This project has received funding from the European Union's Horizon 2020 research and innovation programme under Grant Agreement No 814761

D1.2
Data Management Plan

Safe tolerance zone calculation and interventions for driver-vehicle-environment interactions under challenging conditions

iDREAMS

Project identification

Grant Agreement No	814761
Acronym	i-DREAMS
Project Title	Safety tolerance zone calculation and interventions for driver-vehicle-environment interactions under challenging conditions
Start Date	01/05/2019
End-Date	30/04/2022
Project URL	www.idreamsproject.eu

Document summary

Deliverable No.	1.2
Deliverable Title	Data Management Plan
Work Package	WP 1
Contractual due date	31 October, 2019
Actual submission date	30 October, 2019
Nature	ORDP
Dissemination level	Public
Lead Beneficiary	UHASSELT
Responsible Author	Muhammad Adnan
Contributions from	Tom Brijs, Edith Donders, Elke Hermans

Revision history (including peer review & quality control)

Version	Issue date	% Complete[1]	Changes	Contributor(s)
1.1	10-09-2019	20%	Contributions per partners collected	All Consortium Partners
1.2	10-10-2019	100%	1st draft sent to internal reviewers	Muhammad Adnan (UHASSELT)
1.3	28-10-2019	100%	Revised version after peer review	Muhammad Adnan (UHASSELT)

Disclaimer

Copyright

[1] *i-DREAMS quality assurance process:*
Due date – 3 months: ready for internal review
Due date – 2 months: start review
Due date – 1 month: end review
Due date – 5 days: ready for submission / coordinator submits

Table of contents

List of Figures

List of Tables

Glossary and abbreviations

Word / Abbreviation	**Description**
API	Application programming interface
i-DREAMS	Smart- Driver and Road Environment Assessment and Monitoring System
DMP	Data Management Plan
FAIR	Findable, accessible interoperable and re-usable
GPS	Global Positioning System
OBU	On-board Unit
OBD II	On-board Diagnostics 2nd Generation
EU	European Union
CAN	Controller Area Network
QTMM	Quantitative tabular data with minimal metadata
GEOS	Geospatial data
DVSS	Digital video data / simulator scenarios
REST	Representational state transfer
QTEM	Quantitative tabular data with extensive metadata
QUAL	Qualitative data

1 Introduction

The purpose of deliverable **1.2 – Initial Data Management Plan (DMP)** is to describe the data management life cycle for the data to be collected, processed and/or generated by the i-DREAMS H2020 project. As part of making research data findable, accessible, interoperable and reusable (FAIR), the project's DMP includes information on the handling of research data during & after the end of the project; what data will be collected, processed and/or generated; which methodology & standards will be applied; whether data will be shared/made open access; how data will be curated & preserved (including after the end of the project).

iDREAMS advocates for the effectiveness of openness and sharing, hence we strive to make data collected during the project as available as possible within the limit of personal privacy following the Fair Data Principles. The project will collect the following key types of data:

- Quantitative data generated via on-board devices (Mobileye, CardioWheel system, OBD Units, GPS devices) installed in vehicles
- Video data generated via dashboard cameras
- Quantitative data generated from the driving simulators
- Quantitative data generated from smartphone application
- Qualitative and quantitative data on levels of participation and user experience and opinions collected programmatically by the project website or via interview and questionnaires

These types of data have different characteristics and serve different purposes in relation to achieve project objectives. Almost all types of dataset are characterised as personal data that can be potentially linked to a person or a person's vehicle or location. In order to make this data publicly available, it will be aggregated in order to guarantee the anonymization. Databases will be deposited in recognized, international data repositories so that data will continue to be available. This Data Management Plan will be updated regularly thereafter with the purpose of supporting the data management life cycle for all data that will be collected, processed or generated by the project.

1.1 Deliverable structure

The structure of this deliverable, based on the respective EU guidelines[2] is as follows: Section 2 provides a data summary addressing issues regarding the purpose of the data collection/generation and its relation to the objectives of the project, the types and formats of data generated/collected during the project, the data origin, the expected data size, and the data utility (i.e., to whom might it be useful). Section 3 is about fair data, i.e. about making data findable (metadata, naming conventions, search keywords, versioning, etc.), openly accessible, interoperable, and re-usable. Section 4 addresses the allocation of resources. Section 5 presents provisions in place for data security. Section 6 discusses the ethical and legal aspects related to the data involved in the project. Section 7 discusses maintenance plan for DMP. Section 8 concludes the report.
`

[2] H2020 Online Manual > Cross-cutting issues > Open access & Data management (URL: http://ec.europa.eu/research/participants/docs/h2020-funding-guide/cross-cutting-issues/open-access-datamanagement/data-management_en.htm).

2 Data Summary

2.1 Project objectives and required data

The following paragraph is taken from the Grant Agreement document that summarizes the project and provides a fair idea about the objectives of the project along with some methodological steps that will be followed to achieve those objectives.

*The objective of this project is to setup a framework for the **definition, development, testing and validation of a context-aware 'safety tolerance zone' for on-road driving**, within a smart **Driver and Road Environment Assessment and Monitoring System (i-DREAMS)**. This 'safety tolerance zone' refers to a context-sensitive and dynamic zone in which the driver is within acceptable boundaries of safe operation, and thus not in immediate risk of a crash. The calculation of this zone happens on a continuous real-time assessment by monitoring the driver and environment, taking into account, on the one hand, **driver-related background factors** (e.g. age, driving experience, safety attitudes and perceptions, etc.) and **real-time risk-related physiological indicators** (e.g. fatigue, distraction, stress, etc.), and on the other hand, **driving task-related complexity indicators** (e.g. time of day, speed, traffic intensity, presence of vulnerable road users, adverse weather, etc.). Moreover, **safety-oriented interventions will be developed to prevent drivers from getting too close to the boundaries of unsafe operation and to bring back the driver into the safety tolerance zone**. The interventions will be developed **based on information received from the driver risk profile** (such as stress level, fatigue stored on the smartphone) as soon as the driver enters the vehicle. These interventions will be composed of both **real-time interventions** (i.e. in-vehicle while travelling) **and post-trip interventions** (i.e., aimed at enhancing the knowledge, attitudes, perceptions and behavioural reaction of drivers with respect to safety-related technologies, situations and behaviours). Application areas will include: new road safety interventions, improved driver well-being and transfer of control between human and vehicle. Initial testing will take place in a driving simulator environment, after which promising interventions will be tested and validated under real-world conditions in **a testbed consisting of 600 drivers in total across 5 EU countries** (Belgium, UK, Germany, Greece, Portugal). Market roadmaps will be developed to support smooth transition of the investigated technologies to the market and experience from use cases in different European countries will be used to disseminate best practices.*

The above summary of the project clearly illustrates that assessment and monitoring system that will be developed within i-DREAMS intrinsically follows a data-driven approach. A variety of datasets will be collected and then analysed in a modelling framework to determine/predict the real-time status of the driving vehicle in relation to its safety, based on the definition of a safety tolerance zone. Primarily, this is achieved based on the combination of some threshold values for key safety indicators either from a rule-based or stochastic model. The post-trip interventions part of the project is also data-driven, where trajectories, events (e.g. harsh brake) and types of warnings generated based on driver's behaviour in a safety tolerance zone will be recorded, and processed information will be given to the drivers as a feedback in a gamification manner to promote/encourage safe driving behaviour. To develop such a system and test its efficacy, experiments (off-field/on-field) will be conducted in 5 EU countries across various modes (i.e. trucks, Cars, Bus and Train). The link with the types of data to be collected in the project is therefore quite evident and will be elaborated in the next few sub-sections.

2.2 Dataset types

This section describes the current datasets typologies forecast to be generated during the project, This DMP focuses mainly on the data types, that are collected during the project as raw datasets from instruments and other methods. This section will be updated periodically as

part of the DMP life cycle, that will include details in relation to results (i.e. processed data) generated using a variety of analysis methods.

2.2.1 Quantitative data generated via on-board devices (Mobileye, CardioWheel system, OBD Units, GPS device) installed in vehicles

In order to collect a range of vehicle and driver-related driving attributes, the project uses a CardioID system developed and marketed by CardioID[3] (a technology company, founded in 2014). CardioID is a key partner in i-DREAMS consortium, and is leading WP4 of the project, where the technical implementation is carried out in relation to on-field trials. The CardioID system is available to consortium partners who are responsible for conducting on-field trials.

Based on the *CardioID System API document*[4], this system is composed of several devices that are marketed by CardioID and used for collecting data and then implementing real-time interventions. These are as follows:

- Mobileye in conjunction with DashCam (installed on the windshield)
- CardioWheel system (installed on the steering wheel)
- Connection with On-board vehicle diagnostic units
- GPS device

These devices are connected with a gateway that gathers and centralises information from other components and handles data connectivity and transmission. Additionally, it is capable of local, off-line data processing, and provides an output for a buzzer or a heptic engine. The gateway also has a capability of Ethernet, wifi and mobile data connectivity. Each gateway is conceptually tied to a vehicle (not a specific user), with data being acquired within a trip session. A trip session is defined from the moment the vehicle is turned on until it is turned off. There is a grace period (5 minutes) during which a quick turn off and back on is considered the same trip. The collected data goes through some usual automated processes to remove data noise (i.e. detection of outliers and their removal) and stored in CardioID cloud servers (subject to availability of required connectivity protocol). CardioID also provides a web API to support data access within the i-DREAMS project. The API follows the REST style, based on JSON data (available only for consortium partners).

More details on the quantitative type of data generated by CardioID systems are shown in table 1. The dataset label abbreviation in table 1 can be read as follows: DS stands for a dataset, PD stands for Personal data, QTMM stands for Quantitative data with minimal metadata, GEOS stands for the geospatial dataset and last three digits represents the number of a dataset in the project data library.

Table 1: Quantitative data from the CardioID technology

Datasets Label	Data title	Type of data (Link to Project objectives)	Source	Dataset description
DS_PD_QTMM_005	CardioWheel - Raw ECG	Driver Mental States	CardioWheel	Periodically sampled electrocardiographic signal (1000 Hz, 2 data columns)
DS_PD_QTMM_006	CardioWheel - Inter-beat Intervals	Driver Mental States	CardioWheel	Sequence of time intervals between consecutive heart beats

[3] https://www.cardio-id.com/about
[4] Confidential Document (Only available for Consortium Partners)

DS_PD_QTMM_007	CardioWheel - Hands-On Detection	Driving behaviour	CardioWheel	Event indicating detection of driver's hands on the steering wheel
DS_PD_QTMM_008	CardioWheel - Raw Motion	Driving behaviour	CardioWheel	Periodically sampled signals from inertial unit (200 Hz, 9 data columns)
DS_PD_QTMM_009	CardioWheel - Driver Change Detection	Other	CardioWheel	Event indicating a driver change
DS_PD_QTMM_010	CardioWheel - Fatigue Detection	Driver Mental States	CardioWheel	Event indicating driver fatigue level
DS_PD_QTMM_011	Mobileye - Raw CAN messages	Contextual data	Mobileye or equivalent	Event indicating driver fatigue level
DS_PD_QTMM_012	Mobileye - Warning System	Driving behaviour	Mobileye or equivalent	Decoded Mobileye events
DS_PD_GEOS_013	CardioID Gateway - GNSS	Contextual data	GPS Tracker	Satellite-based geolocation data
DS_PD_QTMM_014	CardioID Gateway - Raw vehicle CAN messages	Vehicle characteristics	OBU/OBD II	Raw CAN bus vehicle data

2.2.2 Video data generated via dashboard camera

Apart from the quantitative data, the CardioID system also provides video data generated from its dash cam which works in support to the Mobileye component. This dash cam is currently triggered by events from Mobileye, in order to better understand, in a post- trip analysis, what led to the event in consideration (e.g. a forward collision warning). The prime objective of collecting such video data is to visualize the real scenario on road because of which a warning (i.e. event) is generated by the Mobileye system. This is important because Mobileye can sense situations causing the system to generate a warning where surrounding vehicles are the main cause of the occurrence of such events. For example, the driver has kept a safe distance from the forward vehicle; however, another vehicle suddenly came in between the subject vehicle and forward vehicle by changing its lane which results in the reduction of the distance/time headway. For developing post-trip interventions in gamification environment such types of situations need to be identified so that appropriate information is sent back to the driver to make the post-trip intervention more effective. Video clips are recorded when specific events occur, and it provides the situation just before, during and after the event. Each video clip size is around 3 MB, and on average is known from past experience of the CardioID that around 50 clips are recorded per 100 km driven by the vehicle. Table 2 summarizes this information. These video clips can be accessed from the CardioID cloud servers via provided web API (only available to consortium partners).

Table 2: Video data from CardioID dashcam

Datasets label	**Data title**	**Type of data (Link to project objectives)**	**Source**	**Dataset description**
DS_PD_DVSS_015	CardioID Gateway - Dash Cam (triggered on specific events)	Contextual data	Camera	Video clips of road context recorded triggered by

				certain events from Mobileye

2.2.3 Quantitative data generated from driving simulator

The driver and road environment assessment and monitoring system that will be developed in this system will undergo a testing phase using driving simulators. Special scenarios will be developed within a simulated environment to test a variety of situations to assess the performance of the preliminary system. Along with the usual driving simulator system, the CardioID system may also be integrated within a driving simulator for real-time intervention testing. Drive Sim Solutions (DSS), a technology provider partner within i-DREAMS consortium will setup identified scenarios and provide driving simulators for consortium partners to conduct these experiments. These driving simulator experiments will be conducted in 5 EU countries (Belgium, Germany, Greece, Portugal, and the UK). A range of datasets that will be generated from these experiments will be stored in partners internal servers. Table 3 provides a list of foreseen datasets generated during the driving simulator experiments (apart from data generated by the CardioID system), with their brief description.

Table 3: Driving simulator-based datasets

Datasets label	Data title	Type of data (Link to Project objectives)	Dataset description
DS_PD_QTEM_002	Raw Data logged from Driving Simulator	Driving behaviour	Data logged from simulator for different participants. Variables included are time/distance headways, speed, lateral movement, acceleration at each time interval, special event information
DS_PD_QTEM_003	Raw Mobileye data from DSS simulator	Driving behaviour	Mobileye Data logged in DSS simulator for different participants
DS_PD_QUAL_004	Raw Cardiowheel data from DSS simulator	Driver Mental States	Cardiowheel Data logged in DSS simulator for different participants
DS_PD_DVSS_031	Driving simulation scenarios	Contextual data	These are different driving simulator scenarios for testing i-DREAMS real time intervention system

2.2.4 Quantitative data generated from smartphone application

OSEVEN PC (OSEVEN Single member private company) is another technology provider within the i-DREAMS consortium. OSEVEN will provide a state-of-the-art android-based

smartphone application that also monitors and collect driving behaviour of individuals using a variety of parameters. The app will use different smartphone sensors to collect such data. The app will be used by drivers recruited for on-field trials. Drivers recruited for the experiments will be provided with smartphone that already contain this app. Within the i-Dreams project, android-based smartphone will be purchased and provided to those partners that are responsible for carrying out these experiments. The datasets collected from this app and the CardioID technology will be fused together for more appropriate prediction of driving behaviour and also to test and validate the performance of i-DREAMS real-time platform. OSEVEN would provide an API to i-DREAMS partners to access these datasets. The smartphone app will also have a functionality to provide post-trip intervention via a feedback mechanism. Datasets collected using this app are listed in table 4.

Table 4: Smartphone app-based datasets

Dataset labels	**Data title**	**Type of data (link to project objectives)**	**Data description**
DS_PD_QTMM_019	Date/Time	Driving behaviour	The recorded date/time/timestamp as Primary data, to associate it with other variables to facilitate analysis
DS_PD_QTMM_020	GPS Data	Driving behaviour	Geographic longitude, latitude and altitude of the vehicle position, horizontal and vertical accuracy of the GPS recording, vehicle movement speed, vehicle movement direction.
DS_PD_QTMM_021	Values of the angles formed by the local axes of the smartphone to the North and to the horizontal plane (ground)	Driving behaviour	Values of the angles formed by the local axes of the smartphone to the North and to the horizontal plane (ground).
DS_PD_QTMM_022	Rate of change of the angles formed by the local axes of the smartphone to the North and to the horizontal plane (ground) versus time.	Driving behaviour	Rate of change of the angles formed by the local axes of the smartphone to the North and to the horizontal plane (ground) versus time
DS_PD_QTMM_023	Accelerometer data	Driving behaviour	Acceleration values on the three local axes of the smartphone, including and excluding the acceleration of gravity
DS_PD_QTMM_024	Gyroscope data	Driving behaviour	Angular velocity values on the three local axes of the smartphone.
DS_PD_QTMM_025	Activity Data	Driving behaviour	Data of the user's activity (indicatively but not limited to, walking, stopping, driving)
DS_PD_QTMM_026	Screen State	Driving behaviour	Data on screen activation (activated or deactivated) as an additional criterion for identifying mobile use. It is noted that OSEVEN does not have any access to the content of the smartphone screen.
DS_PD_QTMM_027	Smartphone device data	Contextual data	It is provided by Google and includes indicatively but not limited to, the manufacturer's brand, the device model, the name and version of the operating

			system and the type of smartphone sensors (e.g. accelerometer, gyroscope, compass, etc.). This data is used to (i) for the support of the User and (ii) to resolve any technical malfunctions / bugs.
DS_PD_QTMM_028	Push Notification Token Data	Other	A unique alphanumeric code produced by Google, which is sent to a smartphone. This code is associated with a single and only installation of the Application

2.2.5 Qualitative and quantitative data on levels of participation and user experience/opinions

Apart from data generated by technological equipment, a pre- and a post-experiment questionnaire survey will also be conducted to obtain driver socio-demographic information and drivers' driving attitudes and feedback in relation to their experience during on-field and off-field experiments. These data provide a meaningful base to conduct analysis to test the performance of the developed system and then improve it before conducting on-field trials. In addition to this contextual information interview-based survey, a few more questionnaire surveys are also planned within the i-DREAMS project. These are online questionnaire-based surveys to get expert/stakeholders opinion to get insight on current state-of-the-practice and the possible advancements suggested by an informed community; these can be incorporated within the development of the i-DREAMS platform. Table 5 presents details of such datasets.

Table 5: User/Participant opinions and information-based datasets

Data Labels	Data Title	Data Type (Link with Project objectives)	Source/ Method	Data Description
DS_PD_QTMM_029	Survey/Questionnaire data from drivers (field trials/simulation experiments	Socio-demographic/health/personality related data	Interview	This data contains subject contextual information (drivers in on-field trials, individuals participating in simulation trials
DS_OT_QUAL_001	Survey/Questonnaire regarding post trip intervention (Companies)	User/Stakeholders Opinion	Web-Platform based questionnaire	This data contains opinions of certain App providers about post trip intervention modalities, what measure they are using, what behaviours they are targeting, to actually know what is the state of the practice in this domain
DS_PD_QTMM_016	Survey response data-for i-DREAMS system priorities	User/Stakeholders Opinion	Web-Platform based questionnaire	Survey data giving opinions on priority areas to address with i-DREAMS system for various transport modes

Apart from the datasets discussed in the above sub-sections, if required, more equipments can also be used to collect further driving behaviour variables such as the eye-tracking movement of drivers while driving etc. More details of such datasets will be provided in the next update of DMP.

3 FAIR Data

3.1 Making data findable, including provisions for metadata

According to FAIR principles [1], to be findable the following requirements need to be met
a) Metadata are assigned a globally unique and eternally persistent identifier
b) Data are described with rich metadata
c) Data (and metadata) are registered or indexed in a searchable resource
d) Metadata specify the data identifier

For the fair access of the project data, a Digital Object Identifier (DOI) will be requested for each artefact. In more details, DOIs from Crossref[5] will be used for research publications, while DataCite metadata schema[6] will be pursued for labelling each dataset of the project that is made available to the public. In addition, a metadata record for each output of the project will be created and stored in the data directory. Amongst other fields, each metadata record will have a set of keywords that will make searches easier for external parties. The data labels as mentioned in tables 1 to 5 are currently being used to identify datasets in the project data library; this will be further extended, so that it can be easily findable. Each data source will be provided with a specific name that is composed by different parts/elements, containing information about data nature, its type, country, transport mode, data type or format and naming structure.:

I-DREAMS_CC_IP_datakind_FORMAT_EXT_FROM:date_TO:date_V1.0

- I-DREAMS: Project initials
- CC: The first letters (three max) of the country where data collected (BE, DE, EL, PT and UK)
- IP : The initials for the partners who have collected the data (i.e. data Controller) (e.g. TUM for Technical University Munich, UH for Universiteit Hasselt
- datakind: The kind of data, The initials for *Data type* as mentioned in table 1 to 5
- FORMAT : Initials for the data format defined in the project (such as QTMM, GEOS, QUAL etc.)
- EXT: Data file extension such as CSV, JSON
- FROM: and TO: fields can be used for specifying the dates that the data are valid.
- V1.0: The version of the dataset.

The data standards related to project activities might vary depending on partners' internal tools and methodologies. However, once certain datasets are publicly available, they will always follow a CSV tabular data standard or JSON format. As part of the project's open data access and sustainability strategy (after the project end), data will be programmatically downloaded as CSV files and metadata will be added following a standard. To ensure the same standards towards consistency and usability are met, all datasets should provide the following metadata as mentioned in the DataCite metadata schema. The list of mandatory elements is as follows:

- Identifier (with mandatory type sub-property)
- Creator (with optional given name, family name, name identifier and affiliation sub-properties)
- Title (with optional type sub-properties)
- Publisher
- Publication Year
- Resource Type (with mandatory general type description sub property)

[5] https://search.crossref.org/
[6] http://schema.datacite.org/

Complete guidance over what exactly is provided in different elements of metadata that are mandatory, recommended and optional within the schema will be provided in the next version of DMP.

3.2 Making data openly accessible

A number of datasets that will be used as part of the project will be offered by third-party providers. Some of these datasets are already open to the public, while others are characterised as personal data and have high sensitivity. In the cases where personal data are processed and aggregated (e.g. as part of a model, or functionality of a component), permission will be requested by the provider prior to making the altered data publicly available. In reference to the nature of the user data involved, some of the results that will be generated by each project phase will be publicly available. As per our Ethics commitment, the access and sharing of data will be rigorously implemented in compliance with the privacy and data collection rules and regulations, as they are applied nationally and in the EU. The project life cycle and flow of data can be seen in Figure 1. Some initial processing of raw collected data will be made at CardioID platform, partner's internal servers and OSEVEN platforms, before it will be integrated and stored in a unified big data fusion platform. More details on this will be provided in the next update of the DMP.

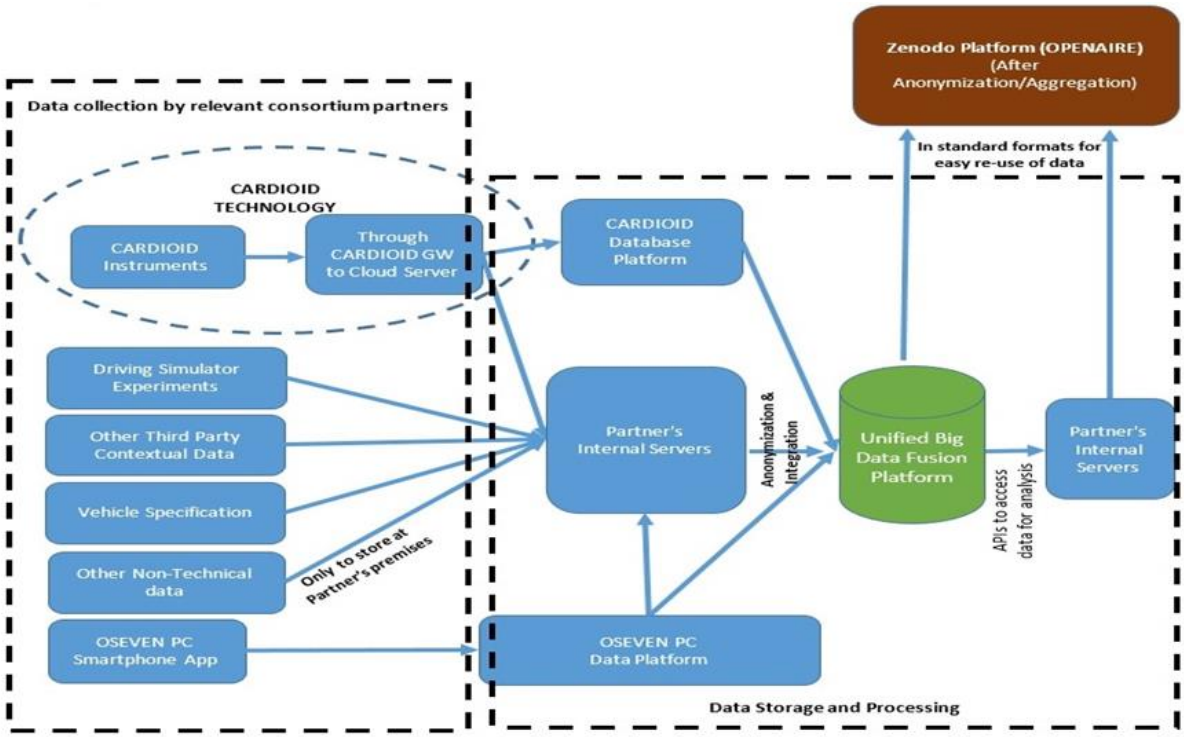

Figure 1: i-DREAMS Project data Life Cycle

In order to reach out to the research and professional community, data will be made available on the Zenodo[7] digital repository following the details specified on the Standards and Metadata in the above section. This includes all of the project data that will eventually be publicly available. This protocol aims at guaranteeing consistency on the use, reference and preservation of the sources, especially when the project finishes. As described above data, archiving and sustainability will be guaranteed by the Zenodo digital repository. As a European

[7] Zenodo was launched at the CERN Data Centre in May 2013 with a grant from the European Commission with a special commitment to sharing, citing and preserving data and code. As a digital repository, Zenodo registers DOIs for all submissions through DataCite. The platform is based on the Invenio open-source software, Zenodo profits from and contributes to the foundation of code used to provide Open Data services to CERN and other initiatives around the world.

Commission supported initiative and technically supported by CERN, we trust this as the best way to ensure access to the generated data remains long after the project ends.

3.3 Making data interoperable

Making data interoperable, means that data exchange and re-use between researchers, institutions, organisations, countries, etc., should be available. The main goal is to facilitate the re-combination of the data produced with different datasets from other origins. In order to do so, the use of standard formats and of available (open) software applications is promoted.

In the i-DREAMS project, the pathway to make data interoperable will be to include the metadata in the project data library. Other researchers within the project will be able to use this information for different research activities. Furthermore, when a dataset becomes publicly available, it can help researchers outside of the project to use and combine the data. In addition, the file extensions and format of the publicly available data followed either CSV or JSON, which are widely used file extension/ formats especially in the domain of road safety.

3.4 Increase data re-use

The intention of the i-DREAMS project is to make as much data as possible re-useable for third parties. The restriction will only apply when privacy, IPR, or other exploitation grounds are in play. All datasets will be cleared of bad records, with clear naming conventions, and with appropriate metadata. All data generated and collected in the project will undergo a quality check in order to analyse its individual plausibility and consistency, making sure that others can directly use it to perform assessments and validate the research carried out by the i-DREAMS project. As in some cases, similar results will be generated for different case studies, data harmonisation will also be of critical importance both for increasing data re-use in general, but also to ease the comparison of i-DREAMS results.

As a default standard, the project will use Creative Commons Licenses for public data sharing. In particular, Attribution 4.0 International[8] will be used to support third party researchers the right to redistribute the material in any medium or format and build upon the material for any purpose, even commercially. However appropriate credit must be given to the original source, provide a link to the license, and indicate if changes are made.

[8] A Creative Commons (CC) license Attribution 4.0 International license file https://creativecommons.org/licenses/by/4.0/

4 Allocation of Resources

This section addresses resources required to make arrangement for FAIR.

Most of the data collection, preservation, curation and its management activities are already budgeted within consortium through different project task and activities. Every partner is responsible for the data they produce. Data collected from on-field trials/simulator experiments/or other datasets, respective partners store and archive such data in their premises and are responsible to manage/secure such data. In addition, data cleaning and making it ready for processing or further analysis is also part of data collection. The i-DREAMS project formed a Data and Knowledge Management Committee (KDM) that include WP leaders, SMEs and Task 3.5 Contributors. The committee will discuss periodically progress in aspects of data collection and its processing and production of results in desired data formats.

Task 3.5 within the project will formulate procedures and logistics for big data handling. Standard i-DREAMS procedures will be written in this task to meet the legal and ethical requirements of collecting, handling and storing such data. In addition to this, task 5.5 will result in the development of a unified Big data fusion framework that will integrate collected data in the project and provide a mechanism for its processing and analysing for extraction of required knowledge through the data-information-knowledge cycle.

UHASSELT as a project coordinator and task leader for DMP has an overview of all datasets collected within the project and will timely request partners towards commitments made for **FAIR.**

In addition to the above, the following items will be further resolved within the consortium.

- Fees associated with the publication of scientific articles containing project's research data in "Gold" Open access journals: The cost of sharing, in case of multiple authors, shall be decided among the authors on a case-by-case basis. For 'Green' open access, consortium partners would make necessary steps to provide a copy of the paper that will be accessed through the project website.
- Project Website operation: to be determined
- Data archiving at Zenodo OPENAIRE plat: free of charge
- Copyright licensing with Creative Commons: free of charge
- Any fee incurred for Open Access through scientific publication of the data will be the responsibility of the data owner (authors) partner(s) in compliance with the CA
- During the Project and for a period of four (4) years after the end of the project, the dissemination of own results by one or several parties including but not restricted to publications and presentations, shall be governed by the procedure of Article 29.1 of the Grant Agreement.

5 Data Security and Protection

5.1 Scope within the project

i-DREAMS involve data sources that require special attention in terms of protection at various levels. Dealing with personal and potentially sensitive information, either collected by partners using the technology provided by CardioID, OSEVEN PC and DSS or collected through ad hoc interaction with individuals (interviews, living labs, etc.) might pose risks for the privacy of individuals: GPS tracker data of private vehicles can potentially be exploited by malicious agents to infer sensitive information (mainly related to locations), including the identity of some of the anonymous users; data from smartphone app also provides geographical location of individuals; data in relation to specific vehicle characteristics/specification in some cases provides links to individual identities; finally, user off-field trial sessions and living lab experimentations might directly provide personal information.

5.2 Data security

To meet standard security requirements in dealing with sensitive and valuable data, the project will apply at least the following basic measures:

- Data storage in a safe location, with physical access, limited to authorised personnel for hardware / software maintenance and partners of the project
- Safe data transfer through secure, encryption-protected connections
- Remote access provided through secure, encryption-protected connections, granting authorisations only to personnel and activities relevant to the project, and within the time frame of the project
- Close monitoring of accesses to the Big data unified platform, including access to data. Based on that, and in compliance with GDPR guidelines. Procedures will be set for timely notification of any kind of harmful data breach to the supervisory authorities.
- Dissemination activities might rely on several third-party platforms mostly operated by US companies (e.g., Google Analytics, MailChimp, Facebook). In order to comply with the EU data protection standards, the companies have to certify their compliance with the EU-U.S. Privacy Shield Framework[9] . Before any of the following services are used within the project, compliance will be verified at the Privacy Shield list[10].

5.3 Data protection and compliance to GDPR

For all project partners involved in the processing of personal data, relevant data protection measures should be applied as prescribed by the General Data Protection Regulation (2016/679)[11]. These may include, depending on each case, either collection of consent forms by the individuals concerned and/or registration with the appropriate (according to the place of establishment) Data Protection Authority (DPA) and/or adherence to the i-DREAMS privacy policy[12].

Personal data will be secured or made inaccessible via transformations (i.e. encryption). All sensitive information, either for individuals or for businesses (i.e. disclosing information that

[9] Commission Implementing Decision (EU) 2016/1250 of 12 July 2016 pursuant to Directive 95/46/EC of the European Parliament and of the Council on the adequacy of the protection provided by the EU-U.S. Privacy Shield http://eur-lex.europa.eu/legal-content/EN/TXT/?uri=uriserv%3AOJ.L_.2016.207.01.0001.01.ENG

[10] List of companies in compliance with the EU-U.S. Privacy Shield Framework https://www.privacyshield.gov/list

[11] Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation) (Text with EEA relevance); https://eur-lex.europa.eu/eli/reg/2016/679/oj

[12] https://idreamsproject.eu/wp/privacy-policy/

might directly or indirectly cause financial damages and/or damages of other nature) will be made inaccessible.

Systems will be designed applying, to the greatest extent possible, an architecture adheres to the general principles of data protection by design, privacy by design, privacy by default. This will include limiting actual or potential access to personal data to only those parties that need to access or process data for project-relevant activities, applying pseudonymisation operations (such as tokenization or data encryption) or complete anonymization, so that personal data is not visible to parties which are not authorized or do not strictly require it.

Partners responsible for collecting the data on-field/off-field will act as data controllers and may work together with other partners who will be data processors. The role of each consortium partner will be explicitly defined in relation to a particular dataset and the project dataset library (available only to the consortium) will keep a record of it. Hasselt University as a project co-ordinator is working on drafting an agreement for joint processing of personal data. The agreement will be signed by all consortium partners. The details of this agreement will be provided in the next update of the DMP. This agreement will provide a legal framework (following the guidelines of GDPR), under which roles of the consortium partners will be defined, such as a separate data controller, joint data controllers and data processors. Furthermore, it will also provide details of the retention policy of the personal data (for each type of data sets) and the steps that will be followed to secure datasets. All processing activities involving personal data will be monitored and recorded, providing information about the purposes of the processing, the agents involved and the time period of the activity. The records will be kept and made available upon request to the supervisory authority. i-DREAMS Data and Knowledge Management Committee (KDM) will regularly monitor the implementation of this framework in its soul and spirit.

Other basic rights granted by the GDRP, namely the right of personal access, erasure and portability of data are delegated to the corresponding data providers, taking care of aligning and processing the data accordingly, e.g. removing specific information that the user asked to be erased.

5.4 Personal data and pseudonymisation / anonymisation

Pseudonymisation means that an individual can still be identified through indirect or additional information. This means that **pseudonymised** personal data is still in the GDPR scope. **Anonymisation** means that you cannot restore the original information, and such data is out of the scope of the GDPR.

In the context of the i-DREAMS project, data analysis will be done by processing pseudonymised data as it might be interesting at a later stage of the analysis to know more attribute of the person, which require linking of a particular dataset (driving behaviour) from another dataset (socio-demographic attributes). This link can only be established if both datasets are pseudonymised. However, necessary data security and protection steps should be made so that such processing is done without misuse of such data.

5.4.1 Consistent replacement[13]

As a project standard, for data that contains individual identity information (such as name, e-mail addresses, Phone numbers, Vehicle ID) based on interview or questionnaire surveys or other dataset collected, the pseudonymisation method known as directory replacement [2] will be followed. The technique is simple and straightforward to implement. Consistent replacement means that we modify data about the registered, while there is still a link between the values. For example, we can use a customer number to identify an individual and store

[13] https://www.24solutions.com/en/blog/pseudonymisation-anonymisation-personal-data-difference/

information that directly identifies an individual, such as a personal identification number, separately. In this way, we pseudonymise sensitive data.

5.4.2 GPS trajectories anonymization

The i-Dreams project also collects raw data from CardioID technology and OSEVEN PC smartphone application. During the project life cycle, such data will provide useful information; however, if at a certain stage to this data is publicly available, these trajectories need to be anonymized especially significant halting points in the trajectories that represent key locations (such as home and work activity locations). A very simplistic algorithm can be used to delete/remove such location information from the trajectories. It involves identifying significant or non-significant halting points based on some threshold value of the duration of halt (e.g. more than 15 minutes). Once halting points are known either significant or non-significant, a zone (rectangular area) is created on the trajectory data containing a specified number of significant and non-significant halting points. The zone corner points are added in the trajectory by replacing significant halting point. When the size of zone increase, it increases the privacy but at the same time, information loss increases. This anonymization algorithm is well described in Rajesh et al. (2018) [3].

6 Ethical Aspects

Ethical aspects related to the activities of the i-DREAMS project will be managed within WP 10, T10.1 “POPD General Requirement - 1”. This task establishes effective ethical management, rooted in the project, with a thorough understanding of both the underlying science as well as the associated ethical principles.
Below are brief summaries of data management related ethical considerations (further details in 10.1)

6.1 Informed consent

It is our intention to be fully transparent with users as to the type and amount of personal data to be collected through stakeholder and community engagement as part of the project. The following actions will ensure informed consent towards this intention:

- At all points in the research engagement process, users will be informed in plain language of the nature and objective of the task, the amount and nature of the data being stored, and the nature of any resulting dissemination
- Users will be fully informed in the documentation supporting research activities about the implications of their activities and the degree of observation included within the activity
- Users will be given the opportunity to examine and correct data relating to them, and will be able to withdraw / delete any user data that is stored in the system
- Data will be stored in compliance with EU & national legislation, as well as in compliance with institutional regulation
- Users will be informed that their participation in any i-DREAMS activities is at all times voluntary and that they may at any time withdraw from any activity without penalty and without giving a reason
- Users will be informed that they can ask questions at any point during any research activity. Withdrawal from an activity can, at the user’s request, include the deletion of any data generated during that activity

6.2 Archiving and preservation

There is no maximum for data storage. However, the data that is pseudonymized will be kept no longer than necessary. We would like to store this data for a maximum of five years. However, in case this duration not sufficient, data controller partners should submit a new application to the respective ethics committee to receive approval to store the data for a longer period. Provisions will be made so that the selected technologies support archiving of data in the form of automated backups or data “dumps” in physical devices that can be located locally, as well as offsite. The goal would be to efficiently restore the data, without any data loss, or corruption.

7 Plan Maintenance

The Data Management Plan will be updated regularly by the UHASSELT and reviewed by WP-leads with the purpose of supporting the data management life cycle for all data that will be collected, processed or generated by the project.

Table 6: Planned review dates

Review Dates	Project Month	DMP Version
October 2020	Month 18	V2.0
April 2021	Month 36	V3.0

8 Conclusions

In this deliverable, we described the data management life cycle for the data to be collected, processed and/or generated by i-DREAMS H2020 project. In particular, as part of making research data findable, accessible interoperable and re-usable (FAIR), we provided information on: the handling of research data during & after the end of the project; the data that will be collected, processed and/or generated; the methodology and standards that will be applied; whether the data will be shared/publicly available; and the way to preserve and curate data (including after the end of the project). The current deliverable is an original version of the project's DMP; this will be revised two times within the project life cycle. The second update is due in M18 and final update is due in M36. Further updates of the DMP will cover the following in detail.

- Details about Joint data processing agreement for Personal data will be presented. This agreement will provide a framework through which it is ensured that the collection and processing of personal data will be carried out as per GDPR guidelines and also following the EU and national laws.
- More details will be presented about types of data being collected with their detailed formats. Furthermore, detail about the use of other technology/equipment for the collection of more detailed driving behaviour data if required to achieve the project objectives.
- A comprehensive life cycle diagram for the i-Dreams project to clearly illustrate the flow and processing of data among the consortium partners and identification of datasets that could be made publicly available.

9 References

[1] Wilkinson, M.D. et al., 2016. The FAIR Guiding Principles for scientific data management and stewardship. Scientific data, 3, p.160018.

[2] PDPC (2018), Personal Data Protection Commission Singapore,

URL:https://www.pdpc.gov.sg/-/media/Files/PDPC/PDF-Files/Other-Guides/Guide-to-Anonymisation_v1-(250118).pdf Last Accessed 10/10/2019

[3] Rajesh N, Abraham S., and Das S. S. (2018) Trajectory Anonymization Through Generalization of Significant Location Points, International Journal of Computer Sciences and Engineering, Vol 6, SI-6., https://www.ijcseonline.org/spl_pub_paper/12-IJCSE-NGIST-2018-54.pdf