Adaptive ADAS to support incapacitated drivers Mitigate Effectively risks through tailor made HMI under automation

Deliverable 10.2 – *Data Management Plan*

Work Package No.	WP10
Work Package Title	Management
Activity No.	A10.2
Activity Title	Technical & Innovation Management
Dissemination level	Public
Main Author(s)	K. Touliou, S. Nikolaou, A. Dimou (CERTH/HIT)
File Name	ADASANDME_10.2_Final.doc
Online resource	Link-to-Deliverable

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 688900

Version History

Date	Version	Comments
28/02/2017	0.1	Version ready for review
13/03/2017	0.2	Received comments and feedback from reviewers
24/03/2017	Final	Final version submitted to EC incorporating all comments made by the reviewers

Authors (full list)

K. Touliou, S. Nikolaou, A. Dimou (CERTH/HIT)

Project Coordinator

Dr. Anna Anund
Research Director / Associate Professor
VTI - Olaus Magnus väg 35 / S-581 95 Linköping / Sweden
Tel: +46-13-20 40 00 / Direct: +46-13-204327 / Mobile: +46-709 218287
E-mail: anna.anund@vti.se

Legal Disclaimer

Table of Contents

Index of Tables

Glossary

ADAS	Advanced Driver Assistance Systems
ADAS&ME	Adaptive ADAS to support incapacitated drivers Mitigate Effectively risks through tailor made HMI under automation
CDB	Central Database
DMP	Data Management Plan
DOI	Digital Object Identifier
HMI	Human Machine Interface
IDF	International DOI Foundation
M	Month
OEM	Original Equipment Manufacturer
PMT	Project Management Team
PTW	Power Two Wheeler
V2X	Vehicle to everything
WP	Work Package

Executive Summary

A preliminary Data Management Plan (DMP) was prepared for the data to be collected during the course of the project following the regulations of the Pilot action on Open Access to Research Data of Horizon 2020.
A data template was circulated to all partners in order to collect information and descriptions of data and metadata types to be collected relevant to the pilot sites, the Use Cases and those involved in defining the indicators for each state, as well as the partners involved in the development of the ADAS&ME algorithms, models and systems. The consolidated data spreadsheet (Annex 2) contains the following information:

- **Data name(s) & description**
- **Metadata name (s) and descriptions**
- **Standards**
- **Reference/compliance to existing standards**
- **Privacy, confidentiality restrictions**
- **Archiving and preservation**
- **Data under closure**

These categories will serve as the basis for creating the ADAS&ME data management repository and the metadata descriptions to accompany any dataset to be shared through or linked to this platform.
This version of the deliverable is a preliminary Data Management Plan encompassing the primary aspects to be addressed within the project with two following updates (on Month 18 and Month 40, respectively). The second update will include refined data descriptions with agreed naming conventions, dataset structures and service and interoperable standards to be applied (in agreement with the ADAS&ME architecture protocols and standards) based on the strategies, the selected indicators per affective state and final selection of sensors and systems. The third update (M40) will contain analytic descriptions of dataset structures with refined restrictions and embargos (if any) for parts/segments of data not only for the models and the systems but for the evaluation indicators per pilot site as well as the surrogate and horizontal impact and metadata indicators/estimators.
In this version the different types of the services and evaluation data are defined and the necessary aspects of the Data Management Plan framework are set:

- **Data types and categories**
- **Data privacy policy (used with the respective section of the "ADAS&ME Ethics Manual" D10.3).**
- **Data documentation**
- **Data access**
- **File naming procedures and ownership**
- **Data sharing, access and re-use**

The Data Management Plan is a deliverable directly connected to forthcoming evaluation and pilot plans for each of the pilot sites (WP7) and the decisions, issues arising in Ethics Manual (D10.3).
Post-processed datasets, free from any private/personal and identifiable information, will reside in the ADAS&ME central data repository which will be described in the next revised version of this deliverable (M18). The final version of the deliverable will include description for the complete and shareable datasets that will be created during the pilots and the analyses

to follow. In addition, the final Data Management Plan will contain the complete structure of the database, descriptions of the metadata files to enable self-explainable use of datasets by external parties.

1 Introduction

ADAS&ME ("Adaptive ADAS to support incapacitated drivers & Mitigate Effectively risks through tailor made HMI under automation") will develop adapted Advanced Driver Assistance Systems, that incorporate driver/rider state, situational/environmental context, and adaptive interaction to automatically transfer control between vehicle and driver/rider and thus ensure safer and more efficient road usage. To achieve this, a holistic approach will be taken which considers automated driving along with information on driver/rider state. The work is based around 7 provisionally identified Use Cases for cars, trucks, buses and motorcycles, aiming to cover a large proportion of driving on European roads. Experimental research will be carried out on algorithms for driver state monitoring as well as on HMI and automation transitions. It will develop robust detection/prediction algorithms for driver/rider state monitoring towards different driver states, such as fatigue, sleepiness, stress, inattention and impairing emotions, employing existing and novel sensing technologies, taking into account traffic and weather conditions via V2X and personalizing them to individual driver's physiology and driving behaviour. In addition, the core development includes multimodal and adaptive warning and intervention strategies based on current driver state and severity of scenarios. The final outcome is the successful fusion of the developed elements into an integrated driver/rider state monitoring system, able to both be utilized in and be supported by vehicle automation of Levels 1 to 4. The system will be validated with a wide pool of drivers/riders under simulated and real road conditions and under different driver/rider states; with the use of 2 cars (1 conventional, 1 electric), 1 truck, 2 PTWs and 1 bus demonstrator. This challenging task has been undertaken by a multidisciplinary Consortium of 30 Partners, including an OEM per vehicle type and 7 Tier 1 suppliers.

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. Communication of research outputs and metadata increases the potentials for inter-disciplinary collaboration among research and practitioners.

This deliverable aims to present the data sources and types per ADAS&ME systems to be evaluated and tested during the pilots and the standards and guidelines followed in order to store and communicate the findings (A preliminary compilation can be found in Annex 2). It was decided to issue another update in M18 in order to include a refined and elaborate account of data gathered during the pilots and their "integration metadata" that might result by the integration of each separate sensor and/or system to the ADAS&ME system as well as final indicators per affective state that will be incorporated to the respective algorithms. 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. Different versions of this deliverable were not defined in DoA and therefore effort used by responsible partner will be used for all three updates. These three updates will better support the identification, collection and management of data across the lifetime of the project.
The collaboration with CARTRE EU project (http://ertico.com/projects/cartre-coordination-automated-road-transport-deployment-europe/) might be an opportunity to discuss further which types of data can be shared with other communities. The deliverable concludes with an operational plan to communicate the data and meta-data agreed by partners through an online database.

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 increases are shared through an archive, 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, industry, who might not be directly involved in research but are interesting in real-life data collection, where they can search and investigate indicators to assist their empirical studies and inferences.

Deliverable D10.2 encompasses the dimensions of the ADAS&ME DMP and the methods followed to address them within the project. It additionally laid the foundations for data categorization, specification and descriptions.
Data management within the project is:
a) **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 created for harmonised data collection across pilot sites
b) **automatic** (through the ADAS&ME systems, sensors, 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.

There are three layers of data management:
a) Locally - data stored at each pilot site which may (or may not) be shared with other sites; these data are used to recruitment and identification of participants (i.e., prerequisite only for arranging follow-up sessions). These data are stored strictly at location and they additionally are stored securely. They are separated from any data that will be aggregated or consolidated for drawing inferences and used for reporting results in related WP7 deliverables.
b) Data control and management of collected and aggregated data (and later metadata) used for common data types' analyses that are categorised, anonymised, harmonised and coherent. High level data are entered only in English to ensure comprehension.
c) ADAS&ME decision center; data collected only at the ADAS&ME center and may relate to any type of addressed user, the center mechanisms (i.e. metrics of use), user clustering, user profiling. This conceptual framework is still under discussion. The conceptual description of the data center will be included at the next version of the deliverable (M18), after the finalisation of the project's architecture.
It is evident that ADAS&ME will collect very diverse data and there is fair representation of data types (e.g. sensor data, notification to various user groups, videos capturing interactions between users and in-vehicle HMI and systems, completion of digitised questionnaires). The plethora in data sources shows that at this stage we can only propose data to be collected (Annex 2). Later these data will be specified and then we will be able to cluster them in an upper level and metadata to support work to be conducted within WP7. 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.

1.1 Data management guidelines

For coherent and efficient data sharing, guidelines exist which were proposed by EC in 2013. Partners should ensure any data they share through SharePoint 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 of time (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 have to keep in mind that if they decide to open they 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. clinicians, vendors). Data are preserved and curated with interoperability in mind across groups of potential users (e.g. countries). Furthermore, data annotation and labelling enables combination and comparison with other datasets outside to the project.

2 Data processes

In order to identify and define the data management procedures, which the ADAS&ME project will follow based on the Guidelines on Data Management in Horizon 2020 document, a template was prepared and completed by the partners. This version of the deliverable is a preliminary Data Management Plan encompassing the primary aspects to be addressed within the project with two following updates (M18). The second update will include refined data descriptions with agreed naming conventions, dataset structures and service and interoperable standards to be applied (in agreement with the ADAS&ME architecture protocols and standards) based on the strategies for integration to the overall architecture. The third update (M40) will contain analytic descriptions of dataset structures with refined restrictions and embargos (if any) for parts/segments of data not only for services but for the evaluation indicators per pilot site and the surrogate and horizontal impact indicators/estimators.

2.1 Data types

Before moving on with categorising the diverse data to be collected in ADAS&ME, first, the definitions of data and metadata are provided, as they are used in this project.

> ***"Data** is a set of values of qualitative or quantitative variables; restated, pieces of data are individual pieces of information. Data is measured, collected and reported, and analyzed, whereupon it can be visualized using graphs or images. Data as a general concept refers to the fact that some existing information or knowledge is represented or coded in some form suitable for better usage or processing." [2]*

Data types can be raw, reduced, categorised, analysed, cluster in surrogates and in many other different forms. They often can be largely categorised in primary clusters of levels of details (e.g. primary, secondary, etc.). Data are categorized in *qualitative* and *quantitative*, and such categorization derives from statistics. Both data types can be *subjective* and *objective*. Neither is exclusive or inclusive. The collection of qualitative data focusses on descriptions like participant's properties, characteristics, features, etc. Quantitative data come from the world of measuring amounts, numbers, sizes, etc. Subjectivity refers to the quality of certain information (i.e. opinion) of the participants referring to experiences, feelings, beliefs, desires, perspectives, etc. Objectivity comes from the attempt of scientists to measure aspects of the natural world without any involvement of emotions, personal biases, and a priori commitments.
The following examples show the possible combinations of qualities [3]:

- **Quantitative/Objective (Q/O):** "The chip speed of my computer is 2 GHz."
- **Quantitative/Subjective (Q/S):** "On a scale of 1-10, my computer scores 7 in terms of its ease of use."
- **Qualitative/Objective (QL/O):** "Yes, I own a computer."
- **Qualitative/Subjective (QL/S):** "I think computers are too expensive."

Occasionally, an affective state might be related to more than one data type collection. Apart from data, metadata will be collected to define the characteristics and in many cases to facilitate processing, storing, and, finally, understanding the data collected during the pilots. Metadata definitions range from quality descriptions of datasets when they are used by analysts who did not participate in the data collection, thus, it is important for them to understand as much as possible about the related processes and procedures to aggregation of data to something different (e.g. values over a threshold might be translated to impairment and volume of impairment). The following definition encompasses more than one use of the term.

> *"**Metadata** is data that describes other data. Meta is a prefix that in most information technology usages means "an underlying definition or description." Metadata summarizes basic information about data, which can make finding and working with particular instances of data easier." [4]*

A template was created and circulated to partners to be completed with information and descriptions about the data and metadata that will be collected either with their systems, at the pilots, and because of potential indicators that will be selected to measure an affective state. The description remains non-technical (i.e. they do not contain information about size, exact type, etc.) as these remain to be clarified later in the project and as soon as integration has reached a functional system level. Therefore, the information provided in this version relies mainly on partner's descriptions and current agreements rather than finalised sets of indicators. The types and characteristics of metadata will alter and will be enriched during the development of the interoperable environment that will host and connect all sensors and systems.

This template aimed to collected data and information about the following:

- **Data (reference and name):** the type of data, the name and the reference they will use during collection.
- **Data description:** Brief description of data; what they are and what they represent.
- **Metadata:** To define any data to be created in order to describe and provide information about the raw data to be collected.
- **Standards:** Internal (and not only) standards they apply for collecting and processing data.
- **Reference/compliance to existing standards:** Give reference to known standards that they comply when collecting, processing, and storing data (they are meant to be universal and known standards besides internal procedures).
- **Privacy, confidentiality restrictions:** restrictions that may govern the data collection, processing and sharing related to both local, national and international data privacy and confidentiality legislation and guidelines. These dimensions are very closely related to the ADAS&ME ethics policy (see D10.2: "ADAS&ME Ethics Manual") which will include and manage any related activities. The Ethics Board will be

responsible for overviewing any data sharing protocols and procedures based on data-specific restrictions, if such apply.

- **Archiving and preservation:** techniques and procedures to categorise and storing and backing-up data. They will be refined based on the technical specification and functionalities of the ADAS&ME database.
- **Patient and healthcare relationship and closure legislation:** relevant mainly to information provided to the ADAS&ME project related to any driver's/rider's health condition and medical history. These are sensitive, private and personal data that will not be shared outside the project based on the code-of-conduct regulations and guidelines governing the established relationship between both parts (i.e. the participant and the organization that will conduct the test). These data are not anticipated and are relevant to incidental findings. The reason for including this category is because data collected are in many cases bio signals that can reveal health problems and conditions the participant might not be aware of.

There was an open-field in the template for each service responsible to customise and add information based on the specificities of each sensor, if any existed at this stage.

2.2 Sensor data

At this stage, data are linearly connected to the ADAS&ME services, whereas no ADAS&ME system data are considered that maybe the result of the operation of the system itself. This list (Annex 2) is not exhaustive but clearly shows the diversity and complexity of data sources and data categories that will be collected during the lifetime of the project. Certain adjustments and improvements may expand the number and change the qualities of recorded and transmitted data but if such changes will be implemented, they will serve as advantages to the project's scope and evaluation objectives.

The data are classified in the following categories:

- ***User profile related data***. This is a local database used to store information related to the user profiles. All data comprising a user profile will be stored locally in the end-user device. Each time a user requests a service by ADAS&ME some information about user profile will be send to the ADAS&ME system (the ADAS&ME architecture will define which management center and how).
- ***User request data***. This data contain the user input that describes his/her request for a particular service.
- ***Algorithm related data***. Data necessary that will be sourced by various sensors in order to model the state of the driver and rider.
- ***Affective state related data***. These types of data are gathered from various content providers and are relevant to the services supported by ADAS&ME. This data include the information the central architecture (as it will be defined by the ADAS&ME architecture) receives from the appropriate web services upon request.
- ***Database internal data***. This category contains all data manipulated by the database and used for internal operations.

A current preliminary list of data can be found in Annex 2. Overall, the vast majority of data are objective. This table presents the main sensors' data outputs and most of them are applications or tools.

The data types are still general and most of the partners have not adopted a common file naming system; implementation is still at a very early stage. A cautious Gestalt approach [5] is followed; the parts of the ADAS&ME system does not equal its parts. This holds true for

data types and sharing prospects. The combined use of certain services might lead to - mainly metadata - additions and modifications to general data presented in the annexed table.

2.3 Evaluation data

Moreover, a large part of data has not been defined yet and these are the data that will be developed for evaluation purposes and will be specified in the evaluation framework as well as the experimental plans. The following table (Table 1) provides an overview of data categories that will potentially be included based on current Use Case description and requirements and for the pilots' requirements, as stated in DoA. This is not an exhaustive list and mainly depicts the clusters of data the projects aims to collect. The list will be updated and refined (i.e. include standards, metadata information, restrictions placed by data owners) as soon as the "Evaluation Framework" is ready (D71.1, M18).

Table 1: *Indicative list of Evaluation Indicators (EIs)*

Indicative list of Evaluation Indicators (EIs) to be collected during the pilots
- Usage - User satisfaction (perceived/rated by users) - User compliance (following the device's advice) - User acceptance (perceived/rated by users) - Usability ratings (perceived by users)
- Efficient unobtrusive monitoring success rate - Adequacy, correctness, sensitivity, specificity, accuracy of algorithms and respective measures - Accurate and appropriate labels for - Carer effectiveness (carer ratings) - Correct effectiveness ratings - Effectiveness ratings (perceived by carers) - Tool effectiveness (perceived carer ratings) - Task efficiency (perceived/rated by carers) - Cost efficiency (users can/want to pay for re-training or aids, measured through WtP)
- Simulated performance of each of functions (technical reliability at simulated events) - Perceived security
- System efficiency improvement (perceived/rated by drivers/riders and other stakeholders)
- Sensitivity - Specificity
- Comparability between similar indicators, metrics, subjective assessments: (Attained Comparability amongst similar context indicators) - Time deviation in finalisation - Numbers of involved users for all user groups - Minimal data loss (e.g. drop outs, lost copies, recordings)
- Generalisability of results from pilot sites: for systems and metrics
- behavioural competence (e.g. measured by indicators of fatigue, cognition, workload, time use, social behaviour); - Affective and psychological factors (negative and positive affect) (e.g. measured by indicators of workload, fatigue, sleepiness, distraction, emotional state, stress, anxiety, etc.) - External, objective (physical) environment (e.g. environmental cues, real or simulated testing environment).

2.4 Data storage and back up

Data collected by the sensors and researchers at the pilots have to 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, have to 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 options 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);
- Adminstration and access requires 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 adminisration 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 infstrastructure might be needed;
- Adminstration and access requires credentials.

2.5 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 2: *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 aggegrate or analyse data utilise import/export functionalities).

2.6 File naming

File naming depends largely on service and the datasets to be derived by this service and/or connected with this service. They have to be **consistent** and **descriptive**.
The creation of the unified database will be based on a common file naming and organizing among partners in order to help partners 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.

1. **Project acronym:** ADAS&ME
2. **Driver state:** Indicate if file is related to any particular driver state (e.g. stress, fatigue, etc.)
3. **Use Case:** Indicate relation to Use Case (A, B, C, D, E, F and G)
4. **System/sensor/data type related related:** e.g. temp for temperature
5. **Location (where it resides):** e.g. CDB
6. **Researcher name/initials:** JS
7. **Pilot identifier:** e.g. SP for Spain Pilot site
8. **Date or range of pilot:** 100117
9. **Type of data:** subjective or biosignals
10. **Conditions:** control
11. **Version number of file:** 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 ADAS&ME project looks like that:

ADAS&ME_sensor_SP.csv

2.7 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.

3 Data sharing and access

At each pilot site a nominated person will be responsible for overseeing that data are safe and secure. A list of relevant EC guidelines that any data management activities should abide to can be found in Annex 1. Further information about relevant legislation and guidelines can be found also in D10.3

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. 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 or transferred to the ADAS&ME database will not include any personal or identification data. These data, of course, cannot be shared with external databases for further (re-)use.

Data collection, storing, accessing, and sharing abide to the international legislation (Data Protection Directive 95/46/EC "on the protection of individuals with regard to the processing of personal data and on the free movement of such data") and guidelines (see D10.3 "ADAS&ME Ethics Manual" 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
- Diagnosis
- GPS coordinates (only metadata or surrogates)
- Raw video and audio recordings

These data are identified based on the initial data pools set by partners responsible for Use Cases and systems. Other data restrictions might arise during the course of the project.

3.2 Data ownership

Any data gathered during the lifetime of the project are the ownership of the beneficiary or the beneficiaries (joint ownership) that produce them according to subsection 3, Art. 26 of the signed Grant Agreement (643442-ADAS&ME-H2020-PHC-2014-2015/H2020-PHC-2014-single-stage). 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 data open-access (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 affective state, therefore this information will be available in an updated version of this deliverable.

3.3 Data Sharing and Re-use

Under Horizon 2020, all publication resulting by work performed within a project have to be in open-access journals [6]. Participating in an open scholar community can help make the work of partners, and the project, more visible to researcher who works in similar disciplines and research areas. Specifically, for ADAS&ME, publishing in open-access journals is sought. Relevant dissemination activities target, organize and manage publishing efforts (WP9).

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 wishes and suggestions.

3.4 Data/meta-data repository (ADAS&ME database)

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. Election of embargo will be made for these datasets that are made available on the central ADAS&ME database.

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 licence or operating restrictions that should be considered (e.g. work only with certain h/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 for facilitating future searches. Partners will decide, and will be reported in the next version of this deliverable, for how long they would like to retain the data.

Decisive factors are different per system and partner. Any related costs for archiving and preserving data -especially for long periods of time- they will be checked for their justification and then incurred during the lifetime of the project. ADAS&ME participates in the Open Research Data Pilot of Horizon 2020 and, thus, three iterations of the Data Management Plan (DMP) are scheduled in ADAS&ME 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 catalogued 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 manner in order to foster 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 (D10.3).

3.6 Open access to journals and scientific publications

Open access addresses not only datasets and metadata produced during the lifetime of the project but also scientific publications based on project developments, outcomes and reports. Journals are open access when readers can access them for free and they are available online.
Scientific publications based on work performed in the ADAS&ME project will always acknowledge the project and the funding programme (i.e. EU Horizon 2020 research and innovation programme under grant agreement No 643442).
Scientific papers will be uploaded at the SharePoint area under the WP folder it belongs. In case of intra-WP collaboration, then the paper will be stored in the leading authors' WP. The authors are always responsible for uploading and storing their papers on the common repository. Partners can also use the Openair repository (https://www.openair.eu).
Project partners are also advised to share their publications at their institutes/ organization's website or online space, wherever this is feasible and allowed.

4 ADAS&ME data privacy policy

Participants' personal data will be used in strictly confidential terms and will be published only as statistics (anonymously).
In addition to the ethical aspects analysed, the following safety provisions will be considered during the project:

- Through a concise screening people suffering from serious psychiatric or substance abusive conditions, will be excluded from the pilots.
- In pilots that are safety related issues (e.g. real traffic conditions) all necessary precautions will be taken (i.e. use of drivers with valid driving license under normal traffic conditions, and , in case of need, using a test car with double pedals and a driving instructor as co-driver).

All pilot data will be anonymised. Only one person per site (relevant Ethical issues responsible) will have access to the relation between test participants'' code and identity, in order to administer the tests. One month after the pilots end, this reference will be deleted, thus safeguarding full anonymisation of results. This is in line with what is decided in D10.3 ADAS&ME Ethics Manual.

The stored data will only refer to users' age, gender and nationality (no other identifier will be kept). Nevertheless, stored date relate only to users' preferences in daily activities or health problems, not to a person's beliefs or political or sexual preferences.

Dynamic data will be limited to storing on pre-defined categories, (i.e. stored events will not include political events, or religious places visited), not all types of each category.

The following data will <u>not</u> be stored:

- Medical info.
- Name, address, telephone, fax, e-mail, photo, etc. of the user (any direct or indirect link to user ID).

- User location (retrieved every time dynamically by the system, but not stored).
- Any other preferences/actions by the users, except the ones motioned explicitly above.
- To whom they communicate, their frequent contacts, etc.

4.1 During pilots

During the ADAS&ME Pilot tests:

1. Drivers participating in the trials will give their names, address, and contact phone, together with age, gender, nationality and, if any, functional problem type (not medical term of impairment), to a single person in each pilot site, to be stored in a protected local database (to contact them and arrange for the tests). The contact person will issue a single Participant ID for each of them. This person will not participate in the evaluation and will not know how each user behaved.
2. The names, address and contact phone will be kept in the database only for the duration of each trial (short term trials-up to 1 week, long term trials- up to 1 month). Such data will not be communicated to any other partner or even person in each pilot site. Once the test ends, they will be deleted.
3. Each month the anonymised data will be re-sorted randomly, to mix participants order by a pre-set mechanism (i.e. easily set up with Excel).
4. Since personal data will be deleted, no follow-up studies with the same people will be feasible.
5. The partners of the consortium agree and declare that participants will not receive any additional medication, related to project research. Personal data will be used in strictly confidential terms and will be published as statistics (anonymously).

5 Conclusion and next steps

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.

Another two versions of this deliverables will be issued during the lifetime of the project (M18 and M40) with detailed descriptions of the datasets and the technical specifications and protocols of the ADAS&ME database.

This deliverable will act as a reference document and later a database manual will be included with stepwise instructions on using the different structures of the database and metadata descriptions of the datasets. 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 ADAS&ME Ethics Board, and the project management team. The final version of the Data Management Plan will also clearly state which datasets will be shared with the public and which parts (or whole datasets) will not.

The overall plan related to data management to be followed during the lifecycle of the project will be further refined with consideration for both pilot planning and respective indicators (D7.1) as well as the ADAS&ME Ethics related policy (D10.3).

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. Data definition. Retrieved 6th July 2015 from: https://en.wikipedia.org/wiki/Data
3. Definitions of objective and subjective data. Retrieved 6th July 2015 from: *http://www.userfocus.co.uk/articles/datathink.html*
4. National Information Standards Organization; Rebecca Guenther; Jaqueline Radebaugh (2004). Understanding Metadata. Bethesda, MD: NISO Press. ISBN 1-880124-62-9. Retrieved 2 April 2014.
5. Jäkel, F., Singh, M., Wichmann, F. A., & Herzog, M. H. (2016), "An overview of quantitative approaches in Gestalt perception.", Vision Research, 126: 3–8.
6. 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*)

Annex 1: 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 web-sites. 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 with regard to e-commerce. First, OECD issued Guidelines governing the protection of privacy stipulating the fundamental principles (OECD, 1980).

In 1998, OECD issued a Recommendation with regard to the implementation of the aforementioned 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 web-sites 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, provided that 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, web-sites should formally state the acceptance of full responsibility for the security and confidentiality of the personal data collected and processed. With regard to 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 is able to 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). As a consequence, 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 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 aforementioned 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 ADAS&ME. The conformance to them will be safeguarded by the PMT.

Annex 2: Preliminary completed Data Management Plan templates (per system/sensor/other)

This a preliminary DMP template for sensors and indicators from the partners involved in testing and partners involved in the development of sensors, algorithms and systems. Certain data descriptions and aspects, especially related to standards and sharing methods, are still to be defined for the later version of this deliverable (M18). Data collection is not relevant and not happening within all WPs. In the second version of this deliverable, we will correlate /match the system, with the cluster of data collected at each pilot site, something that is not feasible at this stage and not addressed in the following table. Any updated information will be added in the deliverable.

Partners will have access to an online version of this template to add at their own convenience, as soon as new information is available and a decision is reached, for each (-meta) data set to be compiled.

Table 3: *Preliminary completed DMP templates for ADAS&ME*

Systems/sensors/other	Partner	Data (reference and name)	Data description	Metadata	Standards	Privacy	Confidentiality	Archiving and preservation	Data sharing	Comments/suggestions
Navigation	TomTom	(HD-) map data	HD map information	Lane information (border, trajectory, connectivity, etc.)	TomTom HAD	none	Confidential	Stored on local device or streamed from TomTom cloud	No	Full specifications of metadata available on request. Map data will be gathered through normal company processes on request and are not exclusively available for ADAS&ME. Use of map data is restricted for use within the consortium by an NDA.
Acoustic sensor (microphone)	OvGU	Audio Signal (Raw, coded)	Digital Signal, PCM (uncompressed), MP3 (compressed)	normally no additional data recorded, but speaker ID, age and gender would be beneficial	RIFE WAVE (WAV) Fraunhofer MP3	Voice can be used to identify the speaker, therefore the name should not be stored	when anonymized (pseudonymized) it can be published, if more privacy needed, only the acoustic features can be shared	Store the uncompressed WAV data together with medatada description (xml or txt)	Internally, share complete (pseudonymized) data, externally use research license or just extracted features	additional driver information (age gender) should be stored within xml
Video		Video Signal (codec, compressed)	two possibilities: compressed sinlge images with motion	Video recording device data (diaphragm, closure time)	MPEG-4, H.264, H.265	Face can be used to identify the speaker, therefore	when anonymized (pseudonymized) it can be internally	Store the video data together with medatada description	Internally, share complete (pseudonymized) data,	additional driver information (age gender) should be stored within xml

Systems/sensors/other	Partner	Data (reference and name)	Data description	Metadata	Standards	Privacy	Confidentiality	Archiving and preservation	Data sharing	Comments/suggestions
			compression (MPEG) or (compressed single images, with i-frames only	additional data recorded, but speaker ID, age and gender would be beneficial		the name should not be stored	published, if more privacy needed, only the acoustic features can be shared	(xml or txt)	externally use research license or just extracted features	
Acoustic sensor (microphone)	OvGU	Audio_RAW,Audio_MP3	Digital Signal, PCM (uncompressed), MP3 (compressed)	normally no additional data recorded, but speaker ID, age and gender would be beneficial	RIFE WAVE (WAV) Fraunhofer MP3	anonymized using subject ID	Confidential	Store the uncompressed WAV data together with medatada description (xml or txt)	Internally, share complete anonymized data, externally use research license or just extracted features	additional driver information (age gender) should be stored within xml
webcam(s)	OvGU	Video_RAW,Video_MP4	two possibilities: compressed single images with motion compression (MPEG) or (compressed single images, with i-frames only	Video recording device data (diaphragm, closure time) additional data recorded, but speaker ID, age and gender would be beneficial	MPEG-4, H.264, H.265	anonymized using subject ID	Confidential	Store the video data together with medatada description (xml or txt)	Internally, share complete anonymized data, externally use research license or just extracted features	additional driver information (age gender) should be stored within xml
cECG Seat	RWTH	Seat_RR	Respiration rate	subject ID, age, gender, body measurements (height, weight, etc.)	timestamped csv file w. header	anonymized using subject ID	Confidential			

Systems/sensors/other	Partner	Data (reference and name)	Data description	Metadata	Standards	Privacy	Confidentiality	Archiving and preservation	Data sharing	Comments/suggestions
cECG Seat	RWTH	Seat_HB_Interval	Heart beat interval	subject ID, age, gender, body measurements (height, weight, etc.)	timestamped csv file w. header	anonymized using subject ID	Confidential	close to metadata file	anonymized data to project members only	
Steering Wheel	Autoliv	SteeringWheel_Hand_Position	Position of the hand at the steering wheel	subject ID, age, gender, body measurements (height, weight, etc.)	timestamped csv file w. header	anonymized using subject ID	Confidential	close to metadata file	anonymized data to project members only	
Steering Wheel	Autoliv	SteeringWheel_Button_States	State of the steering wheel buttons	NA	timestamped csv file w. header	anonymized using subject ID	Confidential	close to metadata file	anonymized data to project members only	
Steering Wheel	Autoliv	SteeringWheel_LED_States	RGB light event	NA	timestamped csv file w. header	anonymized using subject ID	Confidential	close to metadata file	anonymized data to project members only	
Steering Wheel	Autoliv	SteeringWheel_HB_Interval	Heart beat interval	subject ID, age, gender, body measurements (height, weight, etc.)	timestamped csv file w. header	anonymized using subject ID	Confidential	close to metadata file	anonymized data to project members only	
Steering Wheel	Autoliv	SteeringWheel_EMR	muscle activity	subject ID, age, gender, body measurements (height, weight, etc.)	timestamped csv file w. header	anonymized using subject ID	Confidential	close to metadata file	anonymized data to project members only	

Systems/sensors/other	Partner	Data (reference and name)	Data description	Metadata	Standards	Privacy	Confidentiality	Archiving and preservation	Data sharing	Comments/suggestions
Steering Wheel	Autoliv	SteeringWheel_GSR	Galvanic skin response	subject ID, age, gender, body measurements (height, weight, etc.)	timestamped csv file w. header	anonymized using subject ID	Confidential	close to metadata file	anonymized data to project members only	
Driver camera	SmartEye Pro	DriverVideo	Video signal from driver face and eyes	subject ID, age, gender, glasses, facial hair	high quality MPEG-4	anonymized using subject ID	Confidential	close to metadata file	anonymized data to project members only	
vehicle CAN bus	DLR	CAN_Raw	vehicle CAN data raw	NA	timestamped binary file w. header	anonymized using subject ID	Confidential	close to metadata file	anonymized data to project members only	
vehicle CAN bus	DLR	CAN_Data	vehicle CAN data decoded	NA	timestamped csv file w. header	anonymized using subject ID	Confidential	close to metadata file	anonymized data to project members only	
GPS antenna	DLR	GPS_Raw	GPS sensort data raw	NA	timestamped binary file w. header	anonymized using subject ID	Confidential	close to metadata file	anonymized data to project members only	
GPS antenna	DLR	GPS_Data	GPS sensor data decoded	NA	timestamped csv file w. header	anonymized using subject ID	Confidential	close to metadata file	anonymized data to project members only	
V2X communication	DLR	V2X_RAW	V2X sensor raw data: digital map,	NA	timestamped binary file w. header	anonymized using subject ID	Confidential	close to metadata file	anonymized data to project	

Systems/sensors/other	Partner	Data (reference and name)	Data description	Metadata	Standards	Privacy	Confidentiality	Archiving and preservation	Data sharing	Comments/suggestions
platform			manoeuvre of ego vehicle						members only	
V2X communication platform	DLR	V2X_Data	V2X sensor decoded data: digital map, manoeuvre of ego vehicle	NA	timestamped csv file w. header	anonymized using subject ID	Confidential	close to metadata file	anonymized data to project members only	
LCD cluster display	DLR	ClusterDisplayElements_States	State of the different HMI elements of the cluster display	NA	timestamped csv file w. header	anonymized using subject ID	Confidential	close to metadata file	anonymized data to project members only	
Smart device	DLR	SmartDevice_HMI_Elements_States	state of the smart device HMI elements	NA	timestamped csv file w. header	anonymized using subject ID	Confidential	close to metadata file	anonymized data to project members only	
Sensor: Smart Eye Pro (SEP) (multiple camera tracking system)	SEYE	Produced	SEP will collect video recording of the driver and some of the interior of the vehicle. Main data collected and processed by SEP related to: - Head position - Head rotation - Gaze	Head position is a vector in XYZ coordinate system. Measured in mm. Head rotation provides Heading (Yaw), Pitch and Roll. Gaze direction is a vector in XYZ coordinate system.	The following Metadata is not collected by the SEP, but we would like to have it collected and saved by the test leader. Test description: -Who performed	SEP uses infra-red (IR) illumination. In that regard the Smart Eye equipment complies with the appropriate international standard; IEC 62471 "Photo-biological	The Smart Eye Pro camera system will capture and store video images of the driver and some of the interior of the vehicle. We need this video together with EPFL during the project and furthermore we want to be able to use this data for our own	All data collected by the SEP is confidential and should not be made public. It can be shared with the ADAS&ME partners and only them.	SEP produces and stores two types of data: raw video file collected by the HW and log file collected and analysed by the SW. Raw video files will be stored at SEYE internal	???

Systems/sensors/other	Partner	Data (reference and name)	Data description	Metadata	Standards	Privacy	Confidentiality	Archiving and preservation	Data sharing	Comments/suggestions
			direction - Eyelid opening - Pupil diameter - Blink detection - Fixation/Saccade detection - Driver identity - Frame number - Timestamp - Tracking quality Note, full list of data collected and processed by the SEP can be found in the SEP Programmers Guide in the Appendix A on p. 51-53. It can be found at Share Point under WP4-A4.4-Sensors-Manuals.	Measured in mm. Eyelid opening is a distance between lower and upper eyelids. Measured in mm. Pupil diameter is measured in mm. Blinks. For the blink we measure the following characteristics for each eye: Closing amplitude (mm), Opening amplitude (mm), Mid Closing time (SEYE internal time stamp), Mid Opening time (SEYE internal time stamp), Max Closing speed	the study (name, company) -Date of the study - Study environment A (Car, Bus, Truck) -Study environment B (Demo, simulator, desk) -Tests conditions A (Indoor, outdoor) - Test conditions B (Sunny, cloudy, rainy, changing) Data description: -SW/HW version -Units - File size Test subject (participant)	safety of lamps and lamp systems". This standard contains the calculation rules for Maximum Permissible Exposure (MPE) that Smart Eye complies with. Under all normal operating conditions the Smart Eye system has a comfortable 40-fold safety margin relative to the MPE. WARNING! In order to	development purposes after the end of the project. We promise to not distribute the video or camera images to any external parties or publish image(s) of any test participant without his/her consent. This needs to be included in the participant consent form.		servers, while log files will be stored at the ADAS&ME common repository (developed by A4.1, UPatras).	

Systems/sensors/other	Partner	Data (reference and name)	Data description	Metadata	Standards	Privacy	Confidentiality	Archiving and preservation	Data sharing	Comments/suggestions
				(mm per SEYE internal time stamp), Max Opening speed (mm per SEYE internal time stamp). Each blink has assigned BlinkID (the same for both eyes). Blink duration can be also calculated, measured in sec. Fixations/Saccades. The have similar output as the blinks. Driver identity is represented by the set of descriptors with a unique subject ID number. General video image of the subject is processed by the FaceID algorithm and	description: -Head wear (y/n) -Glasses (none, transparent, semi-transparent, dark, dark IR-blocking) -Contact lenses (y/n) -Beard / Moustache (y/n) - Makeup (y/n) - Ethnicity - Age	prevent personal injury and guarantee that the criterions in the IEC 62471 standard are met under all conditions, an operating SmartEye flash should never be placed close to a naked eye for prolonged periods. To meet the criterions in the standard it suffices that either: -The minimum distance between				

Systems/sensors/other	Partner	Data (reference and name)	Data description	Metadata	Standards	Privacy	Confidentiality	Archiving and preservation	Data sharing	Comments/suggestions
				a subject description vector is created. Each subject can have multiple descriptors due to the different view angles. A set of new descriptors is added to the subject database together with an identifier specific to that subject. Memory amount per subject is about 300 kB. Timestamp can be defined by RTC, CPU tick, and User defined. It is measured in sec. Tracking quality is measured for Head position		the eye and the operating flash is limited to no less than 100 mm/4 inches. OR -The duration of the close range exposure of the eye is limited to a maximum of 1 minute, should the eye inadvertently be closer than 100 mm/4 inches to the flash. - Do not exceed a total of 1 minute close range				

Systems/sensors/other	Partner	Data (reference and name)	Data description	Metadata	Standards	Privacy	Confidentiality	Archiving and preservation	Data sharing	Comments/suggestions
				and rotation, Gaze direction, Eyelid opening, and other parameters. It is represented by the value between 0 and 1. Note, full list of the data description can be found in the SEP Programmers Guide in the Appendix B on p. 55-66. It can be found at Share Point under WP4-A4.4-Sensors-Manuals.		exposure (less than 100 mm/4 inches) per any 10 minute period.				
Driver identification system (A4.6)	SEYE and EPFL	Needed	Series of images capturing driver							
		Produced	Driver frontal view Driver identity							
Driver	AUTOLI	Needed	(Filtered)	Gaze direction						

Systems/sensors/other	Partner	Data (reference and name)	Data description	Metadata	Standards	Privacy	Confidentiality	Archiving and preservation	Data sharing	Comments/suggestions
state system monitoring Drowsiness/Sleepiness/Fatigue (A4.3)	V, SEYE, SU, VTI		-Gaze direction -Blinking -Eyelid opening (Filtered) -Pupil size -Head movement (direction)	is a vector in XYZ coordinate system. Measured in mm. Blinks. For the blink we measure the following characteristics for each eye: Closing amplitude (mm), Opening amplitude (mm), Mid Closing time (SEYE internal time stamp), Mid Opening time (SEYE internal time stamp), Max Closing speed (mm per SEYE internal time stamp), Max Opening speed (mm per SEYE internal time stamp).						

Systems/sensors/other	Partner	Data (reference and name)	Data description	Metadata	Standards	Privacy	Confidentiality	Archiving and preservation	Data sharing	Comments/suggestions
				Each blink has assigned BlinkID (the same for both eyes). Blink duration can be also calculated, measured in sec. Eyelid opening is a distance between lower and upper eyelids. Measured in mm. Pupil diameter is measured in mm. Head movement (position) is a vector in XYZ coordinate system. Measured in mm. Head movement (rotation) provides Heading						

Systems/sensors/other	Partner	Data (reference and name)	Data description	Metadata	Standards	Privacy	Confidentiality	Archiving and preservation	Data sharing	Comments/suggestions
				(Yaw), Pitch and Roll.						
			Heart rate							
			Respiratory rate							
			Time of the day							
			Time-on-task	Time-on-task shows total time of driving by the same driver						
			Driving performance	Driving performance refers to driving with the same speed for a long time and driving straight for a long time.						
		Produced	Driver state monitoring Drowsiness/Sleepiness/Fatigue	This driver state will have three levels characterising severity of the state.						
Vitaport II	Scania, VTI	Vita_EOG	Electrooculography (EOG)	Subject ID, age, gender, body measurements, truck driving experience,	Vitaport data file (.vpd)	Anonymous data and metadata	Public	Data and metadata cannot be stored together with test participants	Public	

Systems/sensors/other	Partner	Data (reference and name)	Data description	Metadata	Standards	Privacy	Confidentiality	Archiving and preservation	Data sharing	Comments/suggestions
				medical history, study test condition				identification key		
Vitaport II	Scania, VTI	Vita_ECG	Electrocardiogram (ECG)	Subject ID, age, gender, body measurements, truck driving experience, medical history, study test condition	Vitaport data file (.vpd)	Anonymous data and metadata	Public	Data and metadata cannot be stored together with test participants identification key	Public	
Vitaport II	Scania, VTI	Vita_EEG	Electroencephalogram (EEG)	Subject ID, age, gender, body measurements, truck driving experience, medical history, study test condition	Vitaport data file (.vpd)	Anonymous data and metadata	Public	Data and metadata cannot be stored together with test participants identification key	Public	
Vitaport II	Scania, VTI	Vita_GSR	Galvanic skin response (GSR)	Subject ID, age, gender, body measurements, truck driving experience, medical history, study test condition	Vitaport data file (.vpd)	Anonymous data and metadata	Public	Data and metadata cannot be stored together with test participants identification key	Public	
Written personal journal	Scania, VTI	SleepJournal	Sleep diary	Subject ID, age, gender, body measurements,	Microsoft word text (.doc)	Anonymous data and metadata	Public	Data and metadata cannot be stored together	Public	

Systems/sensors/other	Partner	Data (reference and name)	Data description	Metadata	Standards	Privacy	Confidentiality	Archiving and preservation	Data sharing	Comments/suggestions
				truck driving experience, medical history, study test condition				with test participants identification key		
Digital application on laptop computer	Scania, VTI	PVT_data	Psychomotor Vigilance Task (PVT)	Subject ID, age, gender, body measurements, truck driving experience, medical history, study test condition	Excel file (.xlsx)	Anonymous data and metadata	Public	Data and metadata cannot be stored together with test participants identification key	Public	
Digital application on laptop computer	Scania	MTT_data	Manual Tracking Task	Subject ID, age, gender, body measurements, truck driving experience, medical history, study test condition	unknown	Anonymous data and metadata	Public	Data and metadata cannot be stored together with test participants identification key	Public	
Digital application on laptop computer	Scania	LCT_data	Lane Change Task	Subject ID, age, gender, body measurements, truck driving experience, medical history, study test condition	Lane change test data export file (.txt)	Anonymous data and metadata	Public	Data and metadata cannot be stored together with test participants identification key	Public	
Digital questionna	Scania, VTI	KSS_data	Karolinska Sleepiness	Subject ID, age, gender,	unknown	Anonymous data and	Public	Data and metadata	Public	

Systems/sensors/other	Partner	Data (reference and name)	Data description	Metadata	Standards	Privacy	Confidentiality	Archiving and preservation	Data sharing	Comments/suggestions
ire on mobile device			Scale	body measurements, truck driving experience, medical history, study test condition		metadata		cannot be stored together with test participants identification key		