining the terms and conditions of use for the specific dataset. Embargo is set by the owners.

The database's requirements and specifications (e.g. s/w, standards, guidelines, etc.) are still to be decided. First, it is essential to identify and categorise any inherent restrictions in datasets because of the services or applications they are generated from. In addition, indirect restrictions might apply related to s/w developed for work to be carried out by a specific partner (e.g. in-house s/w, tool, etc.). There might be license or operating restrictions that should be considered (e.g. compatibility and access only with certain h/w or s/w).

3.5. Data Preservation and Archiving

Data will be preserved in the database after the end of the project (for a period of two years) only for complete datasets that partners have agreed to share with other researchers. Datasets could be linked with the European Union Open Data Portal after the end of the project (https://open-data.europa.eu/en/linked-data). Representative keywords will be selected by dataset owners to accommodate for future searches. Data owners will decide the duration of the data retainment period, which will be reported in D9.7.

Decisive factors are different per system and partner. Any related costs for archiving and preserving data -especially for long periods of time- will be checked for their justification and then incurred during the lifetime of the project. Drive2theFuture participates in the Open Research Data Pilot of Horizon 2020 and, thus, one more iteration of the Data Management Plan (DMP) is scheduled in Drive2theFuture project. This obligation is an innovative step compared to previous framework programmes and it presents the processes and data types/ categories undertaken in the project. Additionally, it ensures that we have properly clustered the data collected during the project in a way that will potentially increase the availability and visibility of gathered data. Sharing data and publications can potentially impact and enhance future collaboration of researchers in the same or affiliated areas aiming at a common target; to create reliable, replicable, and transparent data that will further advance similar research initiatives.

The Data Management Plan guarantees systematic data maintenance in a harmonised and coherent fashion and, hence, enables joined publications within the Consortium. Within the project, data will be manipulated in accordance to DMP guidelines, national laws and legislation and the project's ethics policy (D10.2).

3.6. Data Security

Drive2theFuture will provide out-of-the-box security mechanisms and management procedures so as to a) ensure personal (sensitive) data protection through a strict process of data collection, anonymization, harmonization and integration and b) guarantee data integrity and reliability, ensuring system's high performance operation through the exchange of the necessary information.

The Consortium research partners will always fully comply with all applicable data protection legislation and regulation during this project, to ensure the security and protection of individuals' personal information in relation to this project. The Consortium and research partners acknowledge the various new obligations and the new rights granted to data subjects under the GDPR and are aware of the significant fines that may be imposed should a data breach occur.

In terms of **personal data protection**, personal data will be anonymised and strictly used for project's purposes. Before collecting any personal data, the Local Ethics Representative (see D10.2) will be responsible for informing the involved pilot users/participants and collecting their informed consents (templates annexed in D10.2) that data will be maintained and stored based on the Grant Agreement rules and European/National legislation. No personal data will be centrally stored, without anonymization or pseudonymisation. No personal information will be made available by the Local Ethics Representative to the pilot sites, i.e., Drive2theFuture partners participating in the pilots. Only one person per site (the Local Ethics Representative) will have access to the informed consent form containing the personal information and only that person will be aware of the relation between the participant's unique identifier code and their personal identity, in order to administer the tests. In practice, the Local Ethics Representative will collect those data required for contacting the participants and arranging with them the sequence of the current or future tests. The Local Ethics Representative will then issue a single Test ID (unique identifier code) for each of them. This person (Local Ethics Representative) will not participate in the evaluation and will not know how each user behaved. One month before the end of the project, this reference, i.e., the reference between the Test ID and the real-life contact details of the participant, together with any other personal information held on the participant will be deleted, thus safeguarding full anonymization of the results.

The stored data might refer to participant's age and gender, but this information will be safeguarded, stored and processed only in accordance with all applicable data protection laws and regulations. The stored data will not contain any other identifier apart from the Test ID. In no circumstances will a participant be asked for information relating to their beliefs, political or sexual preferences. User-related data will be securely and safely stored. Furthermore, data will be scrambled where possible and abstracted to permit its use to achieve project outcomes while ensuring data integrity and security.

Any party which provides any data or information (the "Providing Party") to another party (the "Receiving Party") in connection with the project will not include any personal information relating to an identified or identifiable natural person or data subject. To this end, the Providing Party will anonymise or pseudonymise all data delivered to other parties to an extent sufficient to ensure that a person without prior knowledge of the original data and its collection cannot, from the anonymised or pseudonymised data and any other available information, deduce the personal identity of participants.

Each party shall be solely responsible for the selection of specific database vendors/data collectors/data providers, and for the performance (including any breach) of its contracts between the party and such database vendors/data collectors, (to which no other project partner shall be a party, and under which no other partner assumes any obligation or liability) and shall further warrant that it has the authority to disclose the information, if any, which it provides to the other parties, and that where legally required and relevant, it has obtained appropriate informed consents from all the individuals involved.

Partners supplying special data analysis tools, shall have the right on written notice and without liability to terminate the license that it has granted these tools to be used during the duration of the project, if the supplying partner knows or has reasonable cause to believe that the processing of particular data through such tools infringes the rights (including without limitation privacy, publicity, reputation and intellectual property rights) of any third party, including of any individual.

It must be pointed out that the data collection legal requirements depend on both EU and national law abidance. The EC Directives and International Agreements annexed in D10.2 are applied in Drive2theFuture. Compliance is monitored by the Ethics Board, the Coordinator and the Technical Manager of the project.

Each pilot site will have its own Ethics Committee and one person will be nominated per site as responsible for following the project's Ethics Board recommendations and data protection (D10.1).

4. Data treatment and management roles

4.1. Assignment of GDPR roles in Drive2theFuture

According to GDPR principles, the following roles and responsibilities are identified.

1. **Data manager** is the natural or legal person that coordinates the actions related to data management, is responsible for the actual implementation of the DMP successive versions and for the compliance to Open Research Data Pilot guidelines. In Drive2theFuture, this role is undertaken by **Dr. Maria Panou (CERTH)**, who is leading the DMP Deliverables in the project (D9.4 and its update) and the **Data Protection Officer (DPO)** liaison between the project and the Ethics Board (A9.3), which is DEUSTO (A5.8 "Pilot Results Consolidation" leader).
2. **Data controller** is the natural or legal person, public authority, agency or other body which, alone or jointly with others, who determines the purposes and means of the processing of personal data. In Drive2theFuture, this role will be undertaken separately and not centrally as each pilot site is a separate pilot centre with different plans, data collected, and results gathered. For online surveys, the partner who controls the data on the respective platform (e.g. CERTH, FIA) is the data controller and potentially the processor. Likewise, each partner planning to organize demonstration and/or dissemination events will be responsible for any data collected.
3. **Data processor** is a natural or legal person, public authority, agency or other body which processes personal data on behalf of the controller and under their guidance. In Drive2theFuture, data processors are all entities participating in user trials as well entities holding modules of the one-stop-shop. In specific, in Drive2theFuture, those are as follows:
   - **Each pilot site:** Each pilot site might process data, especially vehicle or video data before sharing them with the higher-lever processor.
   - **WP5 and WP6 leaders:** Partners who are leading piloting WP, are responsible for data processing.
   - **Data analysis, consolidation and results reporting:** Partners analysing, consolidating and reporting results are data processors in WP1, WP2, WP5, WP6, WP8.

The current allocation of responsibilities will be re-visited and if needed updated in the next version of this Deliverable.

4. **Data Protection Officer (DPO)** is an enterprise security leadership role to oversee data protection strategy and implementation to ensure compliance with GDPR requirements. The DPO assists the controller or the processor in all issues relating to the protection of personal data. As of 25 May 2018, Regulation (EU) 2016/679 has made mandatory for every public authority and corporation that handles personal data in the EU to have a data protection officer.
5. **Supervisory Authority** (or **Data Protection Authority;** https://gdpr-info.eu/art-51-gdpr/) is a public authority in an EU country responsible for monitoring compliance with GDPR. An EU country within the European Union is also referred to as a *member state*. The key role of the Supervisory Authority is to:
   - advise companies about GDPR;
   - conduct audits on compliance with GDPR;
   - address complaints from data subjects;
   - issue fines when companies are deliberately not complying with GDPR.

   In the case of Drive2theFuture, this means that even if the entity is not obliged to have a DPO, each data processor has checked if they are obliged to obtain approval by the respective authority of their country.
6. A **data subject** 'is a natural person whose personal data is processed by a controller or processor'. In Drive2theFuture, those subjects are all those people participating in focus groups, user surveys and pilot activities. However, we use the term user/participant, as it is more appropriate to both their involvement and role in the Drive2theFuture pilots.

GDPR adherence is closely related to the ethical treatment and monitoring of all human related activities, including data collection and processing. Ethics Boards/ Committees may exist at each entity/institution AND/OR on national level. All entities participating in user trials may need to get approval before proceeding with user trials. Ethical approvals address data privacy and ethical issues relevant to the specifications and requirements at each pilot site.

4.1.1. Internal record keeping

To comply with GDPR requirements on record keeping (Article 30), Drive2theFuture asks all data controllers and processors acting on behalf of the data controller (all acting under the auspices of the project Data Manager) to record their (personal) data processing activities in existing standard templates (Annex 5). These forms include field on the contact information of the data controller(s) and processor(s), purpose and categories of processing, a general description of the technical and organisational security measures, etc. These are submitted to the project Data Manager and, in turn, Project Coordinator for presentation whenever need and/or requested. They are treated as living documents and they must be updated whenever it is necessary. At the end of the project, the main processes and outcomes will be shared with the Coordination team.

4.2. Data Protection Officers (DPOs)

Under the GDPR regulation, approval by a Data Protection Officer (DPO) or notification of the Data Protection Authority (DPA), "whichever applies according to the Data Protection Directive (EC Directive 95/46, currently under revision, and the national law" apply to **controllers and processors whose core**

activities consists of operations requiring regular and systematic monitoring of data subjects on a large scale and who processes special categories of data (Article 35).

Drive2theFuture activities do not fall in any of the above categories and it is not mandatory to appoint a DPO or get authorisation from DPAs. Even if no such obligation exists, relevant partners were asked to provide the following information presented in Table 2:

- Contact details of appointed DPO
- Identification of potential Private Data (PD)
- Overview of data types (controlled/processed by partner)

Table 2. Data Protection Officers & Private data per involved partner

Country	Pilot site/Partner	Privacy (DPO/ private data (PD)	Overview of data types controlled/ processed
WP5 Pilots			
Austria	AIT (Vienna)	**DPO:** Michael Löffler **PD:** Name, Age, Gender, Email, Profession, Images.	Pilot data (vehicle, subjective); PD will be anonymized before share. Images will not be shared.
	AIT (PM and Mobility Observation Box; MOB)	**DPO:** Michael Löffler **PD:** Licence plates, Gender, Faces.	Pilot data (vehicle, subjective; only anonymized data will be shared.
	WL	**DPO and PD:** To be decided during implementation stage – GDPR rules will be followed.	Pilot data (vehicle, subjective; only anonymized data will be shared.
Belgium	VUB	**DPO and PD:** To be decided during implementation stage – GDPR rules will be followed	Pilot data (vehicle, subjective; only anonymized data will be shared.
	VIAS	**DPO and PD:** To be decided during implementation stage – GDPR rules will be followed	Pilot data (vehicle, subjective; only anonymized data will be shared.
Denmark	TUC	**DPO:** Christina Grigoriou Dalsgaard **PD:** Age, gender and job function and/or education (incl. level)	Pilot data (vehicle, subjective; only anonymized data will be shared.
France	IFSTTAR	**DPO:** will be appointed soon **PD:** Video sequence (facial expressions of drivers/ passengers)	Pilot data (vehicle, subjective; only anonymized data will be shared.
	VEDECOM	**DPO:** will be appointed soon **PD:** To be defined	Pilot data (vehicle, subjective; only anonymized data will be shared.
Germany	FZL	**DPO:** To be named **PD:** To be defined	Pilot data (vehicle, subjective; only anonymized data will be shared.
	TUB	**DPO:** Annette Hiller **PD:** Gender, age, job status, job title, job experience	Pilot data (vehicle, subjective; only

Country	Pilot site/Partner	Privacy (DPO/ private data (PD)	Overview of data types controlled/ processed
			anonymized data will be shared.
Italy	DBL	**DPO:** To be named **PD:** To be defined	Pilot data (vehicle, subjective; only anonymized data will be shared.
	SWM	**DPO:** To be named **PD:** To be defined	Pilot data (vehicle, subjective; only anonymized data will be shared.
Sweden	TOI	**DPO:** Not appointed yet **PD:** Background variables as age, gender.	Pilot data (vehicle, subjective; only anonymized data will be shared.
	VTI	**DPO:** Louise Dahlgren **PD:** Age, gender (potential)	Pilot data (vehicle, subjective; only anonymized data will be shared.
Spain	DEUSTO (consolidation)	**DPO:** Mikel García Llorente, Internal Audit and Control Manager as a Data Protection Officer; dpo@deusto.es. **PD:** Aggregated data or consolidated data will be anonymized when they reach DEUSTO (Data processor); hence, no PD are expected to be analysed.	Pilot data (vehicle, subjective; only anonymized data will be shared.
Belgium	FIA (WP1) - Data processing & storage, publication of results	**DPO:** Majken Ekam-Nielsen **PD:** None, if possible	Online survey processing (subjective)
Belgium	VUB (MAMCA Workshop)	**DPO:** dpo@vub.be, **PD:** No, disclaimer will be added in the communication with the participants.	Offline/subjective data from stakeholders
Greece	CERTH (WP1)	**DPO:** Ioannis Chalinidis, email: ivchal@certh.gr **PD:** No.	Data from survey (online/ subjective)
Spain	**RACC** (WP8; website)	**DPO:** Data Protection Committee formed by directors and RACC managers. **PD:** First name, Surname, Email address, Company	Data from website traffic and dissemination events and workshops (online/offline and objective/subjective)
Greece	NTUA (WP2; data simulation models)	**DPO:** Panagiotis Katrakazas, email: pankatr@central.ntua.gr **PD:** No, NTUA will handle anonymized data collected from the surveys.	Simulated data (online/offline and objective/subjective)

All research entities participating in the Drive2theFuture project shall ensure that they have entered into an appropriate data sharing agreement prior to any personal data being shared.

5. Data management obligations

The GDPR aims to secure the privacy rights of EU citizens but it is also designed to bolster innovation. This duality has resulted in some key differences between the GDPR and the Data Protection Directive that are relevant to Drive2theFuture personal data processing activities.

5.1. Research privilege and consent

As a research and innovation action, Drive2theFuture processes personal data only for research and evaluation purposes. GDPR has done away with many restrictions on data processing for research purposes. This has resulted in the easing of a number of conditions on secondary data processing (Article 6(4); Recital 50) and, to some extent, on the requirement for data subjects' consent (Article 6(1)(f); Recitals 47, 157), as long as adequate safeguards are put in place for data processing. Just like the broad definition of privacy in the GDPR, 'research' is also interpreted broadly.

Despite the relaxing of conditions on data processing for research, Drive2theFuture will continue eliciting unambiguous consent from participants after giving them the appropriate information in clear and simple terms using the GDPR compliant informed consent forms annexed in D10.2. All pilot sites will have to set their recruitment and consent strategies and ensure that they comply with the procedures set in Chapter 5 of D10.2, before the administration of these consent forms to obtain users' consent. The Data management team ensured -through the Ethics and Data Privacy policies- that all processes, from participants' recruitment to data reporting, are ethical and GDPR compliant.

5.2. Privacy by design

The GDPR states that "the controller shall…implement appropriate technical and organisational measures…in order to meet the requirements of this Regulation and protect the rights of data subjects". Drive2theFuture DMP details the procedures that will be followed to ensure compliance with the GDPR requirement for data processors and controllers to hold and process only the data necessary for its activities (data minimisation), as well as the limitation of access to personal data to those needing it for processing (Article 23).

6. Data Privacy Impact Assessment

The first version of the Data Privacy Impact Assessment (DPIA) has been essentially prepared in order to provide an overview of all the activities that will take place in the project and, as such, anticipate any potential issues. The herein provided version stands as the current version of DPIA. Still, DPIA is an evolving process in the project. As such, continuous updates fed by respective developments in the project as well as the currently undergoing revisions for legal approval will emerge.

Nevertheless, the next and close to final revision of the DPIA will be held in M18 (D9.7). An initial Data Privacy Impact Assessment template was completed by some partners in order to investigate if personal data and/or sensitive data will be collected during the lifetime of the project and to identify any relevant risks. This was completed prior the completion and submission of D5.1 'Pilots Plans' and as such, the included information will be updated and finalized in the updated version of this deliverable.

In addition to the DPIA, the data privacy policy of the website is continuously being updated adhering to the evolving outcome of the DPIA running in the project. The current data privacy disclaimer can be found in Annex 4, but in the future, it will be reached through the project's web site.

The Privacy Impact Assessment is required under Article 35 of the General Data Protection Regulation (EU) 2016/679. A DPIA is a process which helps assessing privacy risks to individuals in the collection, use and disclosure of information. DPIAs help identifying privacy risks, foresee problems and bringing forward solutions.

This is only a segment of the process, reflected in this Deliverable. Design privacy is not a major priority as the focus of the project is not to create new technologies and as such it is briefly touched upon in section 5.2. The focus on the DPIA at this point, is to highlight the key areas private data might be collected/ processed. This first assessment is not exhaustive. This template should be kept and updated by all partners involved in data collecting and processing.

Answers are highlighted with *magenta colour*.

6.1. Do I have to do a DPIA?

Determining if you need to do a PIA - screening questions

Answering yes to **any** of these questions indicates that a PIA is necessary.

- Will the project involve the collection of new information about individuals? *Yes*
- Will the project compel individuals to provide information about themselves? *In the context of focus groups, user surveys and interviews, the project has and will ask information about participants.*
- Will information about individuals be disclosed to organisations or people who have not previously had routine access to the information? *The information about individuals (meaning participants, users, attendees, respondents) will be disclosed only to the project Consortium for research purposes.*
- Are you using information about individuals for a purpose it is not currently used for, or in a way it is not currently used? *Yes. Will be collected to accommodate research purposes. Data minimisation principle will be applied as much as possible and applicable.*
- Does the project involve you using new technology which might be perceived as being privacy intrusive? For example, the use of biometrics or facial recognition. *Non-intrusive technologies will be used. Non direct privacy intrusiveness is anticipated.*
- Will the project result in you making decisions or taking action against individuals in ways which can have a significant impact on them? *No*
- Is the information about individuals of a kind particularly likely to raise privacy concerns or expectations? For example, health records, criminal records or other information that people would consider to be particularly private. *No*.
- Will the project require you to contact individuals in ways which they may find intrusive? *No*

6.2. Step 1: Identify the need for a DPIA

Explain broadly what aims to achieve and what type of processing it involves. You may find it helpful to refer or link to other documents, such as relevant deliverables and other supportive documents that reside in SharePoint. Summarize why you identified the need for a DPIA.

Explain what the project aims to achieve, what the benefits will be to the organisation, to individuals and to other parties.

WP1: *Define the needs and requirements to develop the project's Use Cases through the collection of big volumes of anonymised data across Europe.*

WP2: *The project aims to create a simulation platform for simulating the behaviour of automated driving, estimating and predicting user acceptance upon autonomous vehicle functions, as well as conducting an impact assessment of autonomous driving on traffic, safety and energy. Within the framework of this activity, a technology acceptance model will be developed using data from surveys. These surveys will be conducted during the pilots and participants will be interviewed concerning their experience using and interacting with autonomous vehicles.*

WP3/WP5.3: *In D2F, Fraunhofer IAO (FhG/IAO) and its third-party Uni Stuttgart (USTUTT) will be involved in optimizing Human-Machine Interfaces (HMI) for the project's different transport modalities. They will identify principles for a user-centred, affective and personalized HMI. This will be based on receiving user data collected on the test sites by the responsible test site partners. All test sites will use standard protocols for the use of consent forms to inform users on the purpose of data collection and get their approval for this. The role of FhG/IAO and USTUTT is to analyse and draw conclusions from the collected data with regard to potential improvements of the HMI in iterative steps and make suggestions to better meet the needs of the different users. The need to conduct a PIA was identified since FhG/IAO and USTUTT will receive collected data and personal information on people's preferences, habits and behaviour, which are the basis for their research activity in D2F. This means that FhG/IAO and USTUTT process, analyse and store the shared user data.*

WP8: *Collect feedback from experts and stakeholders in events about the project and AV acceptance.*

6.3. Step 2: Describe the processing

Describe the nature of the processing

WP1/WP8: *Data will be collected to define the needs and acceptance prerequisites for creating the Use Cases in WP1. Data will be collected across EU countries through an online survey. Completion will be anonymous. In WP8, no plans on data collection have been created but it is anticipated to collect data through the official project website (i.e. common traffic data, as they are described in the preliminary Conditions of Use, Data Policy and Disclaimer, and Cookie Policy of the website (see Annex 4). The Data Policy of the website is in complete alignment with the Data Privacy Policy of this Deliverable and the data privacy defined within D10.2. However, data will be collected anonymously to improve the website performance, to gather information and feedback on the project in dedicated project events.*

WP2: *NTUA will use and analyse stated and preference data collected from surveys during the project real experiments. The project partners responsible for the pilots' realization and organization will collect the data taking into consideration that this data will be used as input from NTUA team in* *order to conduct its analysis. Therefore, it is necessary that the type and format of the data obtained from the experiments will be compatible with the type and format of data that NTUA will need. The data will be shared among the other activity participants - CERTH/HIT, IFSTTAR, TUM, FZI – or any other WP2 participant – TOI, DEUSTO, INFILI – and only after contacting the project manager. The data received will be anonymised as any personal information will be deleted. Therefore, no risk involved in any type of data processing is anticipated.*

***WP3/WP5.3:** FhG/IAO and USTUTT itself will not be responsible for the collection of data since the partners responsible for the testing sites will collect them. The shared data will always be stored safely and securely and only the appointed project employee will have access to the data. Moreover, information will be received and stored in anonymous form such that no inferences on individual people are possible. The raw data will only be stored for the purpose of the project and will be deleted once the work within D2F is completed. FhG/IAO and USTUTT will not share the received data from the corresponding partners with anyone outside the consortium unless this was agreed on. As to our knowledge, processing identified as likely high risk pertains the systematic evaluation of personal aspects of natural persons regarding personal preferences, interests or behaviour in order to create optimizations and rules of adaptation for future HMI.*

Describe the scope of the processing

***WP1**: Data will be collected through an online platform addressing different transport mode users.*

***WP2:** The data will be collected from the experiments (pilots) that will take place within the second year of the D2F project. The amount of data highly depends on the participants selected to be interviewed based on their role in the experiment (initial plan is 1000 users and 200 stakeholders/experts). The data will be kept till the project is finished (April 2022).*

***WP3/WP5.3:** The nature of the data collected might include person-related data such as age, gender and origin, as well as personal preferences, opinions, habits as well as emotional responses of individuals during the study. As to our knowledge, no special category nor criminal offence data will be included.*

Study participants involved in the project will not be personally identifiable since FhG/IAO and USTUTT will only receive anonymized data. Where possible, data will be scrambled and abstracted to permit its use to achieve project outcomes. Data will be kept until one month after the end of the project. The amount of data FhG/IAO and USTUTT analyse and store depends entirely on the studies and user tests conducted on the pilot sites by the corresponding partners (VTI, TUB, AIT, PZM, FZI, TUCO, TOI, VUB/VIAS, SWARCO, DBL, IFSTTAR/VEDECOM). For each pilot site, different numbers of individuals will be affected (between 10 and 500). The data collection will take place in different European countries and covers Greece, Germany, Italy, Sweden, Belgium, Denmark, France, Poland and Austria.

***WP8**: Not defined yet.*

Describe the context of the processing
WP1: *Online survey, not personal contact with respondents.* ***WP2:*** *NTUA does not have any relationship with the individuals, i.e. the participants of the project experiments (pilots). Before the interview the participants would be fully informed about the scope of the survey as well as how their answers will be used and exploited for developing a technology acceptance model as well as enhancing AV behaviour. The data provided to NTUA will be anonymised and NTUA will not have any access to personal information of the participants. Concerning the sample composition, apart from drivers, it will also include vulnerable groups like elderly, people with disabilities, young users and so on.* ***WP3/WP5.3:*** *FhG/IAO and USTUTT are not responsible for data collection, therefore no direct relationship to the individuals exist. The partner collecting data are responsible for informing the individuals about the purpose of data collection and the sharing of information with research partners within the consortium. If requested, we will always delete data. The data processing used in this project is common to user studies aiming at a user-centred interface. As to our knowledge, there are no prior concerns over this type of processing. The current state of technology in the area also considers solutions that automatically collect and analyse user data including personal preferences, habits as well as health-related data in order to assist the user in a more adapted and personalized way. FhG/IAO and USTUTT are not signed up to any approved code of conduct, however, we respect and follow the Federal Data Protection Act to protect the collection, use or processing of personal data. In Germany, the Federal Data Protection Act regulates the usage of personal data in research. The Federal Data Protection Act states protection rules and principles, including obligations on the data controller and the consent of data subjects; rights to access personal data or object to its collection; and security requirements. Furthermore, it covers data processing by third parties; and the international transfer of data. Federal Data Protection Act (Bundesdatenschutzgesetz), 30.06.2017. English Version: https://www.gesetze-im-internet.de/englisch_bdsg/englisch_bdsg.pdf)* ***WP8:*** *Will vary. Not defined yet.*

Describe the purposes of the processing
WP1: *As previously reported.* ***WP2:*** *Through the surveys and the data collection and processing, the automated vehicles will be assessed concerning their strong and weak aspects on safety, comfort and acceptance highlighting best practices of attractive and persuasive HMI solutions. Emphasis will be on trust and over-reliance, loss of control, motion sickness, the flow of transitions, situation and system awareness. The pilots, will be divided in three phases where the participants will have the chance to experience automated vehicle function and HMI enhancements and through the data collected it will be possible to observe increase in their acceptance as well as the fields where further improvement is urgent for making AVs more attractive, safe and comfort. Additionally, within the project, NTUA will develop a technology acceptance model based on these data indicating users' attitude towards AVs through their participation in the project pilots. By comparing the user acceptance (using the UAS 9-scale questionnaire) before and after experiencing the automated functions (with different experience levels and tools; from simulation to real world and comparing various HMI's and automation levels),*

as well as the professional operators' efficiency ratings, the market uptake of higher levels of connectivity and automation will be promoted and encouraged.

WP3/WP5.3: *The data collected during pilot activities by the partners are essential to identify best practices and potential problems of the pilot vehicle HMIs and improve them for the user. The intended effect are affective and user-centred HMIs that ensure user acceptance towards automated transport as proposed within the framework of D2tF. For FhG/IAO and USTUTT, the data processing is a prerequisite to fulfil the research activities of D2F appropriately. More broadly, insights gained from the data analysis will be beneficial for future HMI development in different areas and will pave the way for a better user acceptance of different kinds of user clusters and other involved stakeholders upon the introduction of automation for transport.*

WP8: *Will vary.*

6.4. Step 3: Consultation process

Consider how to consult with relevant stakeholders

WP1: *No consultation outside the Consortium.*

WP2: *No consultation has been or will be undertaken.*

WP3/WP5.3: *Seeking individual's views is not applicable. As to our knowledge, there is no need to involve further people.*

WP8: *Will be relevant during dissemination activities and workshops with experts, information and planning not yet available.*

6.5. Step 4: Assess necessity and proportionality

Describe compliance and proportionality measures, in particular

WP1: *Analysis methodology not available yet but administration methodology (i.e. European level large scale anonymous survey has the extrapolation potential required for creating solid and valid Use Cases.*

WP2: *The data processing technique formulated by NTUA is the appropriate for handling stated and preference data collected from surveys and it is the only way for developing technology acceptance model, as well as observing the technology acceptance evolution and users' opinion and attitude towards AV functions, components and behaviours. It is vital to highlight that NTUA will not conduct the survey and collect the data from the experiment participants, but the university will only receive these data collected from other project partner and it will be anonymized, i.e. any personal information will have been previously removed. NTUA will only process and analyse the data which will not be transferred or sent to any external body. The dataset will be kept within the project community.*

WP3/WP5.3: *The first lawful basis for processing is the user's consent since user data will be collected from subjects that have freely given consent for their information to be processed for a specific purpose. Secondly, analysing actual user data is the basis for user-oriented research as specified in the contract of D2F. These insights cannot be gained using other measures. Data will only be used for the purpose it was collected. The test site partners will limit data collection to the information necessary for the project's purpose and will also fully inform the volunteers participating in the study about the use of their data for research activities. The data shared with FhG/IAO and USTUTT will not serve any other purpose than this project's research activities. Data transfer from other partners to FhG/IAO and USTUTT will make use of secure file exchange platforms.*

WP8: *As methods and activities are still to be defined, this section will be completed in the next roll out of the DPIA process.*

6.6. Step 5: Identify and assess risks

Describe source of risk and nature of potential impact on individuals. Include associated compliance and corporate risks as necessary.	Likelihood of harm	Severity of harm	Overall risk
WP2: *Not enough data collected for ensuring a reliable technology acceptance model and drawing safe conclusions for all categories:* • *Passengers* • *Riders* • *VRUs (elderly, youth, etc)*	*Remote*	*Minimal*	*Medium*

6.7. Step 6: Identify measures to reduce risk

For each of the identified risks, the following measures were selected.

Identify additional measures you could take to reduce or eliminate risks identified as medium or high risk in step 5				
Risk	**Options to reduce or eliminate risk**	**Effect on risk** [eliminated; reduced; accepted]	**Residual risk** [low; medium; high]	**Measure approved** [Yes/No]
1 (WP2)	*Sample size should be efficiently monitored and even distributed among the different participant clusters.*	*To be defined*	*To be defined*	*To be defined*
2 (WP3/WP5.3)	*Risk: Leakage of user data during data transfer from pilot site partners. Impact: Anonymized user data could be leaked.* *However, this risk will be kept minimal by using secure file exchange platforms.*	*Remote*	*Minimal*	*Low*

6.8. Step 7: Sign off and record outcomes

1. **Who has approved the privacy risks involved in the project? What solutions need to be implemented?**

Risk	Approved solution	Approved by
E.g. Risk 1	*Data will be deleted when it is no longer necessary to retain such data.*	*E.g. Data Protection Officer. Note, if there is no DPO or National Agency responsible for that, the data manager (CERTH) will be responsible for looking into the privacy risks (with support from the Managing team and proposing the mitigation solution.*

2. **Integrate the PIA outcomes back into the project plan. Who is responsible for integrating the PIA outcomes back into the project plan and updating any project management paperwork? Who is responsible for implementing the solutions that have been approved? Who is the contact for any privacy concerns which may arise in the future?**

Overall: CERTH is responsible for integrating back into the plan and any future data privacy related communication. For implementing the solutions, depends on who has developed the corresponding part.

Action to be taken	*Date for completion of actions*	*Responsibility for action*
Data to be deleted.	*Insert date/description of when.*	*E.g. Data Protection Officer.*

7. Conclusion and next steps

This version of the Data Management Plan addresses the high-level data clusters, the data privacy policy (Chapters 2, 3, and 4), data treatment and management prerequisites according to GDPR, allocation of data related roles, and a preliminary Data Privacy Impact Assessment. Moreover, an initial version of the data disclaimer for the project's website is included.

The Data Management Plan is a deliverable directly connected to forthcoming evaluation and pilot plans for each of the pilot sites (WP5 and WP6 activities) and the decisions and/or issues arising in Ethics (D10.2) have been applied to the Ethics aspect of data privacy as well. Any ethical considerations, especially about data protection, privacy and security will be fore mostly discussed with the partner acting as the data owner, the members of the Drive2theFuture Ethics Board, and the project management team.

The next step will be to define the primary and secondary data sources in order to elaborate the management plan for each data type and for several data types (e.g. per affective state and/ or Use Case). After the data types are defined, the data management repository's technical, content, and quality specifications for storing and communicating the datasets will be created. For data collected during the pilots all security and ethical guidelines and standards will be applied. In addition, data owners will reach a decision upon data visibility and sharing limitations.

Post-processed datasets, free from any private/personal and identifiable information, will reside in the Drive2theFuture data repositories which will be described in the next revised version of this deliverable. It will also include analytic descriptions of the complete and (non) shareable datasets that will be created during the pilots (WP5), the analyses to follow (WP5 and WP6) and other WPs requiring post-processing (e.g. for the preparation of the Use Cases in WP1 and the simulation modelling conducted in WP2). In addition, the final Data Management Plan will contain the complete structure of the database, descriptions of the metadata files to enable self-explainable (re)use of datasets by external parties. Long-term re-usability of these data is of substantial importance, especially in the field of automated driving experience. Embargos (if any) for parts/segments of data, models, evaluation dimensions (e.g. acceptance, trust), the surrogate and horizontal impact and metadata indicators/estimators will be set by the Partners who own these and the Consortium. Additionally, the final location and format of open Drive2theFuture datasets will be defined.

The final DPIA report will be annexed in the final version Data Management Plan Deliverable (D9.7).

In conclusion, this Deliverable is a living document and it will be regularly updated during the lifetime of the project. If further official updates are necessary, then they will be reported in the final Project Management Report (D9.9; M35).

References

1. Pienta, Amy M.; Alter, George C.; Lyle, Jared A. The Enduring Value of Social Science Research: The Use and Reuse of Primary Research Data. “The Organisation, Economics and Policy of Scientific Research” workshop, Torino, Italy, in April, 2010.

2. The EU General Data Protection Regulation (GDPR) website (https://eugdpr.org/). Last accessed: 17/10/19.

3. Guidelines on Open Access to Scientific Publications and Research Data in Horizon 2020(http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-pilot-guide_en.pdf)

4. Panou, M., Gaitanidou, E. (2019). D10.2 H- Ethics Requirement no. 4. Needs, wants and behaviour of “Drivers” and automated vehicles users today and into the future (Drive2theFuture), Grant Agreement no. 815001.

Annex 1. Other relevant EC guidelines and legislation

In January 2012, the European Commission published proposals for a new framework for data protection legislation. The proposals, in the form of the draft Data Protection Regulation is to replace the existing Data Protection Directive, are now being considered and amended by the European Parliament and Council before adoption. The Regulation covers the use of personal data across a wide range of sectors and will affect how patient data are used in research. The original proposals set out a mechanism for protecting privacy, while enabling research and included a requirement for specific and explicit consent for the use and storage of personal data, while providing an exemption for research, subject to strong ethical and governance safeguards.

In October 2013, the Civil Liberties and Home Affairs (LIBE) committee of the European Parliament adopted amendments that would severely restrict the use of personal data for scientific research purposes without specific consent.

Changes in EU Directives affect the conduct of tests with human participants in all European countries but deviation for existing conditions might differ and be relevant to national legislations and code-of-practice.

Specific guidelines from the EFGCP (the European Forum for Good Clinical Practice) and the American Psychological Association (APA) Ethical Code of Conduct are considered.

The Council of Europe Convention for the protection of individuals with regards to automatic processing of personal data is the first European instrument in this field. It laid down the basic principles of a lawful data processing addressing the threats from the invasion of information systems, such as the data aggregation, at that time. In this respect, it concerns the automatic data processing, although the Member Countries could extend its applicability to non-automatic data processing. Art. 6 states that medical data may not be processed automatically unless domestic law provides appropriate safeguards. The Convention is of limited importance for EU countries after the enactment of the EC Directives on data protection.

The Charter of Fundamental Rights dedicates a separate article to the protection of personal data. Article 8 sets out the right to the protection of personal data of an individual and thus the protection of personal data has now its own legal basis apart from the right to respect for an individual’s private life and the protection of the human dignity. Art. 8 of the Charter sets out the rules for the legitimate processing of personal data, notably that the processing shall be fair and for pre-specified purposes based on the consent of the data subject or other legitimate basis laid down by law. Reference is furthermore made to two rights of the data subject: the right of access to the data and the right to have it rectified. Finally, Art. 8 sets out the need for an which shall control the compliance with the data protection rules.

In 1999 the Council of Europe adopted the Recommendation on the Guidelines for the protection of privacy in the information highways. These Guidelines may be incorporated in or annexed to codes of conduct of Internet service provider to obtain legal validity. The Recommendation is in line with the EC Data Protection Directives regarding the principles of the lawful data processing, the duties of the Internet service providers and the rights of the data subject. The Recommendation encompasses a series of detailed information what the users and service providers shall do to reduce the risks arising from the Internet. It is worth mentioning that the users are required to use digital signature and encryption techniques. On the other hand, the service providers are required to use certified privacy enhancing technologies, to ensure data confidentiality and integrity as well as logical and physical security of the network and the services provided over the network. The service providers shall also incorporate detailed privacy statements on the websites. Finally, the communication of sensitive data, for instance medical data, for marketing purposes requires the previous, informed and explicit consent of the data subject.

The OECD (Organisation for Economic Co-operation and Development) is actively participating in the issues regarding the data protection, the data protection on the Internet as well as the protection of consumer rights regarding e-commerce. First, OECD issued Guidelines governing the protection of privacy stipulating the fundamental principles (OECD, 1980).

In 1998, OECD issued a Recommendation about the implementation of the Guidelines on global networks. The Recommendation addresses mainly commercial sites offering various goods and services, such as tourism, air travel ticket sales, finance, etc. It is not legally binding, unless the Internet service providers stipulate this explicitly. Although the Recommendation does not address healthcare applications, its provisions might apply as following:

The Recommendation imposes the obligation to the web-site provider to refer with a hyperlink to the national legislation on data protection and the national Data Protection Authority. Moreover, every Data Protection Authority should be present on the Internet through relevant, well-documented and interactive sites. The websites shall also maintain on-line privacy statements giving details on the kind of data collected, the purpose of, the use of the clickstream data and processing to which they are subject, as well as the opportunity to opt out. In case of on-line payments by cards they should configure their systems in such a way that they ask for the card details once, if they store this information in highly secure files on non-networked computers. Warning messages on the risks of the Internet shall be provided in case of processing of confidential data. For confidential data the highest degree of security shall be implemented. The implementation of privacy enhancing technologies is also required. Moreover, websites should formally state the acceptance of full responsibility for the security and confidentiality of the personal data collected and processed. Data subjects’ rights the Recommendation highlights the right to access on-line the information collected and stored directly or indirectly, i.e. clickstreams or purchased profiles.

Data Protection Directive 95/46/EC

In 1995, the EC Directive on the protection of personal data was adopted by the Council. The Directive was the first attempt on EC level to recognise the right to privacy and harmonise the national laws. Some main characteristics of the Directive are that it applies equally to public and private bodies, to both automatic and non-automatic data processing, and that the protection is restricted to natural persons (as opposed to legal entities). Moreover, the data must form a part of a filing system, which is defined as any structured set of personal data accessible according to specific criteria.

The directive regulates the processing of personal data, regardless if the processing is automated or not.

Scope

Personal data is defined as "any information relating to an identified or identifiable natural person ("data subject"); an identifiable person is one who can be identified, directly or indirectly, in particular by reference to an identification number or to one or more factors specific to his/her physical, physiological, mental, economic, cultural or social identity;" (art. 2 a).

This definition is meant to be very broad. Data is "personal data" when someone can link the information to a person, even if the person holding the data cannot make this link. Some examples of "personal data": address, credit card number, bank statements, criminal record, ...

The notion processing means "any operation or set of operations which is performed upon personal data, whether or not by automatic means, such as collection, recording, organization, storage, adaptation or alteration, retrieval, consultation, use, disclosure by transmission, dissemination or otherwise making available, alignment or combination, blocking, erasure or destruction;" (art. 2 b).

The responsibility for compliance rests on the shoulders of the "controller", meaning the natural or artificial person, public authority, agency or any other body which alone or jointly with others determines the purposes and means of the processing of personal data; (art. 2 d).

The data protection rules are applicable not only when the controller is established within the EU, but whenever the controller uses equipment situated within the EU in order to process data. (art. 4) Controllers from outside the EU, processing data in the EU, will have to follow data protection regulation. In principle, any online shop trading with EU citizens will process some personal data and is using equipment in the EU to process the data (the customer's computer). Consequently, the website operator would have to comply with the European data protection rules. The directive was written before the breakthrough of the Internet, and to date there is little jurisprudence on this subject.

Principles

Personal data should not be processed at all, except when certain conditions are met. These conditions fall into three categories: transparency, legitimate purpose and proportionality.

Transparency

The data subject has the right to be informed when his/her personal data are being processed. The controller must provide his/her name and address, the purpose of processing, the recipients of the data and all other information required to ensure the processing is fair (art. 10 and 11).Data may be processed only under the following circumstances (art. 7):

- when the data subject has given his/her consent;
- when the processing is necessary for the performance of or the entering into a contract;
- when processing is necessary for compliance with a legal obligation;
- when processing is necessary in order to protect the vital interests of the data subject;
- processing is necessary for the performance of a task carried out in the public interest or in the exercise of official authority vested in the controller or in a third party to whom the data are disclosed;
- processing is necessary for the purposes of the legitimate interests pursued by the controller or by the third party or parties to whom the data are disclosed, except where such interests are overridden by the interests for fundamental rights and freedoms of the data subject.

The data subject has the right to access all data processed about him/her. The data subject even has the right to demand the rectification, deletion or blocking of data that is incomplete, inaccurate or isn't being processed in compliance with the data protection rules (art. 12).

Legitimate Purpose

Personal data can only be processed for specified, explicit and legitimate purposes and may not be processed further in a way incompatible with those purposes (art. 6 b).

Proportionality

Personal data may be processed only insofar as it is adequate, relevant and not excessive in relation to the purposes for which they are collected and/or further processed. The data must be accurate and, where necessary, kept up to date; every reasonable step must be taken to ensure that data which are inaccurate or incomplete, having regard to the purposes for which they were collected or for which they are further processed, are erased or rectified; The data shouldn't be kept in a form which permits identification of data subjects for longer than is necessary for the purposes for which the data were collected or for which they are further processed. Member States shall lay down appropriate safeguards for personal data stored for longer periods for historical, statistical or scientific use (art. 6).

When sensitive data is being processed, extra restrictions apply (art. 8). The data subject may object at any time to the processing of personal data for the purpose of direct marketing (art. 14).

A decision which produces legal effects or significantly affects the data subject may not be based solely on automated processing of data (art. 15). A form of appeal should be provided when automatic decision-making processes are used.

Supervisory authority and the public register of processing operations

Each member state must set up a supervisory authority, an independent body that will monitor the data protection level in that member state, give advice to the government about administrative measures and regulations, and start legal proceedings when data protection regulation has been violated. (art. 28) Individuals may lodge complaints about violations to the supervisory authority or in a court of law.

- The controller must notify the supervisory authority before he/she starts to process data. The notification contains at least the following information (art. 19):
- the name and address of the controller and of his/her representative, if any;
- the purpose or purposes of the processing;
- a description of the category or categories of data subject and of the data or categories of data relating to them;
- the recipients or categories of recipient to whom the data might be disclosed;
- proposed transfers of data to third countries;
- a general description of the measures taken to ensure security of processing.
- This information is kept in a public register.

Art. 29-Data Protection Working Party: Working Document on Privacy on the Internet

The Data Protection Working Party has been established by art. 29 of Directive 95/46/EC and is the independent advisory body on data protection and privacy. Its tasks are laid down in art. 30 of Directive 95/46/EC and in art. 14 of Directive 97/66/EC. The opinions and recommendations of the Working Party are not legally binding, reflect, however, the current trends on European level and influence the decisions taken by the European Commission and the Committee established by art. 31 of Directive 95/46/EC.

This working document seeks to raise awareness and to promote the public debate on issues of on-line data protection. It therefore provides detailed information on technical aspects of how the Internet and the communications through the Internet are organised and what are the main privacy risks arising from the use of the Internet. In this context, it aims at the same time to provide an interpretation of the data protection Directives in that field. It follows a "holistic" approach by basing the analysis of privacy risks, the obligations and rights of the involved parties on both the general data protection Directive 95/46/EC and the privacy and telecommunications Directive 97/66/EC.

The risks to privacy arise from the activities of the various intermediaries. For instance, the use of routers, e.g. the telecommunications nodes in the Internet, which have the characteristic that the information may pass through a non-EU country which may or may not have adequate data protection, if this at the time of transmission is the "shortest" way of transmission.

According to the opinion of the Working Party, Directive 97/66/EC applies to telecommunication service providers who connect Internet users and ISPs and access service providers who provide the requested Internet service, transfer the request from the Internet user to proxy server and then to the requested website. It also applies to providers of routers and connecting lines. Moreover, the Directive 97/66/EC shall apply also to Internet Service Providers (ISPs) providing hosting services, such as portal services, who may log the requests, the referring pages and post cookies on the hard disk of the user and make profiles. The latter is, however, arguable since the host service providers transmit content information and thus it should rather come under the general data protection Directive. The working document recognizes that the applicability of the Directive 97/66/EC to the activities of the host service providers is not always clear. When the provider hosts its own portal site, it comes under the general data protection directive whilst it comes under the specific when he plays the role of the access service provider.

The providers of Internet services, dependent on the distinctions, are subject to the obligations to confidentiality and security laid down in both Directives (art. 4, 5 97/66/EC, art. 6 - 8, 16, 17 95/46/EC). Traffic data provided by providers of routers and connecting lines, ISPs and telecommunication providers shall be protected as content data according to art. 5 of Directive 97/66/EC as this is the case in the proposal for an amendment of 97/66/EC.

Interception of communication is unacceptable unless it fulfils three fundamental criteria in accordance with art. 8 (2) EHRC, and the European Court of Human Rights interpretation of this provision: a legal basis, the need for such a measure in a democratic society, and conformity with one of the legitimate aims listed in the Convention.

The Working party strongly recommends the use and offer of encryption tools by the providers of email services at no additional cost. The providers should also offer secure connection for the transmission of the emails. The need of integrity and authentication should be considered as well.

A means for ensuring encryption is the Secure Socket Layer (SSL) which is implemented in most popular browsers and establishes a secure channel between the client and server computers. This is achieved by means of encryption and digital certificates. SSL enables the authentication of the server to whom the information shall be sent and the integrity of the data. It does not ensure the authentication of the client. These difficulties shall be overcome by the protocol SET (Secure Electronic Transactions) that provides for confidential transmissions using encryption, authentication of the parties, integrity and non-revocation (through digital signatures). The Working Party seems to support the use of the SET protocol instead of SSL, especially when sensitive information, such as the credit cards data, will be transmitted. Moreover, if a higher level of security is needed, the digital certificates should be stored on smart cards.

It must be pointed out that concrete legal requirements on the data collection depend also on national legislation. All the above EC Directives and International Agreements will be fully adopted within Drive2theFuture. The conformance to them will be safeguarded by the Coordinator and the Technical Manager of the project.

Annex 2. Data storage, back up, documentation, naming and managing guidelines

Data collected by the sensors and researchers at the pilots must be securely stored and regularly backed-up. Sometimes multiple copies should be made, especially for large datasets that need to be stored in large capacity external hard drives. A separate checklist has been prepared and should be used by all sensors/systems' providers not only during evaluations but when the services are technically validated and/or integrated to the overall architecture. Data that will be stored as a result of regular checks and tests performed by the administrators, that wish to perform regular checks and tests, must create a database and use the following checklist:

Checklist

- ✓ How will the raw data be stored and backed up during the research?
- ✓ How will the processed data be stored and backed up during the research?
- ✓ Which storage medium will you use for your storage and backup strategy? Network storage; personal storage media (CDs, DVDs, USBs, portable hard drives); cloud storage and how reliable as well as long-lasting is it?
- ✓ Are backups made with sufficient frequency so that you can restore in the event of data loss?
- ✓ Are the data backed up at different locations?

Each site and sensor/system responsible should ensure that data are regularly backed-up and they are stored in secure and safe location. There is a common "rule-of-thumb" to only store data that you actually need in three different copies. It is advised that copies can be stored in both local and remote storage units/locations.

The following data storage options can be used:

External hard drives/USB sticks: will be used in long-trials (WP7) and local evaluations. They will serve as backups and intermediate storage units before transferring data to a permanent/long-term storage place.

Advantages

- Offline access;
- Status not affected by external settings and environmental conditions;
- Easy to carry;
- Removable.

Disadvantages

- If information is lost, no other back up system exists;
- Easily corrupted and destroyed.

Personal computers and laptops: Similarly, they will mainly serve as a short-term option and for transferring data after the evaluation sessions to a selected storage place.

Advantages

- Easy access for data analysis and treatment.

Disadvantages

- Not so easily moved;
- Easy to corrupt.

Cloud storage: only aggregated, anonymised and confidential data will be stored on the project cloud storage, depending on the level of agreement between partners who have access to these data. In general, data will be stored that the individual cannot be identified by the shared information and data.

Advantages

- Data difficult to corrupt and lose;
- Might need to pay for storage of data and depends on GB already stored and further data we wish to store;
- Access from different devices;
- Easy to share.

Disadvantages

- Available only online, hence always needs internet access;
- Higher risk of privacy bridging;
- Needs to abide to online sharing and protection protocols (i.e. technical expertise is necessary);
- Administration and access require credentials.

Network/fileservers: large data sets will be stored, and they will serve as the long-term storage solution. Regular backups will ensure data are not lost or corrupted.

Advantages

- Data difficult to corrupt and lose;
- Might need to pay for storage of data and depends on GB already stored and further data we wish to store;
- Might need to create database and have administration team to manage, monitor as well as sustain;
- Access from different devices and remote access;
- Easy to share.

Disadvantages

- Available only online, hence always needs internet access;
- Higher risk of privacy bridging;
- Needs to abide to online sharing and protection protocols (i.e. technical expertise is necessary);
- Technical infrastructure might be needed;
- Administration and access require credentials.

Data Documentation

Data documentation will mostly be metadata and will be used in order to recognize each data type and source. The initial documentation details that will be included for each service are shown below.

Table 3: Data documentation matrix

Basic	Source 1	Source 2	Source n	Advanced
Data name				Definition of variables
Who created/contributed to data				Vocabularies
Date created				Units of Measurement
Data modified				File format and type
Conditions				Methods used/assumptions made/analytical procedural info; description

The role of this metadata file will be to accompany any data to be accessed and used by other than the individuals/organizations that collected them. External users and analysts have to understand any underlying processes in order to understand the data and to be able to re-use them. The details that will define and create the specific dataset profile can be:

Basic: details defining the origin, the type, the creation dates and the way these data were created.

Advanced: include definitions of variables, the methods used and applied in order to gather and capture data based on common standards (incl. any procedural steps taken). A detailed and technical specification of data might be included like the variables that define the specific dataset, the units of measurements and the assumptions applied for the data collection (e.g. zero might mean no activity recorded, therefore the patient is not active). The format and file types used for collecting and storing the data (and size) can be of help if external users show interest in re-using a dataset (i.e. checking compatibility with s/w they use in order to aggregate or analyse data utilise import/export functionalities).

File naming

File naming depends largely on service and the datasets to be derived by this service and/or connected with this service. They must be **consistent** and **descriptive**.

Common file naming and organizing among partners helps to organize effectively and efficiently their work and, of course, ease collaboration with other partners. Additionally, partners using this file naming rationale will find it easier to work (and share) the correct version of data and accompanying metadata files. The following file naming offers a consistent naming of the files in order to make it easier to identify, locate and retrieve the data files.

This file and folding naming system will be used for all data and metadata files. Some elements that can be used are the ones below:

1. **Project short acronym:** D2F
2. **User group:** Indicate if file is related to any driver type (e.g. car or bus driver)
3. **Use Case:** Indicate relation to Use Case
4. **System/sensor/data type related:** e.g. infrastructure
5. **Location (where it resides):** e.g. Local unit
6. **Researcher name/initials:** JS
7. **Pilot identifier:** e.g. OR1
8. **Date or range of pilot:**241019
9. **Type of data:** subjective or logged or simulated
10. **Conditions:** road, rail, etc.
11. **File version number:** Only singular number are acceptable (1, 2, 3)
12. Three letter file extension for application specific files (e.g. csv)

Any spatial characters are avoided because they might not work well with certain programmes and avoid spaces (i.e. use underscores instead).

Each data folder will include a regularly updated README.txt in the directory to explain the codes, abbreviations used and, in general, the coding practices and naming conventions used.

Based on the example used above, an efficient naming convention within the Drive2theFuture project looks like that:

DRIVE2THEFUTURE_sensor_RO1.csv

Use of identifiers

Apart from a common file naming system, a reference number, like the ones used in libraries or journals can offer a long-term and unique identifier that remains the same and will not change over time. A global identifier standard that could be applied also for the datasets to be created (both data and metadata files) is the Digital Object Identifiers (DOI) that now can be used for datasets. Further guidelines for partners on how to automatically assign a DOI can be found in the web site of the International DOI Foundation (IDF): http://www.doi.org/.

An identifier may be assigned to each separate dataset. Dataset is defined as the set of data gathered by each module/service with consideration for included different data types.

Data management guidelines

Partners should ensure any data they share through Dropbox or other online channels are discoverable and identifiable. In addition, datasets and relevant documents should be accessible to targeted audiences and if not for a certain period (e.g. embargos, licenses) clarifications should be provided.

The information within the dataset needs to be comprehensible and self-explanatory. If this is not the case, then accompanying (e.g. read.me) files should be included. Parts of data, software, documentation and results will be available to be reviewed and validated. By such means, the data will be re-used because they will be open to criticism and assessment.

When partners are collected and sharing data, they must keep in mind that if they decide to open them to external / third parties, then they need to be in an intelligible form even for a long time after they have been collected. Metadata can be of use to outside groups that are not necessarily researchers but are associated to this area (e.g. transport professionals, vendors). Data are preserved and curated with interoperability in mind across groups of potential users (e.g. countries). Furthermore, data annotation and labelling enable combination and comparison with other datasets outside to the project.

Annex 3. Data clusters and characteristics

As most activities are currently planning their operations and data to collect and process, this is a very preliminary list of data to be collected during the lifetime of the project across the primary data-related WPs (WP1, WP2, WP5, WP6 and WP8). The aim of this table is, on one hand, to provide a depiction of data sources and, on the other hand, the data characteristics that we must address to accommodate for efficient and GDPR-compliant data management.

Table 4. Data clusters & Sources

Systems/sensors /other	Data (reference and name)	Data description	Metadata	Standards (incl. interoperability)	Privacy	Confidentiality	Archiving and preservation	Data sharing	Comments/suggestions
Interior Camera	Video recording sequences	Video recordings of 5 safety critical situations and interactions with cyclists,	Labels	TBD	Local use	Local use	TBD: But initial recommendation is to share only annotations and notes	TBD	This column will be completed when datasets are defined, and partner have set sharing preferences.
Questionnaire, workshop, interview data (F2F) from surveys, pilots, demonstrations, events, etc. as well as online surveys, direct observations and event diaries	F2F subjective data/ Logged and entered	These are questionnaires and interview hard or electronic copies that they will be collected and store locally	Surrogate data, graphs, artefact data	TBD	Local use	Depends on WP and administration mode	Online database, local storage and use, etc.	TBD	
Infrastructure sensors	Objective, sensor data/ Logged	Collected and stored locally	Raw, surrogates, annotated data	TBD	Not related	Local use	Local storing, metadata and treated data	TBD	

Systems/sensors /other	Data (reference and name)	Data description	Metadata	Standards (incl. interoperability)	Privacy	Confidentiality	Archiving and preservation	Data sharing	Comments/suggestions
							sharing within Consortium		
Simulation data	Combination	Stored locally	Raw data, simulation files	TBD	No related	Local implementation	Local storing	TBD	
Onboard exteroceptive and proprioceptive car sensor data	Objective/ Logged	Collected and stored locally	Raw, surrogates, annotated data	TBD	Not related	Local use	Local storing, metadata and treated data sharing within Consortium	TBD	
Driving (vehicle) performance data (e.g. reaction times)	Objective/ Logged	Collected and stored locally	Raw, surrogates, annotated data	TBD	Not related	Local use	Local storing, metadata and treated data sharing within Consortium	TBD	
Wearable data	Objective/ Logged	Collected and stored locally	Raw, surrogates, annotated data	TBD	Not related	Local use	Local storing, metadata and treated data sharing within Consortium	TBD	
Social media data	Subjective/ Logged	Stored locally	Only surrogated and anonymous	TBD	Not related	Local use	Local storing, metadata and treated data sharing within Consortium	TBD	
TLA app performance data	Objective/ Logged	Collected and stored locally	Raw, surrogates, annotated data	TBD	Not related	Local use	Local storing, metadata and treated data sharing within Consortium	TBD	

Systems/sensors /other	Data (reference and name)	Data description	Metadata	Standards (incl. interoperability)	Privacy	Confidentiality	Archiving and preservation	Data sharing	Comments/suggestions
Retrofitted eye tracking	Objective/ Logged	Collected and stored locally	Raw, surrogates, annotated data	TBD	Not related	Local use	Local storing, metadata and treated data sharing within Consortium	TBD	

Annex 4. Privacy Disclaimer (website and online presence)

Any dissemination outcomes, activities and open source available project products are required to carry the EU and project logos and disclaimer as well as the privacy disclaimer of the project. The basic information statement prepared by the dissemination manager and will be updated for the next version of this Deliverable is the following:

BASIC INFORMATION ON DATA PROTECTION

The Owners of the Portal, which are the entities listed in the section "Owner of the Portal", as Controllers of your data, will use the data you have provided to send you newsletters and information on the project Drive2thFuture within the framework of the H2020 Project, Grant no. 815001 of the European Union. The mentioned data processing will be carried out based on your request. The Rocket Science Group LLC d/b/a Mailchimp will have access to your e-mail-address to send you the requested information. Your data will be kept during the terms established in the applicable regulations. You may exercise the following rights: objection, access, rectification, erasure, limitation and portability at the following e-mail address ……………………... You will find additional information in the privacy policy section of this website.

Conditions of Use and Privacy Policy

These conditions of use and the privacy policy can be revised and/or modified at any time, including a new wording on the website. These changes will be applicable from that moment on.

CONDITIONS OF USE

1. DEFINITIONS

In these conditions of use, the following terms shall have the following meanings:

Owner of the Portal - The owners of the portal are the following entities:

ETHNIKO KENTRO EREVNAS KAI TECHNOLOGIKIS ANAPTYXIS (CERTH),

STATENS VAG- OCH TRANSPORTFORSKNINGSINSTITUT, TRANSPORTOKONOMISK INSTITUTT,

NATIONAL TECHNICAL UNIVERSITY OF ATHENS – NTUA,

UNIVERSITA DEGLI STUDI DI ROMA LA SAPIENZA,

VRIJE UNIVERSITEIT BRUSSEL,

UNIVERSIDAD DE LA IGLESIA DE DEUSTO ENTIDAD RELIGIOSA,

INSTITUT FRANCAIS DES SCIENCES ET TECHNOLOGIES DES TRANSPORTS, DE L'AMENAGEMENT ET DES RESEAUX,

SWARCO MIZAR SRL,

EURNEX e. V.,

DEEP BLUE SRL,

FRAUNHOFER GESELLSCHAFT ZUR FOERDERUNG DER ANGEWANDTEN FORSCHUNG E.V.,

TECHNISCHE UNIVERSITAET MUENCHEN,

STIFTUNG FZI FORSCHUNGSZENTRUM INFORMATIK,

IRU PROJECTS ASBL,

HUMANIST,

FOUNDATION WEGEMT - A EUROPEAN ASSOCIATION OF UNIVERSITIES IN MARINE TECHNOLOGY AND RELATED SCIENCES,

AIT AUSTRIAN INSTITUTE OF TECHNOLOGY GMBH,

UNION INTERNATIONALE DES TRANSPORTS PUBLICS,

FONDATION PARTENARIAL MOV'EOTEC,

PIAGGIO & C S.P.A.,

FEDERATION INTERNATIONALE DE L'AUTOMOBILE,

The Institute of Advanced Motorists Ltd,

POLSKI ZWIAZEK MOTOROWY,

WIENER LINIEN GMBH &CO KG,

TECHNISCHE UNIVERSITAT BERLIN,

TUCO YACHT VÆRFT APS,

INFILI TECHNOLOGIES PRIVATE COMPANY,

STELAR SECURITY TECHNOLOGY LAW RESEARCH UG,

AUTOMOBIL CLUB ASSISTENCIA SA,

VIAS INSTITUTE

Conditions of use: The terms and conditions that regulate the legal relation between the owners of the Portal and the user as regards the access and use of the portal by the user.

Portal: Located in the Internet site (url) www.drive2thefuture.eu

Products: The goods and services offered through the portal to be assessed and/or purchased by users.

Users: The natural person or legal entity that has access to and uses the portal.

2. SERVICE PROVIDED

2.1 The Service

The service provided by the Portal is offering users the free access and use of the portal, in the websites of which they can get information on the Project Drive2theFuture, which is framed within the H2020 Programme, Grant no: 815001 of the European Union.

2.2 Third-party Advertising and Links

Advertising third parties: The transactions between the user and advertising third parties found in or through the portal are made only and exclusively between the user and the third parties. The user acknowledges and accepts that the Owners of the Portal will not be liable for any damages whatsoever arising out of the transactions or the presence of this kind of advertising parties in the portal.

Links: The inclusion of links does not imply in any case the existence of a relationship of any type between the Owners of the Portal and the owner of the respective website. Any bond, link, forwarding or association (“links”) with other websites (www), made available through the portal do not imply any type of guarantee by the Owners of the Portal for the user.

2.3 Contracting Parties and Disputes

Except as otherwise agreed, the Owners of the Portal are neither a party nor are they involved in or liable for the transactions, agreements, contracts or disputes between the users and the respective product seller. In case of disputes the user will hold the Owners of the Portal, its agents and employees harmless from any dispute, demand or liability, of any type, related to or arising out of such transactions, agreements, contracts or disputes.

3. PRIVATE AREA

3.1 Access

The user who chooses to register in the portal will have to enter his/her name and surname, professional contact e-mail address and the company he/she is representing, and therefore needs to enter a password in the registration module of the portal. The user will be responsible for keeping his/her account and all and any transactions made with it confidential. The user undertakes to not giving the password to any other person and to take all necessary measure to keep the password secret. The user is the only and exclusive responsible of the password’s custody and for the use, authorised or not, by third parties. The user must check that the data provided are correct.

The use of the e-mail address and the password of a user will be understood as made on behalf of the user. In this sense, the user will have to notify immediately the Owners of the Portal, about any access restriction, modification of cancellation of the e-mail address or the password he/she wishes to be made.

The user knows that the access to the portal using the identity or the password of another user, as well as obtaining, using or disseminating personal data of other users may be considered as a legal and possible penal infringement. In relation with the registration process in the portal, the user undertakes to provide current, complete and true information and to keep it updated during the whole use of the portal.

3.2 De-Registration

The Owners of the Portal may, at any time and whenever they consider it to be necessary for the good working of the portal, de-register a registered user, regardless of the fulfilment of the pending obligations arising out of these conditions of use or the transactions agreed on.

4. RIGHTS AND OBLIGATIONS OF USERS (registered or not)

4.1 User’s Responsibility

The user acknowledges and voluntarily accepts that the use of the portal is made, in any case, under his/her sole and exclusive responsibility.

The user shall be liable for the damages of any nature that the Owners of the Portal may sustain as a consequence of the infringement of any of the obligations that the user has taken over by virtue of these conditions of use or by law as regards the use of this portal.

4.2 Compliance with Laws

The user undertakes to access and use the portal in accordance with the law, the General Conditions of Use and generally accepted moral standards and good practice. The user likewise undertakes not to use the portal for unlawful or illegal purposes, purposes that harm the rights of third parties, or to use it in ways that damage or disable the Website or the rights of third parties.

4.3 Non-Interference with the Portal

The user declares, guarantees and undertakes expressly not to send, transfer, distribute or publish through the portal, materials that (i) restrict or entirely prevent other users from utilising and enjoying the portal; (ii) are illegal, threatening, abusive, harmful, slanderous, defamatory, hateful, racist, obscene, vulgar, offensive, pornographic, disrespectful towards religions or indecent; (iii) constitute or might constitute facts that could give rise to civil or criminal suits; (iv) injure, violate, plagiarise or infringe the rights of third parties, including intellectual and industrial property rights; (v) contain viruses or other harmful components capable of interrupting, destroying or limiting the functionality of any computer programs or equipment; (vi) the user is not entitled to transfer; and (vii) are unsolicited or subliminal advertising materials; or (viii) contain false or deceitful statements and instructions.

The user declares, guarantees and undertakes expressly (i) not to interfere with the website or the servers connected to the website and to obey the requirements, systems and procedures thereof; (ii) not to obtain unauthorised access to other systems through the website; (iii) not to exert an influence on the website such as to negatively affect the transactions of other users; (iv) not to threaten or coerce other users; (v) not to store or gather personal or professional data of other users; (vi) not to act in a false and unauthorised manner in the name and on behalf of other persons or entities; (vi) not to engage in practices that constitute unfair competition

The user declares, guarantees and undertakes expressly not to show, sell, transmit, use, store, extract or exploit the personal or commercial names, addresses, telephone and fax numbers, e-mail addresses, lists, prices, fees or any other information related to the portal or the users.

4.4 Information on the Processing of Personal Data

See preliminary Privacy Policy of RACC (i.e. the website holder).

5. CHANGE OF THE CONDITIONS OF USE AND THE PORTAL

The Owners of the Portal may change and amend the total or partial content of the conditions of use that are available and accessible at the bottom of each page of the portal at any time. Browsing in the portal implies (i) accepting the conditions of use of the portal and that (ii) the acceptance is permanent, constant and ratified, and that, at any given time, this acceptance shall be understood as referred to the conditions of use that are applicable at any time.

6. PROPERTY RIGHTS

The user acknowledges and accepts that the portal and any software used in relation with it, contains priority and confidential information and that it is protected by intellectual and industrial property rights. Without the express consent of the Owners of the Portal, the user may not alter, rent, provide, sell, copy, reproduce, transfer, distribute or create works deriving from or based upon the portal or upon its content. The user may not connect this portal to another website, nor resell or redistribute any part of the portal, nor provide third parties with access to the portal.

No licence: The Owners of the Portal are not providing any license or authorisation to use of any kind for their intellectual or industrial property rights, or for any other rights related to the portal or its content.

7. INDEMNIFICATION

The user shall defend, hold harmless, and indemnify the Owners of the Portal, its subsidiaries, directors, owners, employees, agents, collaborators, shareholders, commercial partners and suppliers from and against any and all liability and damages of any kind arising out of (i) the use of the portal by the user, (ii) the infringement by the user of the conditions of use, (iii) any dispute or claim between the user and a third party, (iv) the infringement of third party rights by the user.

8. GENERAL

8.1 Communications

All communications to be made between the Owners of the Portal and the user shall be made by e-mail (i) when addressed to the Owners of the Portal to the e-mail address ………………………… and, (ii) when addressed to the user, to the e-mail address provide by the user.

The reply of the Owners of the Portal will depend on the volume of e-mails and messages received, and on the complexity of the questions posed, although it will always stick to the legally established deadlines, without guaranteeing the correct working of the e-mail and of the messages when being received or sent.

The data provided by the user in this e-mail or message will be processed according to the applicable regulations as regards the protection of personal data.

8.2 Applicable Law and Jurisdiction

The Portal and the conditions of use will be regulated and construed according to Spanish and EU Law.

The Owners of the Portal and the user will do any reasonable efforts to solve in a friendly manner, any dispute that may arise out of or in relation with these conditions of use and/or the portal. To this end, any claims may be made by phone, by mail or on the website …………….

Should it not be possible for the parties to solve a dispute, the issue will be submitted to the competence of the respective Judges or Courts, according to the applicable regulations.

8.3 Transfer or Assignment

The rights and obligations of the user about the portal cannot be transferred or assigned, neither totally nor partially, by the user to third parties without the previous consent by the Owners of the Portal.

The user authorises the RACC, from this moment on, to transfer its rights and obligations of the portal to third parties, exclusively for the purposes stated in the Privacy Policy section, and always subject to what is established in the respective regulations in terms of data protection.

8.4 Preservation of the Legal Relation

Should any clause of the Portal or of these Conditions of Use be declared, totally or partially, null or invalid, the nullity or invalidity shall only affect the respective provision. The remaining clauses of the portal or the conditions of use shall continue in force and the respective provision or part affected shall be considered as non-existing, except if it affects the provisions or conditions of use, because it is an essential part.

8.5 Total Agreement

The conditions of use are the total agreement between the parties in relation with the object of their legal relation.

PRIVACY POLICY

Your privacy is very important to Drive2theFuture Consortium.

We have drawn up this privacy policy in order to provide you with information on the personal data compiled by the Owners of the Portal through our interactions with you, how we use the data and with which purpose. Please consider that whenever we refer to "personal data", we refer to any and all information about a natural person (not a company) which has been identified or the identity of which can be established.

Controller: The data controllers are the Owners or the Portal which are the entities listed in the section "Owners of the Portal".

Purpose of the data processing: The purpose of the processing of personal data is managing your registration into the website and sending you newsletters and information dealing with the project Drive2theFuture of H2020 EU Programme, Grant no: 815001.

Legitimation of the processing: The processing of your personal data is carried out based on the request that has been made.

Recipients: The data will not be disclosed to third parties except if required by law.

Please be informed, that in order to be able to send you information on the project DRIVE 2 THE FUTURE, your e-mail address will be forwarded to The Rocket Science Group LLC d/b/a Mailchimp, privacy@mailchimp.com, with registered office at 675 Ponce de Leon Ave NE, Suite 5000, Atlanta, GA 30308 USA, an entity that has been certified within the framework of the EU-U.S. Privacy Shield (EU Decision 2016/1250 of the Commission, of 12 July 2016), that will access your data as Processor of the data. Information available at:

https://www.privacyshield.gov/participant?id=a2zt0000000TO6hAAG&status=Active

Preservation of the Data: The personal data provided will be kept for as long as they are necessary for the purpose they were collected. The cancellation of the data will be made by blocking. We will keep your data blocked during the terms established by the applicable regulations, and we shall proceed to their physical erasure once these terms have finished.

Rights: You may exercise the following rights: objection, access, rectification, erasure, limitation, portability, revocation of consent and contestation of any decision based on the automatic processing of data, by means of a written request to the following e-mail address ………………………. You will have to identify yourself by means of an official document.

COOKIE POLICY

The Owners of the Portal inform the users that the acceptance of these Conditions of Use and Privacy Policy implies the express authorisation by the User for the use of cookies. Cookies are used with the aim of making browsing easier and offering services in an agile and customised way, as well as for your security.

What are cookies?

A cookie is a file that is downloaded into your computer when accessing certain websites. Cookies allow a website to, amongst others, store and recover information about the browsing habits of a user or his/her device, and, depending on the information they contain and the way in which you use your device, they can be used to recognise the user. If you do not wish cookies to be installed in your hard drive, you must change the settings in your Internet browser in order to avoid the reception or to get a message on the screen when receiving them. All cookies can be deleted from the browser at any time. (See *How can I manage my cookies?*)

What types of cookies are used by this website?

Own cookies: are those that are sent to the user's device from another device or domain that is managed by the entity that is providing the service requested by the user.

Third-party cookies: are those that are sent to the user's device from another device or domain that is not managed by the entity, but by another entity that processes the data gathered by means of the cookies.

Session cookies: are cookies designed to gather and store data while the user is accessing a website.

Persistent cookies: are those which keep data stored in the user's device and can be accessed and processed during a specific period by the entity responsible for the cookie. The time period might go from a few minutes to several years.

Technical cookies: are those that allow the user to browse around the website and to use different options or services available on the site, such as controlling the data traffic and communication, identifying the session, accessing sections with restricted access, using security elements while browsing, etc.

Customisation cookies: are those that allow the user to access the service with some general features, which are predefined according to a series of criteria in the user's device like, for example, the language, the type of browser used to access the service, the regional configuration from which the user accesses the service, etc.

Analysis cookies: are those that, well processed by us or by third parties allow to make a follow-up and an analysis of the behaviour of users. The information gathered with this type of cookies is used to measure the activity of the website and to draw-up browser profiles of users, with the aim of introducing improvements based on the analysis of data on the use that the users make of the service.

Advertising cookies: are those that are processed either by us or by third parties to manage the offer of advertising space on the website in the most efficient way, adapting the content of the advert to the content of the requested service or to the use that you make of our website. Thus, we can analyse your Internet browsing habits and show you advertising related to your browsing profile.

Likewise, it is possible that after opening a website or an e-mail with an advert or promotion of our products or services, a cookie is installed in your browser, which will later be used to show you advertising related to the search you made and to control our adverts.

What do we use cookies for?

We use cookies to:

i) show our website, make it work correctly, create your user account, start your session. These technical cookies are necessary for our website to work correctly, to remember your preferences and to help you use our website in an efficient and effective way, such as remembering your username. The portal can be accessed without the activation of the cookie options, but it might hinder the correct working of security mechanism for certain services that need higher security.

ii) understand how our users use the website, detect what is working and what is not, optimise and improve the website, and make sure that our users continue to consider our website as interesting and relevant.

iii) show advertising of the Owners of the Portal on other websites (retargeting). These cookies are stored by trustful third parties. The use of cookies by those companies is subject to their own privacy conditions. (See *"Third party cookies"*)

How can I manage my cookies?

You may accept or reject the use of cookies by changing the settings of your browser. However, the fact of rejecting cookies may hinder the correct function of the security mechanisms for certain services that need higher security, as well as some of the services provided through our website. Below you will find instructions to activate and delete cookies:

Windows Internet Explorer

Firefox

Google Chrome

Safari

Updating of the Cookie Policy

The Cookie Policy may be amended according to the legal demands. Therefore, we recommend users to regularly check the content.

If you wish more information on the use of Cookies, feel free to contact us at the following e-mail address……………………

Annex 5. Data processing - record keeping template

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. These documents should be treated by responsible persons as living, which they are responsible to keep up to date and share them with the management team whenever they are requested.*

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**
5	**Where possible, the envisaged time limits for erasure of the different categories of data**

6	**Where possible, a general description of the technical and organisational security measures for**
a	the pseudonymisation and encryption of personal data;
b	the ability to ensure the ongoing confidentiality, integrity, availability and resilience of processing systems and services;
c	the ability to restore the availability and access to personal data in a timely manner in the event of a physical or technical incident
d	a process for regularly testing, assessing and evaluating the effectiveness of technical and organisational measures for ensuring the security of the processing;

II. Data processor’s record of processing activities

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

4	**Where possible, a general description of the technical and organisational security measures for**
a	the pseudonymisation and encryption of personal data;
b	the ability to ensure the ongoing confidentiality, integrity, availability and resilience of processing systems and services;
c	the ability to restore the availability and access to personal data in a timely manner in the event of a physical or technical incident
d	a process for regularly testing, assessing and evaluating the effectiveness of technical and organisational measures for ensuring the security of the processing;