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

D5.4

Development of a unified Big Data fusion framework for exploiting driving performance data in i-DREAMS

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.i-DREAMSproject.eu

Document summary

Deliverable No	D5.4
Deliverable Title	Development of a unified Big Data fusion framework for exploiting driving performance data in i-DREAMS
Work Package	5
Contractual due date	31/12/2021
Actual submission date	31/12/2021
Nature	Report
Dissemination level	Public
Lead Beneficiary	Technical University Munich
Responsible Author	Md Rakibul Alam
Contributions from	Md Rakibul Alam, Christelle Al Haddad, Kui Yang, Constantinos Antoniou (TUM); Amir Pooyan Afghari, Eleonora Papadimitriou (TU Delft); Christos Katrakazas, Eva Michelaraki (NTUA); Graham Hancox (U Loughborough); Bart Devos, Muhammad Adnan, Evelien Polders (U Hasselt)

Please refer to the document as:

Alam, M Rakibul et al. (2021). *Development of a unified Big Data fusion framework for exploiting driving performance data in i-DREAMS.* Deliverable 5.4 of the EC H2020 project i-DREAMS.

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

Revision history (including peer review & quality control)

Version	Issue date	% Complete	Changes	Contributor(s)
V1.0	25/08/2021	5%	Initial structure	See 'contributions from' above
V2.0	04/10/2021	50%	Partners' input compiled	See 'contributions from' above
V3.0	31/10/2021	100%	Ready for review	See 'contributions from' above
V4.0	28/12/2021	100%	Edited as per reviewers' recommendation	See 'contributions from' above

Disclaimer

Copyright

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

Table of contents

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

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

List of figures

List of tables

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

List of abbreviations and acronyms

Abbreviation/ Acronym	Definition
URL	Universal Resource Locator
API	Application Programming Interface
NDS	Naturalistic Driving Studies
STZ	Safety Tolerance Zone
ECG	Eco-Cardiogram
PPG	Photoplethysmography
ITS	Intelligent Transportation Systems
AID	Automatic Incidence Detection
ADAS	Advanced Driver Assistance Systems
CAS	Collision Avoidance Systems
ACC	Adaptive Cruise Control
GPS	Global Positioning System
IBI	Interbeat interval
HRV	Heart rate variability
O7SDK	OSEVEN Software Development Kit
CAN	Controller Area Network
TTC	Time-To-Collision
KSS	Karolinska Sleepiness Scale
FCW	Forward Collision Warning
LDW	Lane departure warning
PW	Pedestrian warning
KD	Knowledge discovery
RAM	Random Access Memory
RDD	Rapid Distributed Data
SQL	Structured Query Language
ML	Machine Learning
DB	Database
FOT	Field Operational Tests
OA	Open Access
IPR	Intellectual Property Rights

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

EC	European Commission
SHRP2	Strategic Highway Research Program 2
UDRIVE	eUropean naturalistic Driving and Riding for Infrastructure & Vehicle safety and Environment
ANDS	Australian Naturalistic Driving Study
SH-NDS	Shanghai Naturalistic Driving Study
M2M	Machine 2 Machine
DAS	Drivers Assistance Systems
GDPR	General Data Protection Rights
EU	European Union
DPO	Data Protection Officer

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

1 Introduction

1.1 About the project

The overall objective of the i-DREAMS project is to set up a framework for the definition, development, testing and validation of a context-aware safety envelope for driving (‘Safety Tolerance Zone’-STZ), within a smart Driver, Vehicle & Environment Assessment and Monitoring System (i-DREAMS). Taking into account driver background factors and real-time risk indicators associated with the driving performance as well as the driver state and driving task complexity indicators, a continuous real-time assessment will be made to monitor and determine if a driver is within acceptable boundaries of safe operation. Moreover, safety-oriented interventions will be developed to inform or warn the driver in real-time in an effective way as well as on an aggregate to give real timed level after driving through an app and web-based gamified coaching platform. Figure 1 summarises the conceptual framework, which will be tested in a simulator study and three stages of on-road trials in Belgium, Germany, Greece, Portugal and the United Kingdom with a total of 600 participants representing car driver, bus driver, truck drivers and rail drivers.

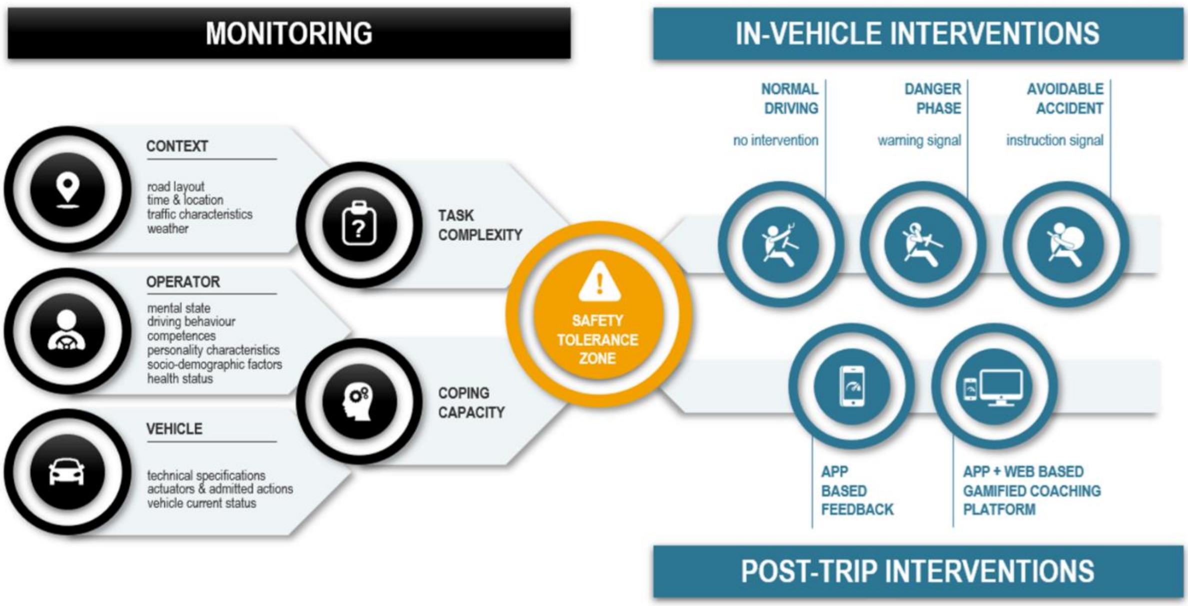

Figure 1: Conceptual framework of the i-DREAMS platform.

The key output of the project will be an integrated set of monitoring and communication tools for intervention and support, including e.g., in-vehicle assistance and feedback and notification tools as well as a gamified platform for self-determined goal-setting working with incentive schemes, training and community building tools. The technology that will be implemented includes a customised LCD capacitive touch display that communicates with the CardioID Gateway to receive the status of the STZ, giving real-time audio and visual alerts. It will also allow for driver identification upon vehicle start-up. Information coming to the CardioID Gateway is from a context-aware road monitoring system (Mobileye), and electrocardiogram (ECG), or photoplethysmography (PPG) technology (CardioWheel/ Wristband), as well as an application installed on the user’s phone to monitor hand-held phone usage.

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

1.2 Deliverable overview and report structure

Within the i-DREAMS project, there are five technical work areas: state of the art (monitoring and interventions), methodological development, technical development, trials, and analysis. This deliverable concerns the analysis part the most among all of those.

This deliverable (D5.4) takes the work outlined in Deliverable 4.3: a back-end database storing raw and processed sensor data (Alam, M Rakibul et al., 2020) and goes on developing a framework for big data collected in the project. This deliverable is meant to describe how the different aspects of big data is managed within i-DREAMS project. Additionally, this deliverable includes the discussion of data sharing aspects within and after project end.

The organization of the document is the following- section 2 provides a literature review of data fusion techniques in transport literature and then discusses the relevant fusion techniques for data fusion in case of i-DREAMS data. This helps in understanding the state-of-the-art in data fusion research and also the nature of data fusion needed for i-DREAMS data to accommodate in the big data framework. Then, section 3 describes the unified big data framework to exploit data collected in the i-DREAMS project. Afterwards, section 4 describes the prospect of the data collected after the end of the project. Here we will explore the aspects of making data open access once the project ends and feasibility analysis will be presented from different aspects. Section 5 concludes the deliverable.

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

2 Data fusion methods and techniques

2.1 Data fusion in transport literature

With the advancement of modern communication, intelligent hardware along with the widely available sensors with cheap prices, collecting and processing data from a number of sources become easier. Data fusion has gained a lot of attention in transport literature. Data fusion can be defined as the process of collecting sensed information from several sources (AlZu'bi & Jararweh, 2020). The collected information is combined in order to reach a better inference and more accurate results. Therefore, data fusion has been considered as inevitable means for Intelligent Transportation Systems (ITS).

Several methodologies have been proposed in the literature for the purpose of multi-sensor data fusion and aggregation under heterogeneous data configurations. Due to the different types of sensors that are used as well as the heterogeneous nature of information that needs to be combined, different data fusion techniques are being developed in order to suit the data and applications (El Faouzi et al, 2011). These techniques were drawn from a wide range of areas, including artificial intelligence, statistical estimation, pattern recognition, etc. Traffic engineering field has naturally benefited from this abundant literature. The following approaches have been suggested in the literature:

- **Statistical approaches**: Weighted combination, multivariate statistical analysis and its most up-to-date form data mining engine (Singh, 2021). Among statistical techniques, the arithmetic mean approach is the simplest which is used for information combination. This approach is not suitable when the available information is not exchangeable or when estimators/classifiers have dissimilar performances (Hashem, 1997; El Faouzi, 1997).
- **Probabilistic approaches**: Bayesian Network and state-space models (Xiao et al., 2019; Cappelle et al., 2008), maximum likelihood methods and Kalman filter based data fusion (Okutani, 1987; Gan & Harris, 2001; Lim et al., 2021), possibility theory (Dubois & Prade, 2021), evidential reasoning and more specifically evidence theory (Du & Zhong, 2021; He et al., 2021) are widely used for the multi-sensor data fusion. The later technique could be viewed as a generalization of Bayesian approach (Yu et al., 2021; Yang & Xu, 2014).
- **Artificial intelligence**: Artificial cognition including artificial intelligence, genetic algorithms and Neural Networks. In many applications, the later approach serves both as a tool to derive classifiers or estimators and as a fusion framework of classifiers/estimators (Zhang et al., 2013).

Although the application of data fusion techniques to complex systems modelling is not new, there is a growing interest in their use in transportation systems. Road traffic engineering could be considered as a field where benefits expected from the application of data fusion techniques are potentially large. However, the benefits come with challenges in assessing feasibility, effectiveness and usefulness of such approaches (Keever et al., 2003; Simon, 2000). With regards to transport literature, the interest in data fusion is quite new and it coincides with ITS advent. Numerous scientific papers exist regarding the application of data fusion in transport engineering. Different techniques have been developed mainly for traffic surveillance and control (El Faouzi, 2004; Kolosz et al., 2013) The most significant applications of data fusion techniques in road traffic engineering area are described below-

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

2.1.1 Automatic incident detection

Incident detection methods for automatic recognition of incidents, crashes and other road events requiring emergency responses have existed for more than three decades. Most of the developed and implemented algorithms rely on loop detectors data. However, these algorithms work with mixed success. Recently, there has been renewed interest in incident detection algorithms partly because of the availability of new sensors and data sources. Thus, automatic incident detection (AID) belongs to the class of problems that can be solved by data fusion techniques. Applications of several data fusion techniques to traffic management to support incident detection have been reported in the literature, and the data fusion algorithms used includes Dempster-Shafer inference (Simon et al., 2008) or Bayesian inference (Jaramillo et al., 2017).

Most of these applications have explored the use of probe vehicles data with the conventional traffic data for incident detection purposes. As an example of such work (Schofer et al., 1994; Ivan, 1995) developed an AID system using surveillance data from two different sources: fixed detectors (e.g. inductive loop detectors) and probe vehicle specially equipped to report link travel time. The neural network approach was considered in the same research and two strategies were tested. The first one combined observed traffic directly to determine whether or not an incident is occurring. In the second, separate incident detection algorithms individually pre-process data from each source, reporting scores which are combined by neural network. Different neural network representations were studied in Ivan (1997) and results indicated that probe and detector based incident detection on arterial networks shows considerable promise for improved performance and reliability. Dempster-Shafer inference or evidential reasoning was also used to perform an operational AID system (Byun et al., 1999; Feng et al., 2009; Basir et al., 2005).

2.1.2 Advanced driver assistance

Passenger safety is considered as one of the important aspects of ITS. Driver assistance techniques are being developed from this point of view. In the past few decades, tremendous progress has been made with regards to vehicle safety and driver assistance. Early safety approaches emphasize precaution and focus on passive devices (i.e. seat belts, air bags and lighting). In spite of crash-related injury severity rate reduction, drivers demand greater improvements in vehicular transportation safety. Research trends show the use of active safety devices which complement the traditional passive ones. ADAS (Advanced Driver Assistance Systems) and collision avoidance systems (CAS) are an illustration of such trends. The main objective assigned to these systems is to provide a more reliable description of the traffic scene surrounding the vehicle to vulnerable road users, in pre-crash situations and to systems like adaptive cruise control (ACC) and CAS.

Operational systems are based on several sensor systems which are complementary and redundant; thus, the data fusion process provides a fused description of the traffic scene. This fusion incorporates the data of the available sensors into a single description. The crucial aspect here is to associate sensor data with environment description, which requires synchronization of the sensor data and associated object state. Whenever there are multiple sensors being used to sense multiple objects, there is a need to associate the measurements with the individual objects (Hall & McMullen, 2004). Once the sensor measurements are associated with appropriate objects, the next step is to remove the sensor bias. This procedure is called sensor registration. Finally, objects are tracked using fused sensor measurements.

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

Kalman filter, its variants and more recently particle filtering become an essential tool to perform this step (Bar-Shalom et al., 2004). Several papers report some results within this topic. For example, Murphy (1998) investigated sensor fusion's role in-vehicle guidance and navigation and proposed general methods for fusing data with sensor-fusion activities within a robot architecture. Another work conducted by Pei and Liou (1998) proposed three-dimensional vehicle motion estimation by fusion of multi-source information. Image point and line features were considered for fusion. Lastly, Langheim et al. (2002) examined data fusion systems for Automatic Cruise Control (ACC) with stop and go phenomenon, and Darms et al. (2009) reported a data fusion framework for obstacle detection and tracking.

2.1.3 Network control

Data fusion techniques were also applied in the road network control issues. In Mueck (2002) and Wang & Papageorgiou (2005), the problem of constructing an adaptive online traffic control in urban or freeway road networks was investigated. In the first one, a model that determines queue length on the basis of vehicle counts from detectors located close to the stop line and on the basis of signal timings was derived. In the second one, the research performed the freeway traffic control using extended Kalman filter. Along this direction, Friedrich et al. (2003) introduced a new approach based on queuing theory models for real time queue length determination. In this later method, Mueck's model serves as a quasi-measurement with Kalman filtering technique.

2.1.4 Crash analysis and prevention

Although there has been a steady reduction in the number of crashes, these continue to sustain heavy losses in both human and economic terms. Reduction in the number of crashes could be due to multiple efforts: road infrastructure improvement, regulations on speeding, alcohol or drug abuse and improvement in-vehicle safety. Many studies were carried out in order to explain the circumstances and the characteristics of traffic crashes. One way to conduct such explorative studies is to utilize retrospective data available such as traffic crash records. Along this direction, Sohn and Shin (2001) employed both neural network and decision tree algorithms to find the classification model for road traffic crash severity (i.e. bodily injury or property damage) as a function of potentially related categorical factors. They noted that the classification accuracy of the individual algorithm was relatively low. Recognizing that, Sohn and Lee (2003) used data fusion and ensemble algorithms in order to increase the accuracy.

Data fusion techniques were used to perform fusion of the results of different classifiers using evidence theory. Data ensemble combined various results obtained from a single classifier fitted repeatedly based on several bootstrap samples (Liang et al., 2011; Wang, 2008; Breima, 2010; Naimi et al., 2018). More precisely, they tried three different approaches: classifier fusion based on the Dempster-Shafer algorithm, the Bayesian procedure and logistic model; data ensemble fusion based on arcing and bagging; and clustering based on the k-means algorithm. Empirical results revealed that a clustering based classification algorithm appeared to have a better performance for road traffic crash classification.

2.1.5 Traffic forecasting and traffic monitoring

Traffic flow forecasting has received increasing attention in the past years and different techniques have been developed mainly for traffic surveillance and control. Many prediction schemes of traffic flow were obtained by means of classic autoregressive models, especially time series techniques. Some authors have tackled this problem in the context of Bayesian

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

framework (Sahu & Mardia, 2005). Some others used Kalman filtering technique (Gastaldi et al., 2004) or neural networks and system identification (Buitrago & Asfour, 2017). However, none of these proposals allowed to achieve highly accurate predictions except in some special situations (for some network configuration and/or with high detector coverage). In the context of traffic operations where highly accurate forecasts are needed, one can obtain different forecasts of the same quantity (the underlined assumption here is that different predictors are measures of the same quantity and/or various aspects of the same thing) by two or more different methods. The set of available methods may consist of alternative models, different forecasters, or a mixture of models and forecasters.

Commonly, the most appropriate approach utilized is to find the single 'best' predictor in some sense (most accurate values, most appropriate models of the underlying process, most cost-effective, etc.) among the available forecasting methods. Another approach consists of combining these individual forecasts. The idea of combining estimators instead of selecting the single 'best' model has a long history and has generated intensive theoretical works since the seminal article of Granger (1989). In this work, it was indicated that the linear combination of several predictors from a single data set can outperform the individual predictors. Methodological and practical issues related to combining forecasts produced by different methods has been investigated extensively in various contexts with notable successes.

With regards to traffic forecasting under heterogeneous data sources configuration, El Faouzi (1999) provided a methodological framework to combine various forecasts of the same quantity. He derived two predictors using nonparametric traffic flow using a kernel estimator and predicting scheme based on the propagation of a lagged upstream traffic flow. The proposed combination strategies exhibit very encouraging results. Data integration and data fusion were applied for other purposes. In Cremer & Schrieber (1996), the integration problem of in-vehicle information and data provided by loop detectors was investigated. The core of the integration step was the extended Kalman filtering. Another research conducted by Sau et al. (2007) examined the traffic monitoring problem within the multi-source data. The particle filter was the estimation technique used in this context. Lastly, Choi (1998) studied the problem of missing data estimation and proposed a framework for missing data inference based on evidential reasoning.

2.2 Data fusion methods and techniques for the i-DREAMS data

The collected data in i-DREAMS can be divided into two broad categories, raw and processed[1] data (deliverable 4.3 of working package 4). These data are collected from various sources including sensors (e.g. GPS, Mobileye, Gateway, and CardioWheel), driving simulator, survey questionnaires, and video cameras. While the nature and frequency of data collection vary across these sources, they all should come together in the back-office for later use in the project. This heterogeneous nature of i-DREAMS data requires data fusion prior to their transfer to the back-office.

Combining the insights from a review of existing data fusion techniques (presented in the previous section) with expert judgment, this section presents data fusion techniques that have been used by different partners in i-DREAMS to bring together all data from various sources. In addition, and to better illustrate the data fusion practice, several examples from data before and after fusion are presented in each subsection.

[1] In this context, "processed data" is distinguished from "analysed data". What we mean by processed data is mainly data pre-processing which aims to prepare the data for the next stage i.e. data analysis

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

2.2.1 Raw trip data

The raw trip data collected in i-DREAMS include GPS coordinates, Mobileye events, acceleration, inter-beat Intervals (IBIs) and travel time. These data are collected via CardioID gateway, Mobileye, and PulseOn wristband or CardioWheel (depending on the transport mode). The frequency of data collection for GPS coordinates and acceleration are 1 Hz and 104 Hz, respectively. The frequency of data collection for headway measurements and traffic sign recognition is 20 Hz. Other Mobileye events and IBIs are collected on an event basis (only available when there is an event) and the travel time is collected continuously.

The trip data are fused using a feature-based data fusion technique, namely geolocation through synchronization and support vector machines. The system provided by CardioID integrates several data streams, generated by the different sensors that make up the inputs of the i-Dreams system. After data validation, confirming that no unordered or repeated points exist in the data set, events generated by the analysis of accelerometer data and by the Mobileye are geolocated through synchronization with the GPS information. Moreover, time on task is fused with heart rate variability (HRV) as features used by CardioID's sleepiness machine learning model to produce the sleepiness state output. An example of such data fusion is shown in Table 1 and 2 and Figure 2.

Table 1: Sample of accelerometer and geolocation data before fusion in i-DREAMS

Accelerometer events					
Timestamp	**Event**	**Max acceleration**	**Duration**		**Level**
2021-08-02T06:23:16.217951Z	ha	0.273136103	1.019230769		Medium
2021-08-02T06:23:16.632684Z	ha	0.273136103	3.201923077		Low
2021-08-02T06:23:17.690376Z	hc	0.368389249	1.221153846		Medium
2021-08-02T06:23:18.184220Z	hc	0.368389249	2.067307692		Low
GPS coordinates					
Time stamp	**Latitude**	**Longitude**	**Altitude**	**HDG**	**Speed**
2021-08-02T06:23:15.085579Z	51.24790152	5.52683285	32.4	73.7	0.0
2021-08-02T06:23:16.091028Z	51.24789797	5.526826383	32.7	73.7	0.0
2021-08-02T06:23:17.096578Z	51.2479156	5.52681035	33.1	331	7.9632
2021-08-02T06:23:18.100690Z	51.24792918	5.526797617	34.2	324.9	7.2216
2021-08-02T06:23:19.104929Z	51.24796712	5.526718383	36	311.4	20.0016

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

Table 2: Fusion of accelerometer and geolocation data in i-DREAMS

Timestamp	Event	Maximum acceleration	Duration	Level	Latitude	Longitude
2021-08-02T06:23:15.632684Z	ha	0.273136103	3.201923077	Low	51.24789984	5.526829471
2021-08-02T06:23:16.217951Z	ha	0.273136103	1.019230769	Medium	51.24789513	5.526821016
2021-08-02T06:23:17.184220Z	hc	0.368389249	2.067307692	Low	51.24791546	5.526810355
2021-08-02T06:23:17.690376Z	hc	0.368389249	1.221153846	Medium	51.2479193	5.52680142

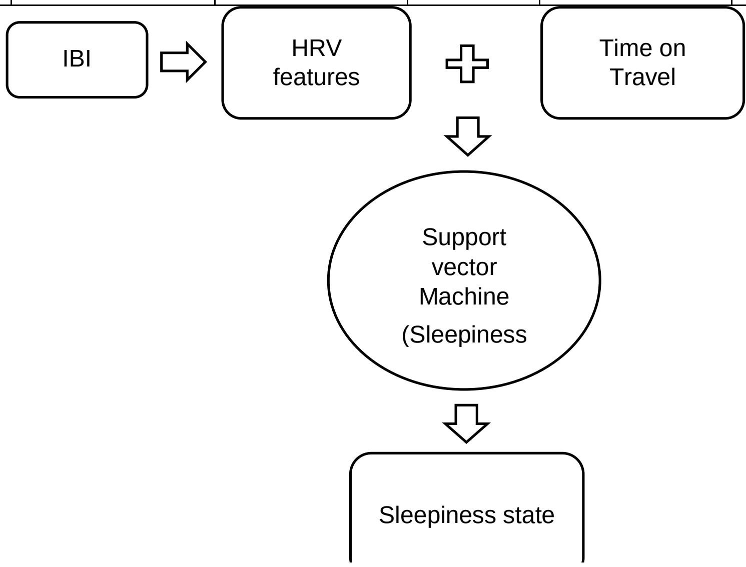

Figure 2: Example of data fusion in i-DREAMS corresponding with heart-rate data using feature fusion

2.2.2 Smartphone data

Another part of raw trip data is collected using drivers' smartphones. These data include date, time, GPS coordinates (longitude, latitude, altitude), speed, accelerometer data, gyroscope data (yaw, pitch, roll), activity (e.g., walking, driving), screen state, and post-trip data such as speed limits (from OpenStreetMaps). All of these data are collected by OSeven software from smartphone sensors with the frequency of 10 Hz. However, it is noted that in several cases the frequency of collected data depends on the specifications of the smartphone manufacturer.

To fuse the aforementioned data, a mix of methods, including mainly feature-based data fusion is used in the O'Seven algorithms within i-DREAMS. For the real time detection of mobile phone use, the data collected from the different sensors are fused in the O7SDK and used in a classification algorithm that reports the mobile phone use to the i-DREAMS system. An example of data subset that is used for the real time mobile phone use detection is outlined in Table 4 below.

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

Table 3: Sample of smartphone data before fusion in i-DREAMS

Accelerometer data			
Time stamp	**Accelerometer X**	**Accelerometer Y**	**Accelerometer Z**
18:40:36.000+1000	0.02179	-0.93841	-0.3508
18:40:37.000+1000	0.02294	-0.94057	-0.34811
18:40:38.000+1000	0.02278	-0.93625	-0.35625
18:40:39.000+1000	0.02475	-0.93976	-0.35121
18:40:40.000+1000	-0.04634	-0.98864	-0.40398
18:40:41.000+1000	0.5104	-0.56752	-0.28987
18:40:42.000+1000	-0.24299	-0.90577	0.12562
18:40:43.000+1000	0.03086	-0.52289	-0.5458
18:40:44.000+1000	0.30064	-0.37375	-0.89348
Orientation data			
Time	**Yaw**	**Pitch**	**Roll**
18:40:36.000+1000	3.47860	-1.21244	0.06389
18:40:37.000+1000	3.43327	-1.21148	0.06741
18:40:38.000+1000	5.68093	-0.212	0.51548
18:40:39.000+1000	3.33776	-1.21242	0.06214
18:40:40.000+1000	1.73611	-0.743	-0.71029
18:40:41.000+1000	5.15209	-0.7443	1.11183
18:40:42.000+1000	0.43563	-0.25123	-0.53457
18:40:43.000+1000	2.05952	-0.97168	0.27707
18:40:44.000+1000	2.95537	-0.39079	0.27006
Screen state data			
Time		**Screen State**	
18:40:36.000+1000		0	
18:40:38.000+1000		1	

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

Table 4: Sample of smartphone data after fusion in i-DREAMS

Time	accelerometer_X	accelerometer_Y	accelerometer_Z	Yaw	Pitch	Roll	Screen State	Mobile usage
18:40:36.000+1000	0.02179	-0.93841	-0.3508	3.47860	-1.21244	0.06389	0	0
18:40:37.000+1000	0.02294	-0.94057	-0.34811	3.43327	-1.21148	0.06741		0
18:40:38.000+1000	0.02278	-0.93625	-0.35625	5.68093	-0.212	0.51548	1	0
18:40:39.000+1000	0.02475	-0.93976	-0.35121	3.33776	-1.21242	0.06214		0
18:40:40.000+1000	-0.04634	-0.98864	-0.40398	1.73611	-0.743	-0.71029		1
18:40:41.000+1000	0.5104	-0.56752	-0.28987	5.15209	-0.7443	1.11183		1
18:40:42.000+1000	-0.24299	-0.90577	0.12562	0.43563	-0.25123	-0.53457		1
18:40:43.000+1000	0.03086	-0.52289	-0.5458	2.05952	-0.97168	0.27707		1
18:40:44.000+1000	0.30064	-0.37375	-0.89348	2.95537	-0.39079	0.27006		1

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

2.2.3 Simulator data

A third part of raw data collected in i-DREAMS is associated with simulator experiments which is collected by driving simulators. Although there are multiple simulator types in i-DREAMS, there are common elements among them in terms of data collection. The main parts of simulator data are collected from the simulator software itself and generally include parameters that are captured by virtual sensors in the simulated environment. In addition, data are collected by the i-DREAMS gateway, which are already fused data from all external sensors (e.g. Mobileye and Cardiowheel). The driving simulator in Germany collects data regarding eye tracking too. Finally, driver demographics and other individual characteristics are collected using survey questionnaires during the simulator experiments. The frequency of data collection in the driving simulators varies between 20 Hz and 30 Hz depending on the simulator, whereas the data frequencies of eye tracker data is 50 Hz. The survey questionnaires are collected per person.

The different parts of data within driving simulator are fused mostly using feature-based data fusion technique. Data fusion is performed in real-time during the simulation experiments. At each simulation step (approximately 30 Hz), parameters from the simulated environment are written to a log file, while any available status messages from the gateway are read from the input buffer, and written to the same log file. The gateway only sends a message when a parameter has been updated, therefore the value of a parameter sent by the gateway can be considered constant until the next update. This is represented by blanks entries in the aggregated log file. An exception to this exist when the parameter is forwarded to the i-DREAMS intervention algorithm, which is done at the end of each simulation frame. In this case the particular parameter has no blank entries (Traffic Sign Recognition, Headway, etc.). An example of driving simulator data-fusion is shown in Table 5 below.

Table 5: Sample of simulator data before fusion in i-DREAMS

<table>
<tr><th colspan="7">Simulator data</th></tr>
<tr><th>Time (s)</th><th>Speed (km/h)</th><th>Lateral Position (m)</th><th>Longitudinal Acceleration (m/s²)</th><th>Distance Travelled (m)</th><th>Headway calculated by simulator (s)</th><th>Throttle input</th></tr>
<tr><td>1.931244</td><td>77</td><td>6.250006297</td><td>1.689551</td><td>0.528308087</td><td>3.261598</td><td>1</td></tr>
<tr><td>1.964577</td><td>77</td><td>6.250006879</td><td>1.685844</td><td>0.571609187</td><td>3.261689</td><td>1</td></tr>
<tr><td>1.997911</td><td>77</td><td>6.250008041</td><td>1.682141</td><td>0.616775216</td><td>3.261763</td><td>1</td></tr>
<tr><td>2.031244</td><td>77</td><td>6.250009785</td><td>1.678442</td><td>0.663802142</td><td>3.261832</td><td>1</td></tr>
<tr><td>2.064577</td><td>78</td><td>6.250010948</td><td>1.674747</td><td>0.712685714</td><td>3.261879</td><td>1</td></tr>
<tr><td>2.097911</td><td>78</td><td>6.250012692</td><td>1.671056</td><td>0.763421934</td><td>3.261922</td><td>1</td></tr>
<tr><th colspan="7">Gateway data</th></tr>
<tr><td>IBI</td><th colspan="2">Headway Calculated by Mobileye</th><th>Heart Rate</th><th colspan="2">Mobileye CAN message</th><th>Mobileye Traffic sign recognition byte</th></tr>
<tr><td>839</td><td colspan="2"></td><td></td><td colspan="2"></td><td></td></tr>
<tr><td></td><td colspan="2">2</td><td></td><td colspan="2">0-0-41-1-0-0-0-1</td><td></td></tr>
<tr><td>847</td><td colspan="2"></td><td></td><td colspan="2"></td><td></td></tr>
</table>

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

		71		
			8-0-254-0-254-0-254-0	8
			8-0-254-0-254-0-254-0	8
	2.5		0-0-51-1-0-0-0-1	

Table 6: Sample of simulator data after fusion in i-DREAMS

Time (s)	Speed (km/h)	Distance Travelled (m)	Throttle input	IBI	Headway Calculated by Mobileye (s)	Heartrate	Mobileye Traffic sign recognition byte
1.797911	77	0.373835212	0.739318	839	0		0
1.831244	77	0.409635461	0.827333		2		0
1.864577	77	0.4473171	0.942857	847	2		0
1.897911	77	0.486876023	1		2	71	0
1.931244	77	0.528308087	1		2		0
1.964577	77	0.571609187	1		2		0
1.997911	77	0.616775216	1		2		8
2.031244	77	0.663802142	1		2		8
2.064577	78	0.712685714	1		2.5		8
2.097911	78	0.763421934	1		2.5		8

2.2.4 Processed data

Data fusion is also implemented on processed data in i-DREAMS. The processed data have already been fused once before processing (i.e. all of the previously mentioned data fusion techniques on raw data). However, such fused data may still need aggregation for specific data analysis needs in i-DREAMS. For example, data may need to be aggregated in finer time intervals (e.g. 30 seconds) for real time analysis and in courser time intervals (e.g. 2 minutes) for post-trip analysis. As such, data aggregation is the final data fusion technique that is applied on the data.

The aggregation method depends on the type of data and their specific purpose in i-DREAMS. For example, the nature of TTC (being continuous; only applicable for simulation experiment data) requires minimum and mean as the aggregation method whereas the nature of KSS (being discrete) requires median as the aggregation method. Nevertheless, what is important to note is that for the identification of the STZ level in real-time, data collected from different sensors are aggregated at 30-second intervals (i.e. mean, min, max, average) and will be used to feed a Dynamic Bayesian Network for multi-level dynamic classification. In addition, for post-trip explanatory analysis of data, they will be aggregated in 2-minute time intervals or higher (i.e. trip level) and the aggregated data will be used in post-trip data analysis methods (i.e. discrete choice models and structural equation models). An example of data fusion for processed data are shown in Table 7 and Figure 3.

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

Table 7: Sample of processed data for post-trip analysis data before fusion in i-DREAMS

Simulator data						
Time	**Headway**	**TTC**	**FCW**	**LDW**	**PW**	**Speed**
0	425666.5	-92.3837				0
0.056967	199483.3	-92.4064				0
0.256967	69791.85	-92.486	False	False	False	0
0.290301	79134.25	-92.4716				0
0.323634	83105.06	-92.4664				48
0.356967	87494.44	-92.4612				0
0.390301	92362.29	-92.4561				0
0.423634	97791.36	-92.451				0
0.456967	103884.6	-92.4458				0
0.490301	81668.91	-92.4682				0
0.523634	94655.27	-92.4539				0
0.556967	100348.3	-92.4487	False	False	False	0
0.590301	79471.91	-92.4711				0
0.623634	91722.68	-92.4567	False	False	False	0
0.656967	97065.26	-92.4516				0

Survey data				
Driver ID	**Gender**	**Nationality**	**Year of Birth**	**Year of driving licence**
1.00	Male	Belgium	1975	1993
2.00	Male	Belgium	1978	2000
3.00	Female	Belgium	1986	2005
4.00	Male	Belgium	1983	2007
5.00	Male	Belgium	1976	2008
7.00	Male	Belgium	1986	2010
8.00	Male	Belgium	1972	2016
9.00	Male	Belgium	1992	2020
10.00	Male	Belgium	1969	1989
11.00	Male	Belgium	1973	1993
12.00	Male	Belgium	1981	2005

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

Table 8: Sample of processed data after fusion for post-trip analysis data in i-DREAMS

ID	Time (s)	Headway (s)	TTC (s)	FCW	LDW	PW	Speed (km/h)	Gender	Age (years)
1	60	971.2058	25.68582	0	0	0	71.21899	Male	46
1	120	37.28229	16.35462	0	0	0	85.1421	Male	46
1	180	53.39157	16.36019	0	0.008452	0	78.76479	Male	46
1	240	53.38296	17.91638	0	0.021886	0	78.38721	Male	46
1	300	76.3441	9.088667	0	0	0	57.63919	Male	46
1	360	66.89072	9.430708	0	0	0	56.10003	Male	46
1	420	87.17356	13.22648	0	0	0	57.68495	Male	46
1	480	67.40645	12.84884	0	0	0	52.74779	Male	46
2	60	3212.413	8.071118	0	0	0	49.39593	Male	43
2	120	69.18868	10.54109	0	0	0	51.57236	Male	43
2	180	90.88912	9.000893	0	0	0	55.35638	Male	43
2	240	99.21179	45.90915	0	0.008869	0	72.28869	Male	43
2	300	16.41718	19.35199	0	0.016439	0	77.56729	Male	43

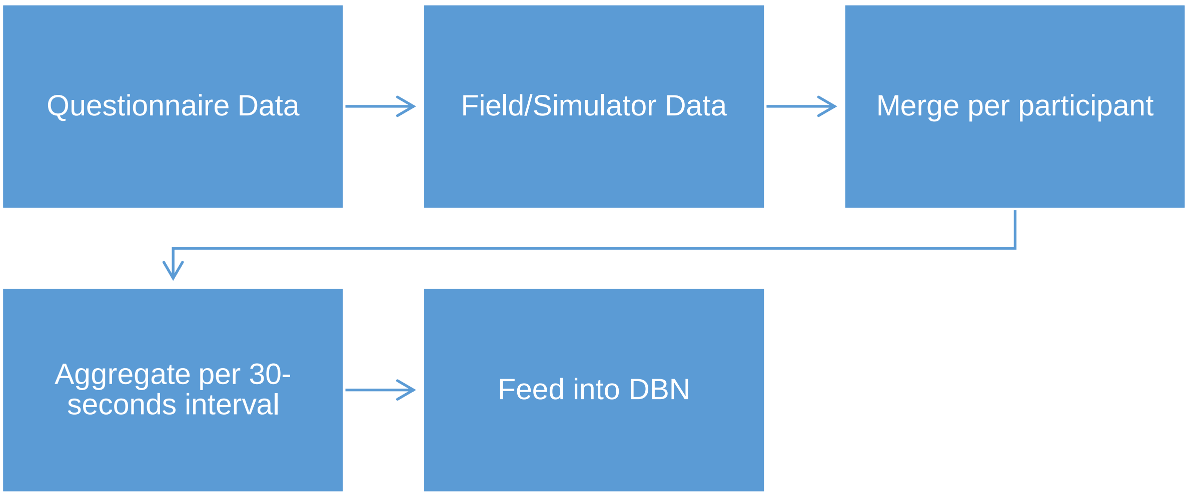

Figure 3:Example of data fusion in i-DREAMS corresponding with processed data for real-time analysis

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

3 Unified big data fusion and analytics framework to exploit the i-DREAMS data

3.1 Context

The i-DREAMS project has a back-office which enables data management and facilitate future data analysis. It empowers consortium partners to store and retrieve data collected through road trials, simulations, surveys etc. In addition, back-office provides a way to perform specific data analysis tasks (defined in future work-packages) on the raw data. The results of such analysis is stored in the back-end database. The back-office system allows all consortium partners to access those results (following a certain access strategy). The architecture of back-office consists of three components- communication component, data storage component, knowledge discovery component.

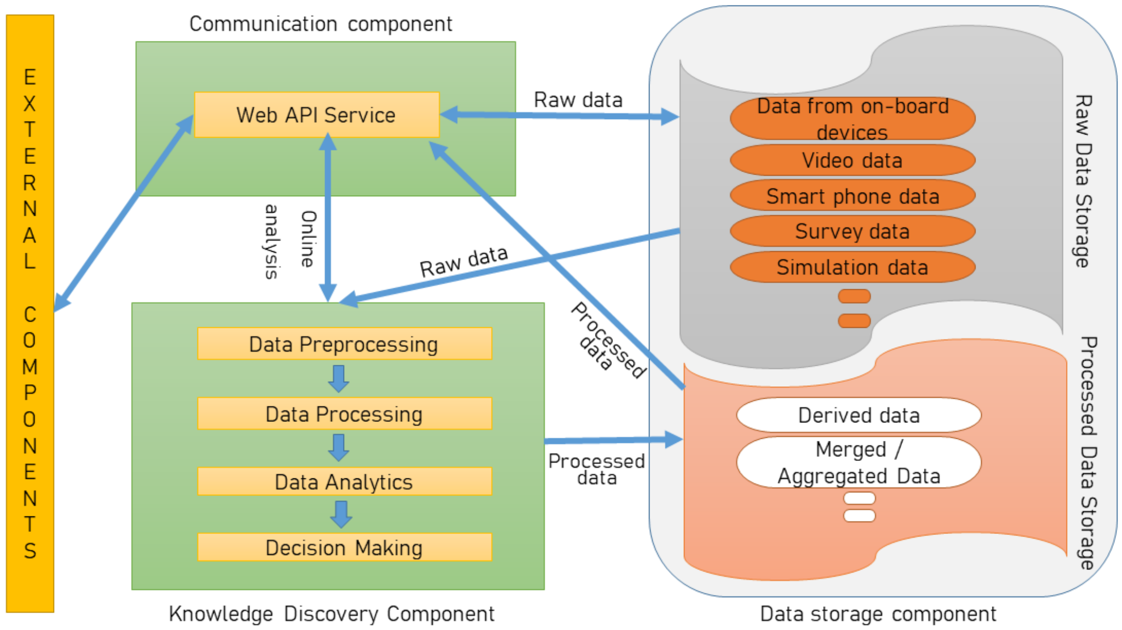

Figure 4: i-DREAMS Back-office components (image taken from i-DREAMS deliverable D4.3)

Among these components, the knowledge discovery component does the data processing and analysis tasks including data fusion. This component needs to have an analytics framework that allows complex data processing and analysis of large datasets which is being collected by the project. The volume and heterogeneity of the data require the framework to have efficient ways to deal with the data.

In the following sections we will discuss the framework for data fusion in analytics in the knowledge discovery and then we will dive into a detailed description of the framework implementation in the back-office component.

3.2 Framework overview

The core objective of a big data analytics framework is to provide a structure for data analytics which aim to benefit from the potential of big data. The content of this section discusses the development of the big data framework which incorporates data fusion and analytics. This framework enables real time data processing (including fusion) and analytics for the i-

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

DREAMS data collected from road trials, surveys and simulations. The data is already stored in key-value storage database (MongoDB). This framework will allow i-DREAMS partners to perform large data analysis and to discover knowledge from the data in real time.

The scope of this framework is limited to loading different data sets from the database, fusing data to accelerate further analysis, performing the analytic tasks and then storing the results back to the database. These result sets then can be visualized in the analytics dashboard which is also part of the framework. For data fusion, we have taken a flexible approach in the framework. Data can be fused together in different ways and in different levels. The i-DREAMS project aims to do numerous research on the large dataset it is collecting. It is anticipated that different research questions would be in need of specialized data fusion techniques designed particularly for the respective analysis to be performed. Therefore, implementing a specific type of data fusion won't be practical in the framework and its implementation. Rather, it is necessary to have the flexibility in the framework that can support different data fusion algorithm implementations in terms of efficient handling of the data.

3.3 Framework components

The big data framework is depicted in the figure 4. It has four major components. The data storage, the analytics service, the analytics API and the dashboard.

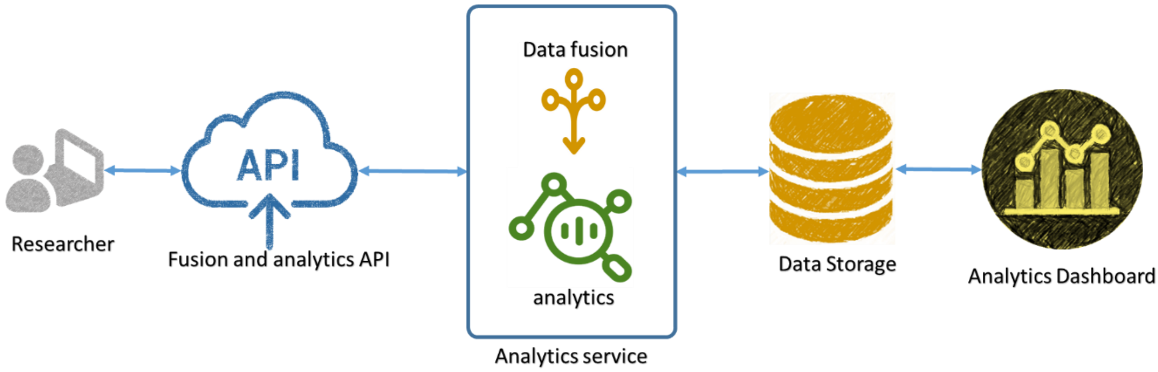

Figure 5: Unified big data fusion analytics framework for i-DREAMS

The data storage component holds the i-DREAMS data for research. The fusion and analytics API serves as communication interface between the researchers of i-DREAMS data and the analytics service component. Researchers use the APIs to call specific data fusion and/or analytics tasks that they want to perform on the data. Requests are sent using specific API endpoints towards the analytics service which in turn send responses back to the researcher. The analytics service is comprised of various data fusion and analytics scripts. Based on the researchers' request this component first loads different datasets from the data storage. Then the appropriate data fusion task is performed on those different datasets to form a fused dataset. This fused dataset is then used in the analytics pipeline. Finally, when the analytics pipeline is finished working on the data then the results are stored back to the data storage for visualization and future usage. The dashboard component is used for visualization of the data stored in the data storage using a number of visualization techniques.

3.4 Framework implementation

We have implemented the framework in the existing back-office architecture. In this section, we will describe the parts of the back-office architecture which correspond to the framework to see how the framework is realised with various available tools and technologies suitable for large data handling.

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

The knowledge discovery component (KD component) implements two of the four components of the framework- the dashboard and analytics service. Fusion and analytics API is part of the

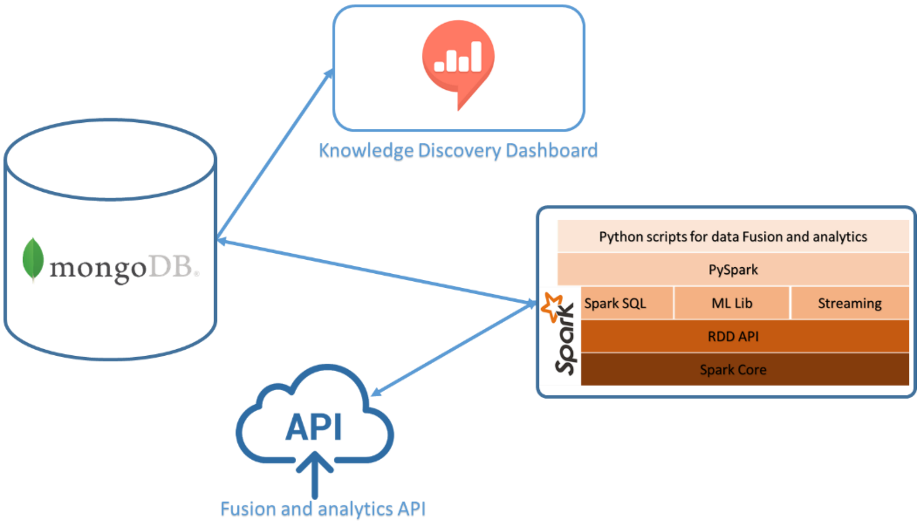

Figure 6: Big data analytics framework implementation in i-DREAMS back-office

communication component which encapsulates other APIs of the back-office architecture while the data storage component refers to the back-office data storage component.

3.4.1 Analytics service

Figure 6 show the architecture of the analytics service. It is based on Apache Spark computing framework deployed in the cloud Apache spark is an open source clustering computing framework. This is currently one of the most popular tools for big data handling due to its scalability, flexibility and fault tolerance. In contrast with Hadoop (another popular tool for big data), Spark can work within RAM and does not need to write to the file system if not necessary.

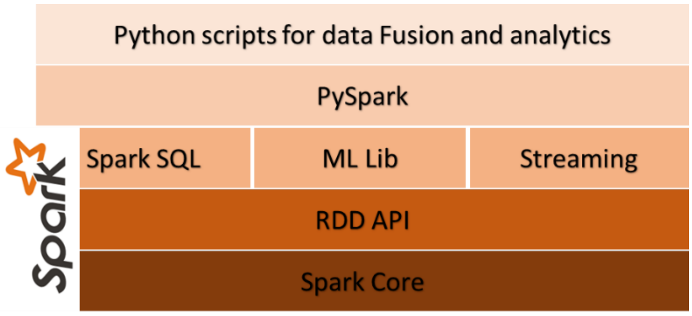

Figure 7: Analytics service for knowledge discovery component in iDREAMS back-office

Spark uses the data structure know as Resilient Distributed Datasets (RDD) which makes it even faster compare to Hadoop while it comes to real-time analysis. Spark is also comprised of several data science and machine learning libraries along with Spark SQL. Spark SQL provides the way to work RDDs for data handling. Spark is written in Java but exposes its API for different programming languages.

PySpark is an interface for Apache Spark in Python. It not only allows to write Spark applications using Python APIs, but also provides the PySpark shell for interactively analyzing large datasets in a distributed environment. PySpark supports most of Spark's features such

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

as Spark SQL, DataFrame, Streaming, MLlib (Machine Learning) and Spark Core. Since the back-office system is implemented in Python, we have chosen Pyspark as the middleware for Spark. Algorithms for data fusion and analytics would be coded in Python according to the needs of specific research tasks (to be defined in next Work Packages of i-DREAMS).

3.4.2 Knowledge discovery dashboard

For visualization of the data in i-DREAMS, we have chosen a dashboard called Redash (Redash website). Redash enables anyone, without sophisticated knowledge of technical aspects, to harness the power of large datasets. It exposes easy ways to explore, query, visualize and share from a variety of data sources. It is a browser based visualization tool. It has a query editor for querying data that supports both SQL and NoSQL queries. Redash has drag and drop visualization tools built in the application along with the capability of saving the visualizations for later and to share with other people. These saved visualizations can be updated when the original data source has more data that the specific query used.

The figures below show a glimpse of different visualizations that already exist in the iDREAM back-office dashboard.

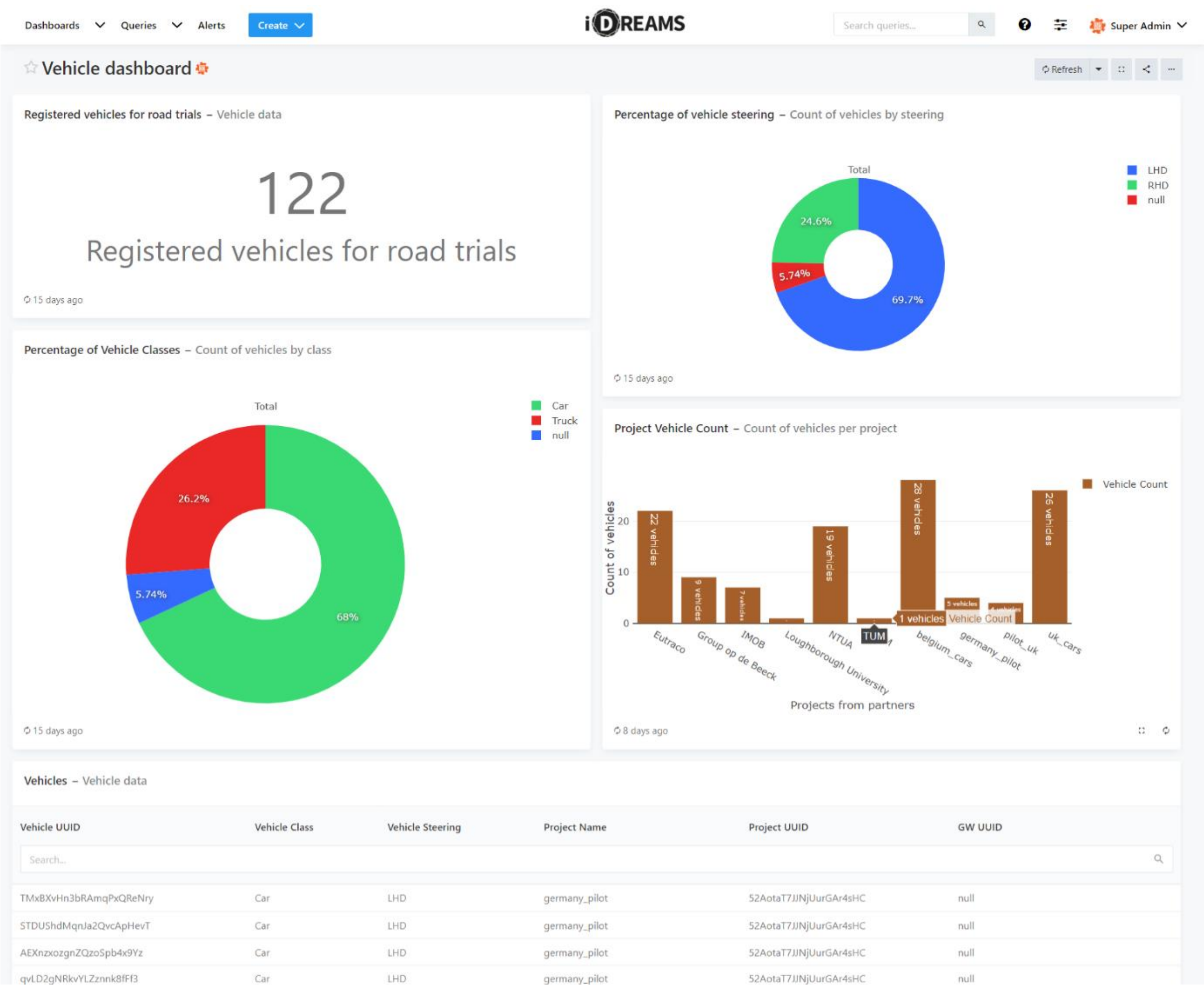

Figure 8: i-DREAMS Dashboard Visualization (a)

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

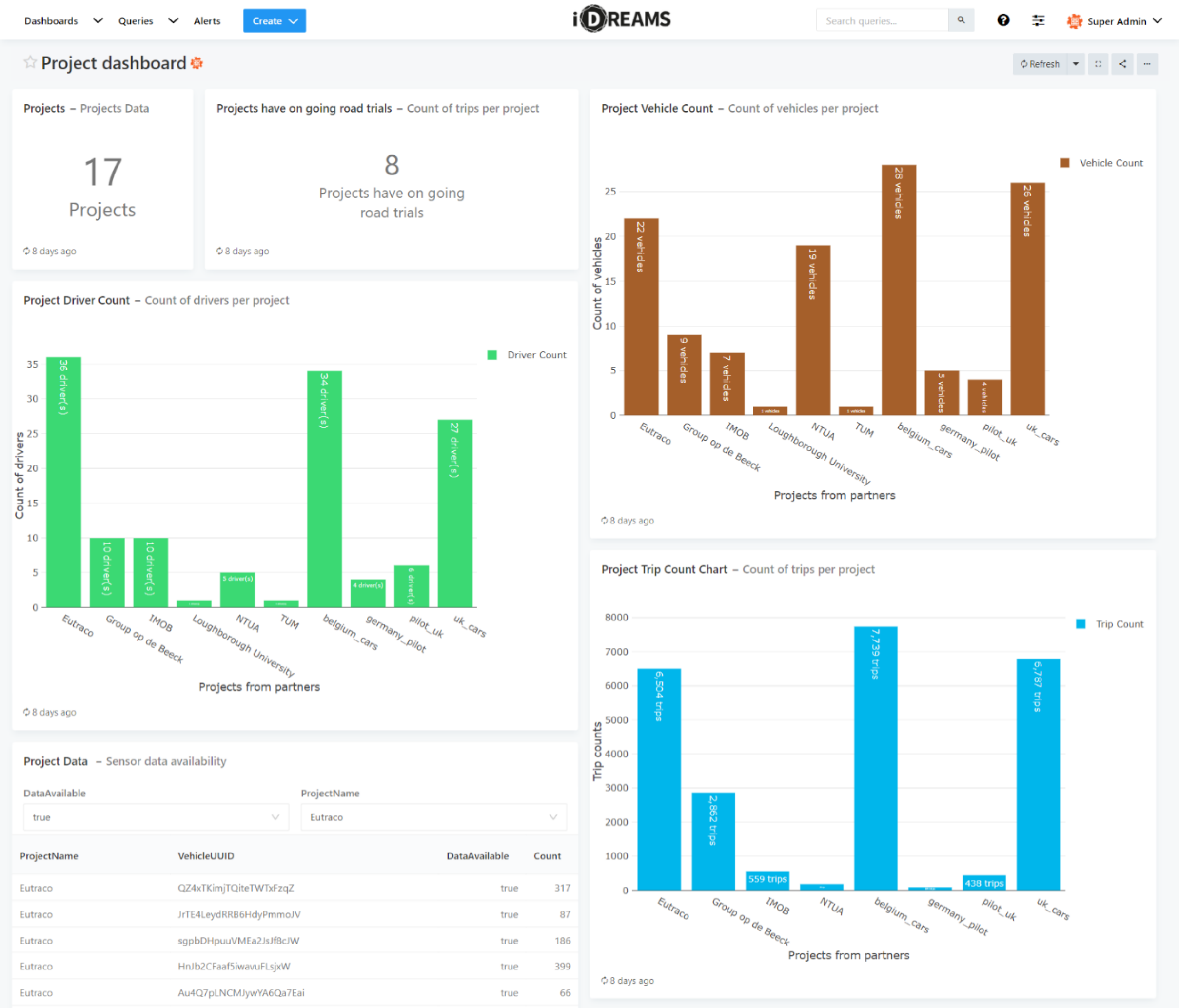

Figure 9: Dashboard Visualization (b)

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

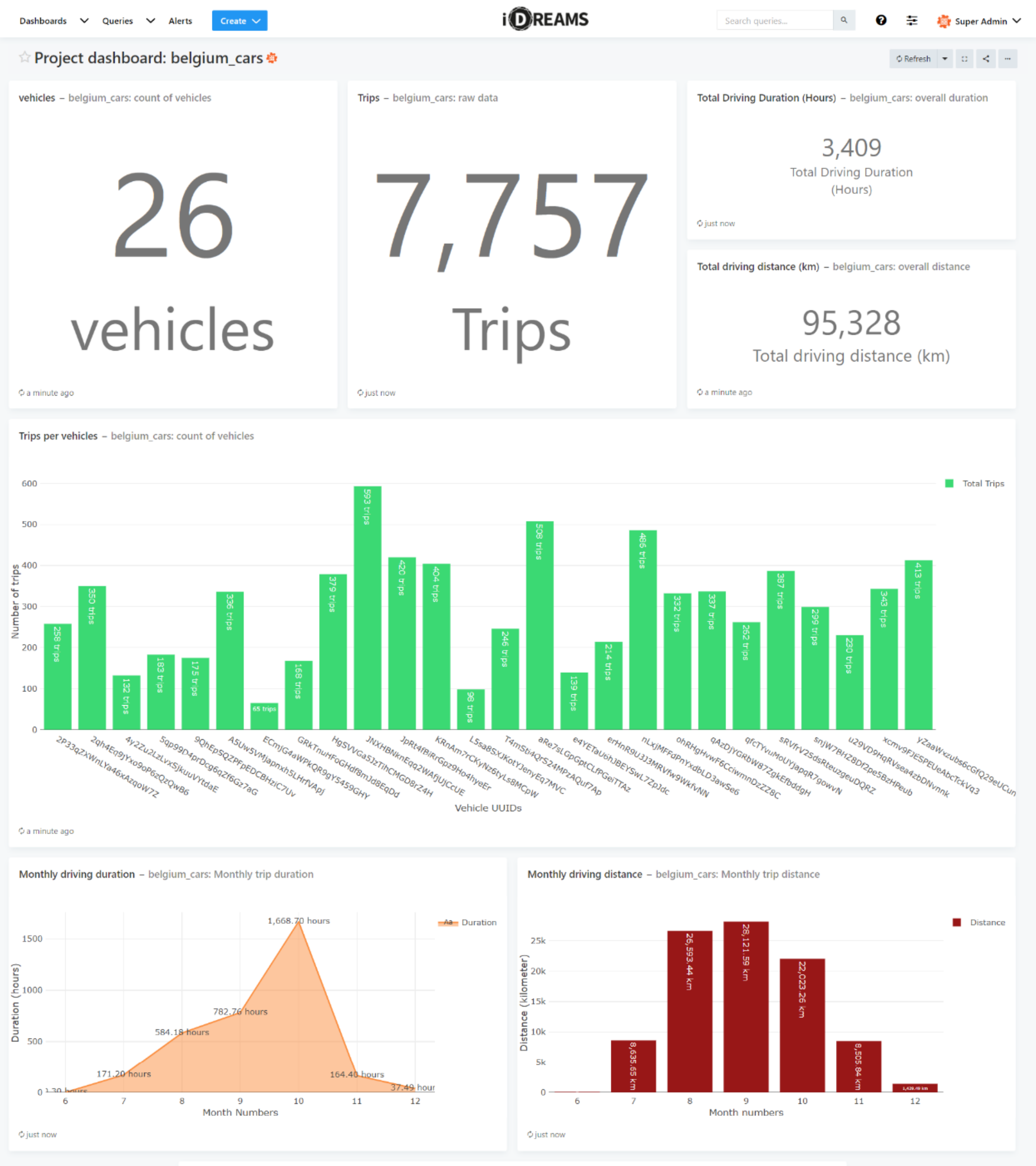

Figure 10: Dashboard Visualization (c)

3.4.3 Fusion and analytics API and data storage

The fusion and analytics API is part of the communication component of back-office which includes other APIs like data persistence API, data access API etc. Back-office APIs are coded in Python which are hosted in a Python-based web server. More details are found in the deliverable D4.3 of i-DREAMS deliverables. The documentation of the back-office data access API is attached in the appendix.

In the back-office, MongoDB is already used as the data storage of raw and processed data. The analytics service component, upon receiving a request via the API, fetches the data stored in the MongoDB and then processes the data accordingly. The results of the analysis then go back to MongoDB for later usage.

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

4 Proposal for the i-DREAMS post-project data sharing model

Transportation research has seen a fast growth in the number of Field Operational Tests (FOT) and Naturalistic Driving Studies (NDS) performed across the world. The main driving force behind the surge of the growth was to better understand the benefits of safety systems and the factors that cause the occurrence of incidents and accidents. Immense volume of data has therefore been collected through conducting naturalistic driving projects and field operational tests. Consequently, due to the need of substantial effort and funding to run these FOT/NDS, the interest in data sharing has got the attention it lacked to optimize the outcome of such large-scale projects. NDS literature is still missing the data sharing related research which could help in understanding the challenges and prospects of data sharing.

In the following section, we will first present the data sharing model and then discuss the associated challenges to overcome to make open access data sharing a reality. Then we will briefly look into previous NDS projects and find out how those projects dealt with those constraints and what model of data sharing they have considered. Finally, we will look into the current status of i-DREAMS data to check the feasibility of adopting the open access data sharing model.

4.1 Open access model

Open access (OA) refers to a set of principles and practices that enables access to research outputs through online distribution without any monetary cost or other access barriers. According to the Budapest Open Access Initiative –

"…There are many degrees and kinds of wider and easier access to this literature. By open access to this literature, we mean its free availability on the public internet, permitting any users to read, download, copy, distribute, print, search, or link to the full texts of these articles, crawl them for indexing, pass them as data to software, or use them for any other lawful purpose, without financial, legal, or technical barriers other than those inseparable from gaining access to the internet itself. The only constraint on reproduction and distribution, and the only role for copyright in this domain, should be to give authors control over the integrity of their work and the right to be properly acknowledged and cited…"

In the light of the definition above, Open Access (OA) data means making the data available for the public through distributing the data online without any technical, monetary or strict copyright barriers that hampers the use of the data.

4.2 Challenges of data sharing in NDS

Challenges in data sharing come from many directions. Volume and structure of the data plays a crucial role for designing/choosing systems that would host the data. Policy regarding the access of the hosted datasets is influenced by legal and ethical aspects which vary depending on the scale and locality (within or across country borders) of the data sharing. The availability and accessibility of data itself does not make the data usable since a proper description of the dataset is usually required to understand the context and reasoning of data collection and quality of the dataset. Data analysis would heavily rely on these aspects. So a proper metadata needs to be available along with the actual datasets. A substantial effort is required to produce such metadata. So, funding specifically focusing on the data sharing aspects would then be crucial to make this possible. Since extensive efforts are made to collect data and build up the data infrastructure and tools, researchers often become reluctant to share the data with third

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

parties to retain the ownership of the data and to be able to exclusively utilize the data. Such a mentality is quite common in the research world and often works as an influential force to decide on data sharing. It is important to find mutually beneficial sharing models for both data collectors and end users, for re-use of the data to become a reality.

For many of the previous and on-going projects, the concept of data sharing has been an internal matter and the re-use of data after the project has not been in focus. Therefore, some projects lack the portion of the funding in the project agreements to make it possible to share data. Data documentation level remained quite low (just enough to make sure the project partners can re-use the data) which makes it hard for outsiders to properly utilize the data. New laws putting a stronger emphasis on privacy preservation have also enhanced the need for a larger extent of data protection. Funding problems at the end of the project have been quite common which quite commonly leaves the fate of the data to the partners' goodwill.

In the following subsections, we will discuss the constraints of data sharing in naturalistic driving studies in detail.

4.2.1 Data ownership

NDS projects in general have several partner organizations that invest their resources to collect and process the data. All involved partners need to agree on sharing the data within a specific model- in this case, open access (OA). If one or more of the partners are negative about providing OA to the data, then an intervention would be needed otherwise OA won't be possible for the data. Oftentimes, licensing could be a way to deal with particular rules that partners may want to impose. But it is important that licensing needs to be under similar conditions as already available OA licenses. The initial stage of the project is important to set the way to share data while keeping the OA model in perspective after project lifetime. At this stage the awareness of the topics and issues regarding OA data sharing and re-use of data helps getting ready to face the challenges of data sharing from many aspects including reserve the right funding portion to facilitate sharing related costs, distributing the tasks leading to smooth data sharing among partners, building tools in a way that supports the sharing of data without too much additional effort etc. The set of topics that should be addressed includes - the ownership, distribution and access to data and data tools, the nature of data usage that would be allowed during and post-project (re-)use of data and last but not the least, post-project financing to host the data and tools supporting the data.

4.2.2 Availability of sufficient metadata

NDS projects collect a large volume of raw data with continuous data collection over a long period of time. Depending on the aim of data re-use, both raw data and transformed/aggregated data are useful. The latter requires data to be cleaned-up, derived, annotated before any analysis on the data. Usually, partners who are in charge of such tasks have enough knowledge to understand the raw and derived data. An end-user who was not part of the project would not know the way to assess the quality of the data. That's why data documentation is in the core of data sharing. In other words, good metadata is vital to successful data re-use. Extensive and high quality metadata, e.g. the conditions in which it has been collected, for which purpose, how it has been stored, processed, and how it can be accessed etc. are vital. In the case of OA, the targeted population of researchers could be even bigger. So more attention to the metadata production would be required.

Metadata can be divided into four different categories, surrounding the actual data:

- FOT/NDS study design and execution documentation, which corresponds to a high level description of a data collection: its initial objectives and how they were met, description of the test site, etc.
- Descriptive metadata, which describes precisely each component of the dataset, including information about its origin and quality.

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

- Structural metadata, which describes how the data is being organized.
- Administrative metadata, which sets the conditions for how the data can be accessed and how this is being implemented.

When we talk about data, we think about a dataset that is the result of a data collection process. In reality, data comes from an iterative process that comprises pre-processing, integration of multiple data sources, derivation of different measurements, data reduction etc. Data takes different forms in these iterations. The data can take subjective or objective forms of contextual, raw, derived (e.g. aggregated) data.

The study design and experimental procedures must be well documented to enable analysis. The main goal of this documentation is to describe the scope and purpose of the data collection, the experimental procedures and the important details of the actual execution including description of the test site, which are important to know before interpreting the data. The document should give an overview of the purpose for which the data collection was made, research questions, sample selection criteria and description of participant recruitment, description of used equipment, chronological record of different phases of study, test plan and execution, how contact was kept during the study.

Descriptive metadata refers to definition of the actual dataset and would include detailed description on measures, performance indicators, time and locations segments, and their associated values. Also external data sources, subjective data from self-reported measures and situational data from video coding must be described in detail. The knowledge of data generation and processing is equally important to preserve as the output of the data. So both of these need to be described well. This translates to building trust on the data for external users who don't know about the data. The description of the data processing should include information on data type, precision, unit, sample rate, error codes and quality indicators along with synchronization policies, re-sampling filters and harmonization rules etc.

Structural metadata describes how the data is being organized in a system (e.g. a database or a file system, with a certain structure or database schema, etc.). It aims to facilitate the initial phase of data re-usage by providing the necessary know-how on the data organization. It is important to also preserve tools that can read and process the data and how to use these tools as the cost for building up the tools again might exceed the data value when the project ends.

Administrative metadata is information that is collected for effective operation and management of data storages and catalogues. Administrative information is stored along with the datasets, covering various topics. The key role of administrative metadata would be to cover access conditions, rights, ownership and constraints. The administrative metadata has a role also in data protection, defining processes, personal data management, access rights and keeping track of e.g. periodic backups.

4.2.3 Data protection

OA data sharing model allows researchers from across the world to use the data in various ways. NDS data sets contain privacy sensitive information that can be traced back to individual participants; this will constrain the publishing of the data from a legal perspective. Willingness to share data increases if the data provider (e.g. participant of a road trial) knows that the researchers have defined and proved procedures in place to keep control of who is accessing the data and that the researchers have knowledge in the legislation surrounding the handling of personal and IPR data. Two types of data need specific protection: personal data and proprietary data. Personal data is classified as sensitive personal data, and more general personal data (European Directive 95/46/EC Art. 2 1995). Personal data could include video of the driver or video of persons outside a vehicle, GPS, or other attributes that can help identify a person. Proprietary data, if revealed, could potentially harm a commercial company. The provision of proprietary data is usually accompanied by agreements, stating the conditions for access and use. Other data is by his classification considered non-sensitive data. The data

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

protection requirements of this data are of course less restrictive. If the dataset consists of personal, sensitive personal or confidential data, it is mandatory to take actions to ensure data protection, even for a minor dataset. If the data is classified as non-sensitive there are fewer requirements that are mandatory, but it is still recommended to investigate all requirements.

So, the agreements with the participants of the project have substantial influence on the possibility to re-use data after the project within the OA model. These agreements come from both the project partners and by external parties like participants and external data providers. It is vital to decide and act on these issues because once the project ends it could be difficult to reach participants and external data providers to come to an agreement. As each participant allows the project to follow the participant´s private life for a period ranging from a few weeks up to more than a year, it is important that they have a solid understanding of what the data could be used for. The participant should make an active consent to the most vital topics for data sharing. Additionally, to overcome privacy constraints, data anonymization by filtering of sensitive information and aggregation of data are useful which lead to only publishing a selection of data properties and values.

Data access can be achieved in four different ways. The data could be downloaded via a public or restricted website, transferred on hard drives to the research organization, remotely accessed at the data provider or only accessed at the premises of the data provider. Each method has its own implications and it is usually the data type that has a large impact on the conditions for selecting a method.

4.2.4 Data sharing platform

Data sharing requires a platform where the data will reside and will be served from. This platform should be flexible to meet the needs of individual end-users; OA data sharing model implementation should take into account the privacy and legal aspects and incorporate input from different partners. Development of such platforms often requires extensive resources in terms of man-power, money and time. A data sharing platform needs to be monitored and managed properly to ensure continuous sharing of data.

4.2.5 Support and research services

The result of re-using third-party data is dependent on the level of documentation and usually also on the available support, especially if the researcher is not so familiar with the type of data and in the case of OA, researchers from a very different field could be interested in using the data. Having research and support services to assist the researchers would be fruitful to achieve the data sharing goal. Support services are there to assist researchers during the analysis, whereas research services are more targeted towards doing part of the analysis work, such as extracting usable datasets for the data re-user or even perform the whole research itself and provide results. The support starts already at the application stage, where the need for support during the analysis is established. The services could be divided into these areas- information and data provision, supporting tools, assistance on dedicated research needs and data protection and analysis facilities. The research services go beyond the initial start-up provided by the support services. They could be divided into the following levels; (1) research advice on methodology; (2) research involvement / research support; and (3) complete research performance. As part of such services, training for all personnel handling NDS data would be highly beneficial to ensure efficiency and reliability in the data sharing process.

4.2.6 Funding

The collection of large datasets involves a huge amount of effort and resources. To further benefit from these datasets and make better use of the invested resources, the datasets should be made available for further re-use. If the dataset is large, the funding becomes crucial, as the cost of maintaining and providing the data becomes substantial. The areas where data management costs are incurred include data selection, enhancement of documentation (metadata), creation of entries in relevant data catalogues; management & coordination

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

personnel costs; IT operations; data handling facilities; support services; promotion and advertisement; and standardization and collaboration regarding dataset formats (optional). Strict requirements to uphold user privacy and product IPR require secure facilities and processes and thereby raising management costs of such datasets.

4.3 Data sharing in previous NDS

A lot of NDS have happened over the last two decades. In this section, we will review some of the biggest NDS projects to understand the data sharing scenarios that took place in those projects. We will specifically discuss how those projects dealt with the topics discussed in the previous section. The projects that are discussed includes- Second Strategic Highway Research Program (SHRP2) NDS, U-DRIVE, Australian NDS (ANDS), Shanghai NDS (SH-NDS) and Track&Know.

Information presented in the next sections are collected in numerous ways. We have gone through research papers, publicly accessible project deliverables, websites of projects where available etc. In some cases, we have personally contacted project partners and sent out a list of queries which were answered and sent back to us. To have a focused discussion, in each of the following sections, we will present a set of questions and the discuss the information we have against those questions.

4.3.1 Data ownership

An NDS project is commonly undertaken by a group of partners. This brings about questions regarding technicalities including ownerships and distribution of responsibilities or resources between partners during and after the project. In this section we look into the “Project Agreements” of several NDS projects to see how they envisioned the ownership of the data.

Australian Naturalistic Driving Study (ANDS), an NDS project which was mainly coordinated by three Australian universities, formed an NDS governance group to oversee further use of the data acquisition system and Mobileye units and data in future studies. This group was chaired by the two main coordinators of the project and involved representatives of each university and the industry/government partners. All equipment and data collected are owned by the University of New South Wales (UNSW) and Monash University. Exclusive licenses were granted to each Australian university to access and use the equipment and the data during the study for research purposes. Each Australian university partner and the Virginia Tech University held a copy of the raw dataset at the end of the project.

There is no available information regarding ANDS’s method for data distribution between partners, however, to access the data, each Australian research team would need to generate data queries (using Client software installed on their own computers) that were transmitted to the Data Center at Virginia Tech University in the United States via an Internet portal. The results would then be transmitted back to the researchers using the same Client software. The data pre-processed by Virginia Tech University would be downloaded regularly and stored securely in a Database Server at UNSW. Trip data of ANDS were only accessible through the OzSight web page maintained by Virginia Tech University for the consortium members and approved analysts who were directly supervised by the consortium members, and only a limited number of analysis workstations approved by the ANDS Data Access committee would be allowed data access. The OzSight website provided general data access and various trip summaries in a more general sense using a browser. Essentially, non-sensitive data could only be accessed from this website.

Another major NDS project, Second Strategic Highway Research Program (SHRP2) NDS by Virginia Tech Transportation Institute (VTTI) unfortunately did not publish publicly the detailed data and ownership and distribution between partners. However, given that SHRP2 gathered

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

the data with the purpose to facilitate road safety and driver behavior analysis in the framework of SHRP2, planning of such procedures should exist.

Regarding the usage of the data, ANDS limited the research and commercial areas where the data were allowed to be used. ANDS's project agreements defined seven allowed key research themes such as- safety at intersections, speed choice, interactions with vulnerable road users, fatigue, distraction and inattention, crashes and near-crashes, interactions with intelligent transport systems (ITS) etc. After the project was over, each university partner (and through them each government/industry partner) continued to have access to the data to undertake further separately-funded analyses. Unfortunately, information regarding the arrangements for access to the data by others are not available publicly.

Similarly, SHRP2 also allows (re-)use of the data after the project was concluded. VTTI has envisaged that the data would be heavily used by researchers for several years after the end of the program, providing insights in driver behavior and safety and creating impact worldwide. As stated in the project's final report ref: "The goal of this unprecedented data collection effort was to create a rich data resource for researchers, regulators, advocates, students, and other interested parties all over the world to analyse and address many of the key transportation safety research questions for at least a generation to come."

In contrast to SHRP2's openness to share the data, the data in the project Track&Know (Big Data for Mobility Tracking Knowledge Extraction in Urban Areas, coordinated by INLECOM) came from the three industrial partners which therefore have the full ownership of the raw data. The data used in this project will remain within the premises of these industrial partners and are retained for continuity purposes. There is a repository that exist (within a project website) which contains the software tools developed during the project and majority of the software can be freely accessed, however, data sharing is restricted to an extent.

As an exception, the agreement also allowed the consortium to not provide open access to data that would jeopardize achievement of some objectives. However, data management plan needed to provide reasons for this. According to the data management plan, data preservation of at least 1 year after the project was required by the Grant Agreement. The associated costs for dataset preparation for archiving would be covered by the project itself. Long-term preservation of code and open datasets would be provided, and associated costs covered by a selected disciplinary repository. Proprietary and sensitive datasets would remain the property of the owners.

It should be noted that the project contained several deliverables that are 'confidential' primarily due to the nature of the project that involved development of a big data management tool that could be exploited and potentially patented.

4.3.2 Availability of metadata

Metadata involves any piece of information necessary to use or properly interpret the data. In the context of NDS, it may include information on study design and execution, data organization, etc. SH-NDS does not provide any descriptions of the data. Contrariwise, SHRP2 and ANDS offer several documentations related to the data. The following lists the data documentations published by SHRP2:

- project objectives and framework
- the study preparations
- human subjects' protections and participant management
- DAS design and equipment issues
- data collection site

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

- data quality, management, and processing.

Moreover, further documentations of these types can be found with the processing and manipulation of the raw data before it concludes to specific variables in the final "product", e.g., radar data pre-processing. Also, a thorough data dictionary with all the variables (and units) are explained.

On the other hand, ANDS provides considerable information on the study design and execution. This includes project commencement date (2013 and 2014) and the study regions (New South Wales and Victoria). In addition, the sample size (400 participants) and general descriptions (e.g. each driver will have their own car instrumented with data recording equipment for 6 months, allowing the researchers to record continuously their driving behaviors and those of others with whom they interact) are also made available to the public. As for the objectives, they have mentioned that the focus was not primarily on crashes, and the research themes of interest are ones involving behaviors that occur often and regularly.

Moreover, ANDS also provides documentation on the data organization in this project. The said documentation explains that a data management office at two of the participating Australian universities (UNSW and Monash University) securely housed the data transfer and storage equipment: a "Drive Bay" (which read the raw, encrypted, data recorded by each DAS hard drive and transferred it to a "Staging Server"); a "Staging Server" (which stored temporarily the data); a high bandwidth internet connection (for transfer of the raw data to a Central Staging Server located at UNSW); and USB drives (which were transported to UNSW, if there was data loss during internet data transfer). In addition, the management office at UNSW had a machine-to-machine (M2M) mobile telephone link to enable remote communication with each DAS unit in the field.

The raw data from vehicles were sent to a central computer at UNSW. It was then sent via the internet to the Virginia Tech Data Centre in the US, where it would be pre-processed into a form ready for analysis by the Australian research teams and securely stored in a database. To prepare the data for storage and analysis, the data needed to be de-encrypted, synchronized (as not all data were collected at the same rate), and filtered (to eliminate outliers, etc.). This was done by the Virginia Tech University under contract to UNSW. To access the analysis-ready data, each Australian research team (NSW, VIC, QLD, SA) generates data queries (using Client software, installed on their own computers), that are transmitted to the data center at Virginia Tech via an Internet portal. The results were then transmitted back to the researchers using the same Client software. The data pre-processed by the American university would be downloaded regularly and stored securely in a Database Server at UNSW. Every 4 to 6 months, the near-full DAS hard disk, containing the continuous data, were removed manually from the host vehicle and swapped for a fresh one (unless the participant has finished the study).

Most of the technical deliverables are confidential in the Track&Know project. However, public deliverables and scientific papers published on the website provide useful information about study design and execution of several analysis methods where dataset acquired in this project were used to develop and further validate them. An example is the development of a machine learning-based crash prediction model. The model utilized trip features extracted from raw GPS dataset for vehicles to predict involvement of a vehicle in an accident in the near future.

Track&Know also provides the data management plan that contains the information regarding the three major sources of the datasets, as well as how these data were obtained (primarily raw data collected from On-Board Devices (OBD) from vehicles). There was one dataset which was manual and contained patient appointments, referrals, and records of their journeys to

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

available medical services. The data management plan also describes metadata of such a dataset, especially all the variables information and their units of measurements. Additionally, from many open access sources other datasets were acquired and fused with the dataset to further enrich the dataset. There are a few deliverables that are confidential including ones that contain the method to analyze the quality of dataset by visually inspecting them and applying various cleaning/filtering mechanisms to prepare the raw data for other meaningful analysis. The project developed a Big Data processing toolbox that contains many processes/algorithms to perform such processes. A few of them are as follows:

- GPS data cleansing and map-matching
- GPS data enrichment (weather, POI, etc.)
- NoSQL data access operators – the NoDA API
- Distributed trajectory joins
- Distributed sub-trajectory similarity search

Similar documentations also exist for the U-DRIVE project. Some documents were publicly available, while others were restricted. Unfortunately, the access to these deliverables is difficult as the project's website no longer exists.

The documentation of the U-DRIVE project was flawed as an interview with one of the project contributors revealed that most meta information was lost. These include aggregation levels of the data, detailed processes that were applied on the data, general information about the sensors and how the data was treated, data dictionary, codes which link the pseudo-anonymized questionnaires (related to driving behavior) to the drivers, etc.

4.3.3 Data protection

As briefly mentioned, data in NDS projects are sensitive and require the projects that employ these data to implement data protection measures. As with the other projects in this review, SHRP2 included data protection clauses in the data sharing agreements and implemented thorough rules for data security and management. Part of the process included paperwork with the IT services of the university to reply to questions and confirm there was the necessary equipment and actions have been taken to be able to receive and store the data appropriately.

Data protection procedures were also implemented by U-DRIVE. Contract agreements were written as to how data could be used and accessed, which every data analysis site had to agree to. There is also an agreed procedure if an institution wants to become a data holding site after the completion of the project, in which approval needs to be given by two existing sites, Charmers and DLR. The procedure and regulations are detailed and the aspiring data holding site needs to show evidence that the conditions could be met along with the local GDPR.

Meanwhile, during the Track&Know project, the data provider (in this case, also the data owner) was mainly setting the terms on data protection. For analysis purposes they were bound to provide raw datasets (to become an EU consortium beneficiary in the project) to technical partners. Therefore, a data sharing agreement between each technical partner and data provider partner had to be established. The following data protection measures were ensured in the data management plan of Track&Know:

- Remote access would be provided through secure, encryption-protected connections, granting authorizations only to personnel and activities relevant to the project, and within the time frame of the project.

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

- Data storage in a safe location, with physical access limited to authorized personnel for hardware /software maintenance and partners of the project. Safe data transfer through secure, encryption-protected connections.
- For all project partners involved in the processing of personal data, it would be required to apply relevant data protection measures prescribed by the General Data Protection Regulation (2016/679). 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 Track&Know project privacy policy.
- There was a data protection officer appointed in the project with key responsibility to ensure that data collection and processing within the scope of the project, was carried out according to EU and national legislation. The DPO reports to the Track&Know Project Management Board (PMB) through the coordinator.
- All processing activities involving personal data were monitored and recorded, providing information about the purposes of the processing, the agents involved, and time period of the activity. The records were kept and made available upon request to the supervisory authority.
- In order to design security procedures, it is common to classify the data into different groups so that the measures are well-suited to the sensitivity of the data. For example, in ANDS, the data are grouped into three different levels:

  Personal Data: data related to an identified or identifiable person, e.g. a participant; an identifiable person is one who can be identified, directly or indirectly, in particular by reference to an identification number (ID number) or to one or more factors specific to their physical, physiological, mental, economic, cultural, social identity or their travel patterns

  Confidential Commercial Data: 'Open' where all approved analysts and the project consortium partners can access the data freely; 'Closed' where confidential commercial data is open to all consortium partners during the project on a per project approval provided by the Commercial stakeholder and the consortium that includes the Core Research Group; 'Confidential' data owned by the stakeholder that is never shared, as the commercial value of the data is too high for data sharing. The latter will apply to Seeing Machines data.

  Non-sensitive data: completely anonymized data and do not include any confidential elements; no personal identifiable data is available in the dataset (i.e. video, images or GPS traces) can identify the participant.

SHRP2 and SH-NDS also conduct similar classification practice, however no information of such detail is available. On the other hand, all the data in the U-DRIVE fell under the "sensitive" classification and the access to such data was restricted due to lack of GPS clipping or face blurring.

Moreover, role classifications were also at place at some of the projects. As an example, data storage in the U-DRIVE project was defined as e.g., the central data center or local data center. There were also operational roles such as being an operations site (test site) or a data analyst. Meanwhile, the consortium of the Track&Know project was built in a way that some partners were:

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

- **Data provider**: could also be considered as clients that wanted to solve specific problems that could bring new opportunities for enhancing their business using their datasets
- **Data analysts**: mostly academic institutes which had expertise in different domains such as machine learning, data visualization, big data management and processing, etc.

After the project lifespan, roles are not clearly defined.

These data and role classifications are closely related to the categorization of the access methods as a class of data may be only accessible by certain groups of users or through certain access channels. In this regard, access to the SHRP2's data access website is controlled by user access level. The information available for someone to view and manipulate is determined by a user status. By federal law, VTTI must monitor usage of these data to assure compliance with the data privacy requirements established by the responsible institutional review boards (IRBs). This website supports three user access levels:

Guest User is granted access to- project background documentation, aggregated and summary graphs, access to data dictionaries. The registered user is granted access similar to the guest user access plus access to User forums, access to personal profile, ability to apply for qualified researcher. Qualified Researcher is granted access similar to registered user plus access to all available datasets, access to custom queries, access to crash, near crash, and baseline events. To become a Qualified Researcher (QR), one needs to upload a valid IRB Training Certificate (attend the related training course) and apply for Qualified Researcher status.

Meanwhile, regarding access methods, ANDS ensures that trip data are only accessible through the OzSight web page that is maintained by Virginia Tech University for ANDS Consortium members and approved analysts who are directly supervised by ANDS Consortium members, and only a limited number of analysis workstations approved by the ANDS Data Access Committee will allow data access.

Concerning ANDS, it was also possible to download specific trip data that includes both non-sensitive and personal data, using different programs such as Matlab, Python or R programs, and other programs, by using an API (Application Programming Interface) Key Portal. However, only a selected number of approved computers where security was ensured would be registered for data access. It should also be noted that there is not any public access to the data other than the information provided in publications that have been approved.

During the duration of the Track&Know project, various methods were used for partners to access the available data for analysis purposes depending on the data owner (company rules/regulations), sensitivity of the data, etc. For example: Health related data was made available after signing a special agreement with each partner and this data was then shared through a restricted website. Data from Systematica (Trajectory data) was only accessible only on premises of the data owner because of the strict privacy agreement with the company customers. Only a specific aggregated part of this data can be made available after a formal request for third parties. Datasets from Vodafone Innovus were made available to technical partners for analysis using a restricted website. The data is historic and not available as a real time data stream. The real-time data stream was available to specific partners using a remote access at the data provider for specific duration only.

As previously mentioned, the data in the U-DRIVE project sat in one of the partner's servers and were accessible remotely through a special access software/tool that required secure

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

connection. Although eventually the cost of such a method became prohibitive that the data were then transferred physically between partners using hard drives. In spite of this, agreements were still in place to ensure data security.

4.3.4 Data sharing platform

As aforementioned, the data from SH-NDS and ANDS are not publicly shared. Hence, no data sharing platform was initiated. However, to facilitate data management and transfer between partners, ANDS The raw data from vehicles are sent to a central computer at UNSW. It is then sent via the Internet to the Virginia Tech Data Centre in the US, where it will be pre-processed by them into a form ready for analysis by the Australian research teams; and securely stored in a database. To prepare the data for storage and analysis, the data need to be de-encrypted, synchronized (as not all data are collected at the same rate), and filtered (to eliminate outliers, etc.). This is done by the Virginia Tech University under contract to UNSW. Every 4 to 6 months, the near-full DAS hard disk, containing the continuous data, is removed manually from the host vehicle and swapped for a fresh one (unless the participant has finished the study).

On the other hand, the data from the SHRP2 project are publicly obtainable by request. The data become accessible and downloadable (according to the specific data request) through a link that is obtained per email from the coordinator. More related information including data collection equipment, installation fixtures and software, assessment equipment and software, and data ingestion equipment are discussed in the Naturalistic Driving Study: Technical Coordination and Quality Control report. The Track&Know project utilizes the project website (https://trackandknowproject.eu/file-repository/) as a platform for data sharing. All metadata information is available with complete description on anonymization and synthetic generation of data that is similar to reality.

4.3.5 Research and support services

Concerning research services (e.g. providing advice on methodology or evaluating research performance), none of the NDS projects that are reviewed are known to provide such assistance. Though, regarding technical support, SHRP2 provides assistance mostly through email contact on anything relevant with the data access process, data request, technical details, practicalities, etc. Moreover, once one reaches the stage to download the requested data, support is provided throughout their project for possible enquiries or other issues. The technical partners of Track&Know have also deposited many algorithms/methods codes to GitHub repository to facilitate other developers to reuse it.

4.3.6 Funding/Financial models

An NDS is most probably a resource-intensive and costly project. To know the best financial model to fund the project would be valuable for future NDS. (Gellerman et al., 2016) classified the financial model of such projects into three models based on the funding source: direct governmental, project-based, or end-user. Unfortunately, no detail is publicly available regarding the financial models employed in SH-NDS, ANDS, and SHRP2. Yet, given that a fee is applied to every data request to SHRP2, it can be assumed that this could be used for data maintenance and technical support services by the SHRP2's coordinators.

Concerning Track&Know, it is not clear where to fit the project based on the 3 financial models mentioned above, as a few features of direct governmental and project-based model were applicable. Though, there were partners involved that acted as raw data providers (which were also responsible to come up with research questions) and some technical partners to solve these research questions using the available dataset. The Horizon 2020 EU funding made it possible for industrial partners and technical partners to work together in a consortium that

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

developed methods for big data management/processing and analysis which were demonstrated through the specific research questions that industries have with use of their own datasets. The industrial partners were sole proprietors of the dataset.

4.3.7 Research findings and impact

The projects that are discussed in this review are some of the most important natural driving studies that have been conducted. The impact of these studies may be quantified by the number of publications that these studies gave rise to (i.e., citations). To the date when this review was written, the SH-NDS study has published 9 publications and been cited with many papers, while the 7 papers published from ANDS have been referred to by a total of 51 studies.

Majority of the Track&Know papers were published during the lifespan of the project. However, there were some planned publications that were published after the official end of the project. There are around 39 publications including journals, conference proceedings, and book chapters as stated in the project's website. These publications are estimated to have been cited by more than 50 studies.

This deliverable is also interested in whether or not the NDS projects made any code or prepared data that were used in the research findings accessible to lay a foundation for further studies. The feasibility of applying and/or verifying the research findings on (and/or, using) data from sources is also reviewed in this deliverable. Track&Know project discussed this topic in its data management plan. Most of the code that was generated in this project is available on Github and other repositories. Code descriptions with the input and output are also mentioned on the project website. There are in total 18 items that are currently openly available. Although the prepared data is not readily available, there are some open data shared publicly that are synthetic and far from reality. Interested researchers could, however, understand completely the type and structure of the data used in the project as the meta data are described sufficiently. They could also validate the tools developed in the project. It is worth noting that the open data from Track&Know are not meant for any decision making.

SHRP2 NDS is one of the most impactful NDS projects due to the availability of the data. Plenty of papers from global researchers and other parties have been published in several journals and conferences. SHRP2 NDS also lists in their website several SHRP2 publications, projects, extracted reduced datasets, reports and other details.

4.3.8 Other important issues

Firstly, among the study challenges in SHRP2 NDS, participants that asked to view or obtain a copy of their data are mentioned. Protocols and processes were established in order for the video to be shown only to the participant and without releasing it to anyone.

ANDS raised the issue that since the study involved video recording of driver behavior and modifications to participants' vehicles, the main challenge or issue was how to handle illegal driving behavior of participants. For example, whether to share or report to authorities, insurance companies, etc. if drivers speed or violate traffic regulations. Legal advice was that if serious traffic offences are detected when going through video analysis, they would need to be reported to relevant authorities but that this should be done on a case-by-case basis.

A significant potential ethical issue in ANDS that caused a number of concerns relating to privacy was that the person driving the instrumented vehicle was readily identifiable on video. This problem has been solved by pixelation of driver obscuring faces in all images that will be shown publicly and ensuring that analysis of video data is conducted by researchers who have signed confidentiality agreements and under conditions in which only authorized researchers have access to the video material

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

SH-NDS also agreed that data privacy was a significant issue that could obstruct data sharing. SH-NDS mentioned that the most challenging parts to make the data sharing a reality is the data protection and commercial value. Track&Know is not sharing any data that is representing reality due to a variety of reasons:

- The data used in the project is based on position information and also contain health related information which is very personal data therefore, it is not shared publicly
- the industrial partners are the sole proprietor of the data, and their business is tightly coupled with collecting/managing and analysing such data
- Only a small quantity of data is shared publicly, which was anonymized and synthetically developed by manipulating the original data in various ways. This was done with an objective that researchers could have a sense of what type and structure of data used in the project and to validate a variety of tools/software/methods developed within the project.

A U-DRIVE representative mentioned in an interview conducted for this review that the project was hindered due to poor camera quality. It was challenging to obtain the footage of the driver's eye gaze which would have been valuable for the analysis. Lessons could also be learned from the experience of U-DRIVE that much meta information was lost due to the aggregation levels and processes that were applied to the data. This became significant when a tool designed to access the data was used as opposed to simple access to the raw data. Said use of the tool involved automatic processing of the data which was unclear in terms of what and how. General information about the sensors used in U-DRIVE and how the data were treated by the sensors was also lost.

Another issue that the U-DRIVE project stumbled upon was the lack of a comprehensive data dictionary. There was documentation of codes used, however it was unclear what these codes do. Therefore, it is important for NDS projects to have a ‘corporate memory’ for data collection and understanding the data, the technology, the sensors, the limitations, etc. This would ensure that all this information is understandable should the data be used by different personnel in the future.

4.3.9 Data sharing models adopted by different NDS projects

SHRP2 offers a large sample of data available on the SHRP2 NDS data website, information of the variables available and a relevant forum. Nevertheless, the data is not downloadable. Moreover, to get access to the site, one should have an IRB Certificate and have been provided with a Qualified Researcher (QR) status. Aside from that, SHRP2 also makes the following possible on the SHRP2 NDS data access website- background information; detailed data collected from driver assessments, vehicles, trip summaries, and critical driving events; query the database of detailed data, create cross tabulations, and assess NDS database content.

In contrast to SHRP2, ANDS is committed that any information obtained in connection with the study and could be identified with participants will remain confidential and would be disclosed only with their permission or if required by law (e.g., if subjected to a valid search warrant or a subpoena). Meanwhile, SH-NDS does not make any data and data description from the project available publicly.

On the other hand, Track&Know makes some partial data available, but not the full datasets as these are proprietary datasets. Industrial partners of the project such as Systemetica S.p.A (Italy), Vodafone Innovus (Greece) and Royal Pathworth hospital (England) have the ownership rights of the full datasets and for this reason only partial datasets are made available as an open access dataset. Furthermore, these partial datasets have also gone through a special process of anonymization and some steps were followed to make them synthetic. As

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

a result, these partial datasets cannot be considered as a part of original datasets, therefore, the re-usability of such open datasets for research purposes is very limited. However, a request can be made to the owner of the data and there may be a possibility to acquire datasets after arranging a data sharing agreement.

In the case of U-DRIVE, one could request data access to SWOV (coordinator) or other partner data centers. However, since the project was concluded, this may be put behind paywalls and procedures. One may require the partner institution to do analysis of the data on one's behalf or access of the data may be done under one of the partners' supervision.

4.4 Feasibility of the Open Access model for the i-DREAMS data

With respect to the i-DREAMS project, previous published deliverables contain relevant information regarding the project data, study design, but also data sharing and long-term plans for data after the project end. Of course, some things are still not completely updated or completed, and it is the purpose of this activity (the review of previous projects and the initiatives they took), to come up with a set of suggestions when it comes to the i-DREAMS project and the open data initiative.

4.4.1 Data privacy

As already discussed in the various deliverables, partners were required to get approvals from their respective ethical committees prior to starting their data collection. In addition, partners were requested to consult with their local data protection officers, to make sure the protocols they are using for data collection and storage complies with GDPR. Overall, this has been specified in D3.5, but also in the data management plan (D1.2). In the latter, GPS trajectory replacement is also discussed to prevent sharing participants' locations, which could be considered as sensitive data, and a violation of their privacy.

Within the i-DREAMS project, the protocols for handling data have been defined in Deliverable 3.5. This applies for handling sensitive and personal data. The personal integrity and IP rights have been also discussed in the joint agreement for processing personal data (signed between i-DREAMS partners), and being discussed (ongoing) in the framework agreement governing the i-DREAMS platform exploitation.

Overall, partners have been informed in the various meetings to also be aware of data handling and report breaches that may occur (even if this happens by mistake). For example, a partner accessing data that sees personal information that they are not supposed to see (e.g. a participant's name) are to notify the partners responsible for data storage and access for this breach, so that the corrective action takes place as soon as possible.

4.4.2 Support and research services

Support and research services are currently provided by TUM within the project lifetime as TUM is managing the back-end access and storage. It is to be decided however how the resources are going to be managed after the project ends.

4.4.3 Availability of metadata

The data management plan (DMP- D1.2) describes 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). D1.2 describes the dataset types, but also how the data can be made accessible after the project ends. However, D1.2 does not really provide a specific appendix for the specific variables collected. These are available in various documents, e.g., sensor data is available in the API provided by Cardio ID to partners.

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

In relation to the driving simulator data, the appendix of D5.1 contains an overview of the driving parameters and their units (however perhaps not in the same way the variable names appear when the files are exported from the driving simulator). Moreover, questionnaires are provided in D3.4; however, these are perhaps not the most updated ones that were eventually used by each partner (for both simulator and field experiments) and therefore might need to be updated. In addition, when eventually sharing the questionnaire data, there needs to be a unification of the variable names for questionnaire answers, mostly when referring to specific constructs or agreement statements, to be able to better merge data and make use of it.

In terms of access and sharing of the data, this has been defined within D4.2 and the project back-end office documentation. Deliverable 3.4 provides an overview of the development of considerations for the simulator and field trials including the high-risk scenarios used for the simulator experiments. For the latter, D3.4 provides the event types investigated for the different scenarios designed for the driving simulator experiments, but these are not specified in a uniform format, where a tabular format is given with distances and occurrence of events, which would be ideal to provide were the simulator data to be shared with external parties.

D3.4 is further expanded, and the best practices for the implementation of both simulator and road experiments are provided in D5.1. Later, D5.3 describes in detail the on-road trials based on the development from design recommendations and specifications that were presented previously in D3.4, and therefore describes in detail the different stages of the road experiments.

4.4.4 Funding

As mentioned already in D1.2, the project website operation will still need to be determined (in terms of who could financially bear the costs). However, both data archiving at Zenodo (OPENAIRE platform) and copyright licensing with Creative Commons are free of charge. According to the framework agreement with respect to the i-DREAMS platform (still ongoing), following can be done with regards to financing:

- Industrial partners could finance the data infrastructure as set forth in article 4, last paragraph of this Framework Agreement, which is considered to be a fair and reasonable compensation to the Knowledge Institutions; and
- With respect to the Data in particular, the Industrial Partners explicitly acknowledge the fundamental right of the Knowledge Institutions to use the Data for academic purposes. In order to facilitate said academic use, the Industrial Parties agree to finance the infrastructure maintenance and storage of the analytical database for a duration of three (3) years upon the termination of the iDreams Project. Within said period, full access of the Knowledge Institutions to said analytical database is guaranteed.

4.5 Recommendations for open access data sharing model

The iDREAMS project has taken necessary steps to ensure data privacy in the OA model. This includes data anonymization according to GDPR as well as controlling access to the data through data access API. The post-project ownership of data allows OA adopting possibilities. Different deliverables (both publicly available and restricted) contains information about study design and execution, data descriptions and organization, metadata and other administrative metadata while the project is going on. Different partners have taken the responsibilities for the support and research services, although these can be extended further. A proper documentation of the processes can be immensely helpful for post-project usage of data. Several funding options are being explored. Therefore, depending on the funding availability to maintain the platform and infrastructure for data sharing that already exist, decision will be taken whether or not i-DREAMS project can adopt an open access data sharing model for its data once the project ends. With proper funding, data storage will remain online to ensure data

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

availability for third party researchers. API server will ensure restricted access control to the data. Along with purely data centric services, analysis services can remain online for external researchers to make use of. Significant effort will be saved if external researchers don't need to write their own programming processes which may already be in use by the i-DREAMS partners while conducting research within the project lifetime.

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

5 Conclusion

The purpose of this deliverable was to describe the unified big data fusion framework used for exploiting the data collected in i-DREAMS. To understand the necessity and requirements of such framework, we first discussed data fusion through a literature review and then looked at the fusion tasks being used in the i-DREAMS project.

After that, we discussed the big data fusion framework which builds on the architecture of back-office in i-DREAMS (outlined in D4.3). The big data framework for i-DREAMS incorporates data fusion and analytics while enabling real time data processing (including fusion) and analytics for the i-DREAMS data collected from road trials, surveys and simulations. This framework aims to serve i-DREAMS partners to perform large data analysis and to discover knowledge from the data in real time or through batch processing. The framework can fuse heterogeneous data sets from the database and store it back the database to accelerate further analysis and visualization. The framework has the flexibility to support different data fusion algorithm implementations for better support of research tasks. It is implemented in the knowledge discovery component of the back-office to facilitate big data fusion and analysis for researchers.

Another contribution of the deliverable is the recommendation of the Open Access data sharing model once the project ends. The recommendation came from extensive literature review of previous projects and also by discussing the major challenges in data sharing faced by NDS projects.

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

6 References

De Vos, B. et al. (2020). Description of the driving simulator experiment for identifying safety tolerance zones and the performance of in-vehicle interventions. Deliverable 5.2of the EC H2020 project i-DREAMS

AlZu'bi, S., & Jararweh, Y. (2020). Data fusion in autonomous vehicles research, literature tracing from imaginary idea to smart surrounding community. In 2020 Fifth International Conference on Fog and Mobile Edge Computing (FMEC) (pp. 306-311). IEEE.

Bar-Shalom, Y., Li, X. R., & Kirubarajan, T. (2004). Estimation with applications to tracking and navigation: theory algorithms and software. John Wiley & Sons.

Basir, O., Karray, F., & Zhu, H. (2005). Connectionist-based Dempster-Shafer evidential reasoning for data fusion. IEEE Transactions on Neural Networks, 16(6), 1513-1530.

Breima, L. (2010). Random Forests. Machine Learning.

Buitrago, J., & Asfour, S. (2017). Short-term forecasting of electric loads using nonlinear autoregressive artificial neural networks with exogenous vector inputs. Energies, 10(1), 40.

Byun, S. C., Choi, D. B., Ahn, B. H., & Ko, H. (1999). Traffic Incident Detection Using Evidential Reasoning Based Data Fusion. In Proceedings of 6th World Congress on Intelligent Transport Systems (ITS), Held Toronto, Canada, November 8-12, 1999.

Cappelle, C., El Najjar, M. E., Pomorski, D., & Charpillet, F. (2008). Multi-sensors data fusion using dynamic bayesian network for robotised vehicle geo-localisation. In 2008 11th International Conference on Information Fusion (pp. 1-8). IEEE.

Choi, D. B. (1998). On multisensor data fusion using attribute association for intelligent traffic congestion information inference. In Towards the New Horizon Together. Proceedings of the 5th World Congress on Intelligent Transport Systems, Held 12-16 October 1998, Seoul, Korea. Paper No. 4115.

Cremer, M., & Schrieber, S. (1996). Monitoring traffic load profiles with heterogeneous data source configurations. In Transportation and Traffic Theory. Proceedings of the 13th International Symposium on Transportation and Traffic Theory, Lyon, France, 24-26 July 1996.

Darms, M. S., Rybski, P. E., Baker, C., & Urmson, C. (2009). Obstacle detection and tracking for the urban challenge. IEEE Transactions on intelligent transportation systems, 10(3), 475-485.

Du, Y. W., & Zhong, J. J. (2021). Generalized combination rule for evidential reasoning approach and Dempster–Shafer theory of evidence. Information Sciences, 547, 1201-1232.

Dubois, D., & Prade, H. (2021). Possibility theory and possibilistic logic: Tools for reasoning under and about incomplete information. In 4th IFIP International Conference on Intelligence Science-TC 12: Artificial Intelligence (ICIS 2021) (Vol. 623, pp. 79-89). Springer.

El Faouzi, N. E., Leung, H., & Kurian, A. (2011). Data fusion in intelligent transportation systems: Progress and challenges–A survey. Information Fusion, 12(1), 4-10.

El Faouzi, N. E. (2004). Data fusion in road traffic engineering: An overview. In Multisensor, Multisource Information Fusion: Architectures, Algorithms, and Applications 2004 (Vol. 5434, pp. 360-371). International Society for Optics and Photonics.

El Faouzi, N. E. (1999). Combining predictive schemes in short-term traffic forecasting. In 14th International Symposium on Transportation and Traffic Theory Transportation Research Institute.

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

El Faouzi, N. E. (1997). Heterogeneous data source fusion for impedance indicators. IFAC Proceedings Volumes, 30(8), 1307-1312.

Feng, R., Zhang, G., & Cheng, B. (2009). An on-board system for detecting driver drowsiness based on multi-sensor data fusion using Dempster-Shafer theory. In 2009 international conference on networking, sensing and control (pp. 897-902). IEEE.

Friedrich, B., Matschke, I., Almasri, E., & Mück, J. (2003). Data fusion techniques for adaptive traffic signal control. IFAC Proceedings Volumes, 36(14), 61-66.

Gan, Q., & Harris, C. J. (2001). Comparison of two measurement fusion methods for Kalman-filter-based multisensor data fusion. IEEE Transactions on Aerospace and Electronic systems, 37(1), 273-279.

Gastaldi, M., Lamedica, R., Nardecchia, A., & Prudenzi, A. (2004). Short-term forecasting of municipal load through a Kalman filtering based approach. In IEEE PES Power Systems Conference and Exposition, 2004. (pp. 1453-1458).

Granger, C. W. (1989). Invited review combining forecasts—twenty years later. Journal of forecasting, 8(3), 167-173.

Hall, D. L., & McMullen, S. A. (2004). Mathematical techniques in multisensor data fusion. Artech House.

Hashem, S. (1997). Optimal linear combinations of neural networks. Neural networks, 10(4), 599-614.

He, A. X., Zhang, J. F., Zhang, D., & Gan, W. D. (2021). Collision risk assessment in Jiangsu section of the Yangtze River based on evidential reasoning. In IOP Conference Series: Materials Science and Engineering (Vol. 1043, No. 5, p. 052050). IOP Publishing.

Ivan, J. N. (1997). Neural network representations for arterial street incident detection data fusion. Transportation Research Part C: Emerging Technologies, 5(3-4), 245-254.

Ivan, J. N. (1995). Real-time data fusion for arterial street incident detection using neural networks.

Jaramillo, V. H., Ottewill, J. R., Dudek, R., Lepiarczyk, D., & Pawlik, P. (2017). Condition monitoring of distributed systems using two-stage Bayesian inference data fusion. Mechanical Systems and Signal Processing, 87, 91-110.

Keever, D. B., Shimizu, M., & Seplow, J. (2003). Data fusion for delivering advanced traveler information services.

Klauer, K., Pearson, J., Hankey, J. (2018). An overview of the Canada naturalistic driving and Canada truck naturalistic driving studies. In 7th International Symposium on Naturalistic Driving Research. https://www.vtti.vt.edu/PDFs/ndrs-2018/s4/Klauer.pdf.

Kolosz, B., Grant-Muller, S., & Djemame, K. (2013). Modelling uncertainty in the sustainability of Intelligent Transport Systems for highways using probabilistic data fusion. Environmental Modelling & Software, 49, 78-97.

Langheim, J., Buchanan, A. J., Willhoeft, V., Lages, U., & Gyory, G. (2002). CARSENSE-sensor fusion for das. In 9th World Congress on Intelligent Transport Systems (ITS) America, Japan, ERTICO (Intelligent Transport Systems and Services-Europe).

Liang, G., Zhu, X., & Zhang, C. (2011). An empirical study of bagging predictors for different learning algorithms. In Twenty-Fifth AAAI Conference on Artificial Intelligence.

Lim, H., Bae, B., Han, L. D., Chin, S. M., & Hwang, H. L. (2021). A Data-Fusion Method using Bayesian Approach to Enhance Raw Data Accuracy of Position and Distance

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

Measurements for Connected Vehicles. In 2021 IFIP/IEEE International Symposium on Integrated Network Management (IM) (pp. 1018-1023). IEEE.

Mueck, J. (2002). Estimation methods for the state of traffic at traffic signals using detectors near the stop-line. Traffic Engineering and Control, 43(11), 429.

Murphy, R. R. (1998). Sensor and information fusion improved vision-based vehicle guidance. IEEE Intelligent Systems and their Applications, 13(6), 49-56.

Naimi, A. I., & Balzer, L. B. (2018). Stacked generalization: an introduction to super learning. European journal of epidemiology, 33(5), 459-464.

Okutani, I. (1987). The Kalman filtering approaches in some transportation and traffic problems. Transportation and traffic theory.

Pei, S. C., & Lin-gwo, L. (1998). Vehicle-type motion estimation by the fusion of image point and line features. Pattern Recognition, 31(3), 333-344.

Perez, M. A., Sudweeks, J. D., Sears, E., Antin, J., Lee, S., Hankey, J. M., & Dingus, T. A. (2017). Performance of basic kinematic thresholds in the identification of crash and near-crash events within naturalistic driving data. Accident Analysis & Prevention, 103, 10-19.

Sahu, S. K., & Mardia, K. V. (2005). A Bayesian kriged Kalman model for short-term forecasting of air pollution levels. Journal of the Royal Statistical Society: Series C (Applied Statistics), 54(1), 223-244.

Sau, J., El Faouzi, N. E., Ben Aissa, A., & de Mouzon, O. (2007). Particle filter-based real-time estimation and prediction of traffic conditions. In Recent advances in stochastic modeling and data analysis (pp. 406-413).

Schofer, J. L., Koppelman, F. S., Bhandari, N., Sethi, V., & Ivan, J. N. (1994). Calibration of Data Fusion Algorithm Parameters with Simulated Data.

Simon, C., Weber, P., & Evsukoff, A. (2008). Bayesian networks inference algorithm to implement Dempster Shafer theory in reliability analysis. Reliability Engineering & System Safety, 93(7), 950-963.

Simon, C. (2000). Travel time estimation by evidential data fusion. Recherche-Transports-Securite (French), (68).

Singh, P. (2021). Data Mining Techniques and Its Application in Civil Engineering—A Review. Advances in Interdisciplinary Research in Engineering and Business Management, 175-183.

Sohn, S. Y., & Shin, H. W. (2001). Data mining for road traffic accident type classification. Ergonomics, 44, 107-117.

Sohn, S. Y., & Lee, S. H. (2003). Data fusion, ensemble and clustering to improve the classification accuracy for the severity of road traffic accidents in Korea. Safety Science, 41(1), 1-14.

Wang, Y., & Papageorgiou, M. (2005). Real-time freeway traffic state estimation based on extended Kalman filter: a general approach. Transportation Research Part B: Methodological, 39(2), 141-167.

Wang, C. W. (2008). New ensemble machine learning method for classification and prediction on gene expression data. In Encyclopedia of Healthcare Information Systems (pp. 982-989). IGI Global.

Xiao, Q., Zhao, Y., & Huan, W. (2019). Multi-sensor data fusion for sign language recognition based on dynamic Bayesian network and convolutional neural network. Multimedia Tools and Applications, 78(11), 15335-15352.

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

Yang, J. B., & Xu, D. L. (2014). A study on generalising Bayesian inference to evidential reasoning. In International conference on belief functions (pp. 180-189). Springer, Cham.

Yu, Q., Teixeira, A. P., Liu, K., Rong, H., & Soares, C. G. (2021). An integrated dynamic ship risk model based on Bayesian Networks and Evidential Reasoning. Reliability Engineering & System Safety, 107993.

Zhang, S., Du, B., & Du, N. (2013). Mer-Gesh: A New Data Fusion Framework to Estimate Dynamic Road Travel Time. In Geo-Informatics in Resource Management and Sustainable Ecosystem (pp. 1-15). Springer, Berlin, Heidelberg.

Dingus, Thomas A.; Hankey, Jonathan M.; Antin, Jonathan F.; Lee, Suzanne E.; Eichelberger, Lisa; Stulce, Kelly; McGraw, Doug; Perez, Miguel and Stowe, Loren, 2015. SHRP2 REPORT S2-S06-RW-1: Naturalistic Driving Study: Technical Coordination and Quality Control, Transportation Research Board of the National Academies, Washington, D.C.

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

Appendix

1. Back-office API documentation

Back-Office_API_Specification.pdf

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

/static/swagger.json Explore

iDREAMS Back office API Specification

2.0.0 OAS3

/static/swagger.json

This is a specification for the iDREAMS backoffice web api.

Overview:

- You need a valid jwt-token to send requests to the API endpoints except the authentication endpoints. You can get your jwt-token by sending request with your credentials to authentication endpoints.
- The Backoffice API will only respond to secured communication done over HTTPS. HTTP requests will be sent a `301` redirect to corresponding HTTPS resources.
- Response to every request is sent in JSON format. In case the API request results in an error, it is represented by an `"error": {}` key in the JSON response.
- The request method (verb) determines the nature of action you intend to perform. A request made using the GET method implies that you want to fetch something from Backoffice, and POST implies you want to save something new to Backoffice.
- The API calls will respond with appropriate HTTP status codes for all requests. A 200 indicates all went well, while 4XX or 5XX response codes indicate an error from the requesting client or our API servers respectively.

Terms of service

Contact the developer

Apache 2.0

Find out more about Project iDREAMS

Servers

https://api.idreams.tum-tse.eu/

Authorize

UploadData Endpoints for uploading data to backoffice

POST `/data/dataprocessor/` Upload data from dataprocessor

Parameters Try it out

No parameters

Request body required application/json

Upload data from data proecssor

Example Value Schema


```
{
  "_id": "string",
  "uuid": "string",
  "trip_start": "2021-12-30T11:44:44.114Z",
  "trip_end": "2021-12-30T11:44:44.114Z",
  "distance": 0,
  "duration": 0,
  "dtypes": "string",
  "vehicle_uuid": "string",
  "project_uuid": "string",
  "trip_data_collected": true,
  "trip_score_collected": true,
  "trip_video_collected": true,
  "safety": {
    "health": {},
    "vehicle_control": {},
    "road_sharing": {},
    "speed_management": {},
    "safety_devices": {}
  },
  "received_at": "2021-12-30T11:44:44.114Z"
}
```


Responses

Code	Description	Links
200	successful operation Media type **application/json** Controls `Accept` header. **Example Value** Schema `{ "message": "string" }`	*No links*

Code	Description	Links
500	Upload failed Media type **application/json** **Example Value** Schema `{ "message": "string" }`	*No links*

POST `/data/simulation/` Upload simulation data. Requires read_write access

Parameters

Try it out

No parameters

Request body required

multipart/form-data:

Needs to be a CSV (.csv) file. The CSV File name should have this syntax- <p_id>*<DD-MM-YYYY-HH-MM-SS>*<country_code>_<mode(singular)>.csv (e.g. AfdgTgZDuY4UzsdP8cEjQ_distraction_A_01-01-2021-23-59-59_DE_car.csv). CSV file should contain the simulation results.

sim_file
`string($binary)`

Responses

Code	Description	Links

Code	Description	Links
200	successful operation Media type **application/json** Controls `Accept` header. **Example Value** Schema `{ "message": "string" }`	*No links*
500	Upload failed Media type **application/json** **Example Value** Schema `{ "message": "string" }`	*No links*

POST `/data/survey/` Upload survey data. Requires read_write access

Parameters Try it out

No parameters

Request body required **multipart/form-data:**

File containing survey results. Should be CSV file. The CSV File name should have this syntax- <p_id><*country_code*><mode(singular)>.csv (e.g. AxUTgZDuYzrJHP8cEjQ_DE_car.csv)

survey_file
`string($binary)`

Responses

Code	Description	Links
200	successful operation Media type **application/json** Controls `Accept` header. **Example Value** Schema `{ "message": "string" }`	*No links*
500	Upload failed Media type **application/json** **Example Value** Schema `{ "message": "string" }`	*No links*

Vehicles Endpoints for vehicles

GET `/vehicles/` List all vehicles

List all vehicles from the backoffice

Parameters

Try it out

No parameters

Responses

Code	Description	Links
200	successful operation Media type **application/json** Controls `Accept` header. **Example Value** Schema	*No links*


```
[
  {
    "_id": "string",
    "uuid": "string",
    "gw_uuid": "string",
    "name": "string",
    "class": "string",
    "steering": "string",
    "received_at": "2021-12-30T11:44:44.119Z",
    "owner": {
      "uuid": "string",
      "name": "string"
    }
  }
]
```


Code	Description	Links
500	Request failed due to server error Media type **application/json** **Example Value** Schema	*No links*


```
{
  "message": "string"
}
```


GET `/vehicles/{vehicle_uuid}` Find vehicle by vehicle uuid

Find vehicle by vehicle uuid from the backoffice

Parameters

Try it out

Name	Description

Name	Description
vehicle_uuid * required string *(path)*	ID of vehicle to return vehicle_uuid - ID of vehicle to return

Responses

Code	Description	Links
200	successful operation Media type **application/json** Controls `Accept` header. **Example Value** Schema	*No links*


```
{
  "_id": "string",
  "uuid": "string",
  "gw_uuid": "string",
  "name": "string",
  "class": "string",
  "steering": "string",
  "received_at": "2021-12-30T11:44:44.120Z",
  "owner": {
    "uuid": "string",
    "name": "string"
  }
}
```


GET `/vehicles/filter/class/{class_name}` Find vehicle by vehicle class

Find vehicle by vehicle class from the backoffice

Parameters

Try it out

Name	Description

Name	Description
class_name * required string *(path)*	class of vehicles to return class_name - class of vehicles to return

Responses

Code	Description	Links
200	successful operation Media type **application/json** Controls `Accept` header. **Example Value** Schema	*No links*


```
[
  {
    "_id": "string",
    "uuid": "string",
    "gw_uuid": "string",
    "name": "string",
    "class": "string",
    "steering": "string",
    "received_at": "2021-12-30T11:44:44.121Z",
    "owner": {
      "uuid": "string",
      "name": "string"
    }
  }
]
```


GET `/vehicles/filter/steering/{steering_type}` Find vehicle by vehicle steering type

Find vehicle by vehicle steering type from the backoffice

Parameters

Try it out

Name	Description

Name	Description
steering_type * required string *(path)*	steering type of vehicles to return steering_type - steering type of vehicles to re

Responses

Code	Description	Links
200	successful operation Media type **application/json** Controls `Accept` header. **Example Value** Schema	*No links*


```
[
  {
    "_id": "string",
    "uuid": "string",
    "gw_uuid": "string",
    "name": "string",
    "class": "string",
    "steering": "string",
    "received_at": "2021-12-30T11:44:44.124Z",
    "owner": {
      "uuid": "string",
      "name": "string"
    }
  }
]
```


GET `/vehicles/filter/owner/{owner_name}` Find vehicle by vehicle owner name

Find vehicle by vehicle owner name from the backoffice

Parameters

Try it out

Name	Description

Name	Description
owner_name * required string *(path)*	owner name of vehicles to return owner_name - owner name of vehicles to retu

Responses

Code	Description	Links
200	successful operation Media type **application/json** Controls `Accept` header. **Example Value** Schema	*No links*


```
[
  {
    "_id": "string",
    "uuid": "string",
    "gw_uuid": "string",
    "name": "string",
    "class": "string",
    "steering": "string",
    "received_at": "2021-12-30T11:44:44.125Z",
    "owner": {
      "uuid": "string",
      "name": "string"
    }
  }
]
```


GET `/vehicles/{vehicle_uuid}/driver/` Find driver of a vehicle by vehicle uuid

Find who is driving a vehicle by vehicle uuid from the backoffice

Parameters

Try it out

Name	Description

Name	Description
vehicle_uuid * required string *(path)*	vehicle uuid of the vehicle vehicle_uuid - vehicle uuid of the vehicle

Responses

Code	Description	Links
200	successful operation Media type **application/json** Controls `Accept` header.	*No links*

Example Value Schema


```
[
  {
    "uuid": "string",
    "idreams_id": "string",
    "name": "string",
    "gender": "string",
    "age_30": true,
    "pro_driver": true,
    "owner": {
      "uuid": "string",
      "name": "string"
    },
    "vehicles": [
      {
        "_id": "string",
        "uuid": "string",
        "gw_uuid": "string",
        "name": "string",
        "class": "string",
        "steering": "string",
        "received_at": "2021-12-30T11:44:44.126Z",
        "owner": {
          "uuid": "string",
          "name": "string"
        }
      }
    ]
  }
```


Drivers Endpoints for drivers

GET `/vehicles/{vehicle_uuid}/driver/` Find driver of a vehicle by vehicle uuid

Find who is driving a vehicle by vehicle uuid from the backoffice

Parameters

Try it out

Name	Description
vehicle_uuid * required string *(path)*	vehicle uuid of the vehicle vehicle_uuid - vehicle uuid of the vehicle

Responses

Code	Description	Links

Code	Description	Links
200	successful operation	*No links*

Media type

application/json

Controls `Accept` header.

Example Value Schema


```
[
  {
    "uuid": "string",
    "idreams_id": "string",
    "name": "string",
    "gender": "string",
    "age_30": true,
    "pro_driver": true,
    "owner": {
      "uuid": "string",
      "name": "string"
    },
    "vehicles": [
      {
        "_id": "string",
        "uuid": "string",
        "gw_uuid": "string",
        "name": "string",
        "class": "string",
        "steering": "string",
        "received_at": "2021-12-30T11:44:44.128Z",
        "owner": {
          "uuid": "string",
          "name": "string"
        }
      }
    ]
  }
```


GET `/drivers/` List all drivers

List all drivers from the backoffice

Parameters

Try it out

No parameters

Responses

Code	Description	Links

Code	Description	Links
200	successful operation Media type **application/json** Controls `Accept` header. **Example Value** Schema	*No links*


```
[
  {
    "uuid": "string",
    "idreams_id": "string",
    "name": "string",
    "gender": "string",
    "age_30": true,
    "pro_driver": true,
    "owner": {
      "uuid": "string",
      "name": "string"
    },
    "vehicles": [
      {
        "_id": "string",
        "uuid": "string",
        "gw_uuid": "string",
        "name": "string",
        "class": "string",
        "steering": "string",
        "received_at": "2021-12-30T11:44:44.129Z",
        "owner": {
          "uuid": "string",
          "name": "string"
        }
      }
    ]
  }
```


Code	Description	Links
500	Request failed due to server error Media type **application/json** **Example Value** Schema	*No links*


```
{
  "message": "string"
}
```


GET `/drivers/{driver_uuid}` Find driver by driver uuid

Find driver by driver uuid from the backoffice

Parameters Try it out

Name	Description
driver_uuid * required `string` *(path)*	ID of driver to return driver_uuid - ID of driver to return

Responses

Code	Description	Links
200	successful operation Media type **application/json** Controls `Accept` header. **Example Value** Schema	*No links*


```
{
  "uuid": "string",
  "idreams_id": "string",
  "name": "string",
  "gender": "string",
  "age_30": true,
  "pro_driver": true,
  "owner": {
    "uuid": "string",
    "name": "string"
  },
  "vehicles": [
    {
      "_id": "string",
      "uuid": "string",
      "gw_uuid": "string",
      "name": "string",
      "class": "string",
      "steering": "string",
      "received_at": "2021-12-30T11:44:44.130Z",
      "owner": {
        "uuid": "string",
        "name": "string"
      }
    }
  ]
}
```


GET `/drivers/filter/gender/{gender}` Filter drivers by gender

List all drivers from a particular gender from the backoffice

Parameters

Try it out

Name	Description
gender * required string *(path)*	gender of driver to return gender - gender of driver to return

Responses

Code	Description	Links
200	successful operation	*No links*

Media type

application/json

Controls `Accept` header.

Example Value Schema


```
[
  {
    "uuid": "string",
    "idreams_id": "string",
    "name": "string",
    "gender": "string",
    "age_30": true,
    "pro_driver": true,
    "owner": {
      "uuid": "string",
      "name": "string"
    },
    "vehicles": [
      {
        "_id": "string",
        "uuid": "string",
        "gw_uuid": "string",
        "name": "string",
        "class": "string",
        "steering": "string",
        "received_at": "2021-12-30T11:44:44.131Z",
        "owner": {
          "uuid": "string",
          "name": "string"
        }
      }
    ]
```

Code	Description	Links
500	Request failed due to server error Media type **application/json** **Example Value** Schema `{ "message": "string" }`	*No links*

GET `/drivers/filter/pro_driver/` Filter drivers by pro_driver

List all drivers from who are pro drivers from the backoffice

Parameters

Try it out

No parameters

Responses

Code	Description	Links

Code	Description	Links
200	successful operation	*No links*

Media type

application/json

Controls `Accept` header.

Example Value Schema


```
[
  {
    "uuid": "string",
    "idreams_id": "string",
    "name": "string",
    "gender": "string",
    "age_30": true,
    "pro_driver": true,
    "owner": {
      "uuid": "string",
      "name": "string"
    },
    "vehicles": [
      {
        "_id": "string",
        "uuid": "string",
        "gw_uuid": "string",
        "name": "string",
        "class": "string",
        "steering": "string",
        "received_at": "2021-12-30T11:44:44.132Z",
        "owner": {
          "uuid": "string",
          "name": "string"
        }
      }
    ]
  }
```


Code	Description	Links
500	Request failed due to server error	*No links*

Media type

application/json

Example Value Schema


```
{
  "message": "string"
}
```


GET `/drivers/filter/age_30/` Filter drivers by age_30

List all drivers from who are of age 30 or higher from the backoffice

Parameters

Try it out

No parameters

Responses

Code	Description	Links
200	successful operation	*No links*

Media type

application/json

Controls `Accept` header.

Example Value Schema


```
[
  {
    "uuid": "string",
    "idreams_id": "string",
    "name": "string",
    "gender": "string",
    "age_30": true,
    "pro_driver": true,
    "owner": {
      "uuid": "string",
      "name": "string"
    },
    "vehicles": [
      {
        "_id": "string",
        "uuid": "string",
        "gw_uuid": "string",
        "name": "string",
        "class": "string",
        "steering": "string",
        "received_at": "2021-12-30T11:44:44.133Z",
        "owner": {
          "uuid": "string",
          "name": "string"
        }
      }
    ]
  }
```

Code	Description	Links
500	Request failed due to server error Media type **application/json** **Example Value** Schema `{ "message": "string" }`	*No links*

Projects
Endpoints for Projects

Videos
Endpoints for videos

GET `/videos/` List all videos

List all video metadata from the backoffice

Parameters

Try it out

No parameters

Responses

Code	Description	Links

Code	Description	Links
200	successful operation Media type **application/json** Controls `Accept` header. **Example Value** Schema	*No links*


```
[
  {
    "trip_uuid": "string",
    "vid_start": "string",
    "vid_end": "string",
    "duration": "string",
    "url": "string",
    "events": [
      {
        "type": "string",
        "evt_start": "string",
        "evt_end": "string",
        "duration": 0,
        "delta": 0
      }
    ]
  }
]
```


Code	Description	Links
404	No video metadata found in back office Media type **application/json** **Example Value** Schema	*No links*


```
{
  "message": "string"
}
```


GET `/videos/filter/trip/{trip_uuid}` Find video metadata by trip uuid

Find video metadata by trip uuid from the backoffice

Parameters

Try it out

Name	Description

Name	Description
trip_uuid * required string *(path)*	trip ID of video metadata to return trip_uuid - trip ID of video metadata to return

Responses

Code	Description	Links
200	successful operation Media type **application/json** Controls `Accept` header. **Example Value** Schema `{}`	*No links*

GET `/videos/filter/event/{event_type}` Find video metadata by trip uuid

Find video metadata by event type from the backoffice

Parameters

Try it out

Name	Description
event_type * required string *(path)*	event type of video metadata to return event_type - event type of video metadata to

Responses

Code	Description	Links
200	successful operation Media type **application/json** Controls `Accept` header. **Example Value** Schema `{}`	*No links*

Trips Endpoints for trips of vehicles

GET `/trips/` List all trips

List all trips (with default max limit) from the backoffice

Parameters

Try it out

Name	Description
limit * required `string` *(query)*	max number of trips to fetch limit - max number of trips to fetch

Responses

Code	Description	Links

Code	Description	Links
200	successful operation	*No links*

Media type

application/json

Controls `Accept` header.

Example Value Schema


```
[
  {
    "_id": "string",
    "uuid": "string",
    "trip_start": "2021-12-30T11:44:44.139Z",
    "trip_end": "2021-12-30T11:44:44.139Z",
    "distance": 0,
    "duration": 0,
    "dtypes": "string",
    "vehicle_uuid": "string",
    "project_uuid": "string",
    "trip_data_collected": true,
    "trip_score_collected": true,
    "trip_video_collected": true,
    "safety": {
      "health": {},
      "vehicle_control": {},
      "road_sharing": {},
      "speed_management": {},
      "safety_devices": {}
    },
    "received_at": "2021-12-30T11:44:44.139Z"
  }
]
```


Code	Description	Links
404	No trip found in back office	*No links*

Media type

application/json

Example Value Schema


```
{
  "message": "string"
}
```


GET `/trips/all` List all trips

List all trips from the backoffice. Requires superadmin access.

Parameters

Try it out

No parameters

Responses

Code	Description	Links
200	successful operation	*No links*

Media type

application/json

Controls `Accept` header.

Example Value Schema


```
[
  {
    "_id": "string",
    "uuid": "string",
    "trip_start": "2021-12-30T11:44:44.140Z",
    "trip_end": "2021-12-30T11:44:44.140Z",
    "distance": 0,
    "duration": 0,
    "dtypes": "string",
    "vehicle_uuid": "string",
    "project_uuid": "string",
    "trip_data_collected": true,
    "trip_score_collected": true,
    "trip_video_collected": true,
    "safety": {
      "health": {},
      "vehicle_control": {},
      "road_sharing": {},
      "speed_management": {},
      "safety_devices": {}
    },
    "received_at": "2021-12-30T11:44:44.140Z"
  }
]
```


Code	Description	Links
404	No trip found in back office	*No links*

Media type

application/json

Example Value Schema


```
{
  "message": "string"
}
```


GET `/trips/{trip_uuid}` Find trip by trip id

Find trip by trip id from the backoffice

Parameters

Try it out

Name	Description
trip_uuid * required `string` *(path)*	ID of trip to return trip_uuid - ID of trip to return

Responses

Code	Description	Links
200	successful operation	*No links*

Media type

application/json

Controls `Accept` header.

Example Value Schema


```json
{
  "_id": "string",
  "uuid": "string",
  "trip_start": "2021-12-30T11:44:44.142Z",
  "trip_end": "2021-12-30T11:44:44.142Z",
  "distance": 0,
  "duration": 0,
  "dtypes": "string",
  "vehicle_uuid": "string",
  "project_uuid": "string",
  "trip_data_collected": true,
  "trip_score_collected": true,
  "trip_video_collected": true,
  "safety": {
    "health": {},
    "vehicle_control": {},
    "road_sharing": {},
    "speed_management": {},
    "safety_devices": {}
  },
  "received_at": "2021-12-30T11:44:44.142Z"
}
```

GET `/trips/before/{trip_uuid}` Find trip by trip id

Find trip by trip id from the backoffice

Parameters

Try it out

Name	Description
trip_uuid * required `string` *(path)*	ID of trip to return trip_uuid - ID of trip to return
limit * required `string` *(query)*	max number of trips to fetch limit - max number of trips to fetch

Responses

Code	Description	Links

Code	Description	Links
200	successful operation	*No links*

Media type

application/json

Controls `Accept` header.

Example Value Schema


```
{
  "_id": "string",
  "uuid": "string",
  "trip_start": "2021-12-30T11:44:44.143Z",
  "trip_end": "2021-12-30T11:44:44.143Z",
  "distance": 0,
  "duration": 0,
  "dtypes": "string",
  "vehicle_uuid": "string",
  "project_uuid": "string",
  "trip_data_collected": true,
  "trip_score_collected": true,
  "trip_video_collected": true,
  "safety": {
    "health": {},
    "vehicle_control": {},
    "road_sharing": {},
    "speed_management": {},
    "safety_devices": {}
  },
  "received_at": "2021-12-30T11:44:44.143Z"
}
```


GET `/trips/after/{trip_uuid}` Find trip by trip id

Find trip by trip id from the backoffice

Parameters Try it out

Name	Description
trip_uuid * required `string` *(path)*	ID of trip to return trip_uuid - ID of trip to return
limit * required `string` *(query)*	max number of trips to fetch limit - max number of trips to fetch

Responses

Code	Description	Links
200	successful operation Media type application/json Controls `Accept` header. **Example Value** Schema	*No links*


```
{
  "_id": "string",
  "uuid": "string",
  "trip_start": "2021-12-30T11:44:44.144Z",
  "trip_end": "2021-12-30T11:44:44.144Z",
  "distance": 0,
  "duration": 0,
  "dtypes": "string",
  "vehicle_uuid": "string",
  "project_uuid": "string",
  "trip_data_collected": true,
  "trip_score_collected": true,
  "trip_video_collected": true,
  "safety": {
    "health": {},
    "vehicle_control": {},
    "road_sharing": {},
    "speed_management": {},
    "safety_devices": {}
  },
  "received_at": "2021-12-30T11:44:44.144Z"
}
```


POST `/trips/filter/daterange/` List all trips with specified time range

List all trips from the backoffice for the given time range

Parameters

Try it out

Name	Description
limit `string` *(query)*	max number of trips to fetch limit - max number of trips to fetch

Request body required

application/json

Optional description in *Markdown*

Example Value Schema


```
{
  "startdate": "2021-04-27T09:07:00.488Z",
  "enddate": "2021-04-27T09:07:00.488Z"
}
```


Responses

Code	Description	Links
200	successful operation	*No links*

Media type

application/json

Controls `Accept` header.

Example Value Schema


```
[
  {
    "_id": "string",
    "uuid": "string",
    "trip_start": "2021-12-30T11:44:44.146Z",
    "trip_end": "2021-12-30T11:44:44.146Z",
    "distance": 0,
    "duration": 0,
    "dtypes": "string",
    "vehicle_uuid": "string",
    "project_uuid": "string",
    "trip_data_collected": true,
    "trip_score_collected": true,
    "trip_video_collected": true,
    "safety": {
      "health": {},
      "vehicle_control": {},
      "road_sharing": {},
      "speed_management": {},
      "safety_devices": {}
    },
    "received_at": "2021-12-30T11:44:44.146Z"
  }
]
```

Code	Description	Links
404	No trip found in back office Media type **application/json** **Example Value** Schema `{ "message": "string" }`	*No links*

GET `/trips/filter/vehicle/{vehicle_uuid}` List all trips by vehicle uuid

List all trips by vehicle uuid from the backoffice

Parameters

Try it out

Name	Description
vehicle_uuid * required `string` *(path)*	ID of vehicle associated with trips vehicle_uuid - ID of vehicle associated with tr

Responses

Code	Description	Links

Code	Description	Links
200	successful operation Media type **application/json** Controls `Accept` header. **Example Value** Schema	*No links*


```
[
  {
    "_id": "string",
    "uuid": "string",
    "trip_start": "2021-12-30T11:44:44.148Z",
    "trip_end": "2021-12-30T11:44:44.148Z",
    "distance": 0,
    "duration": 0,
    "dtypes": "string",
    "vehicle_uuid": "string",
    "project_uuid": "string",
    "trip_data_collected": true,
    "trip_score_collected": true,
    "trip_video_collected": true,
    "safety": {
      "health": {},
      "vehicle_control": {},
      "road_sharing": {},
      "speed_management": {},
      "safety_devices": {}
    },
    "received_at": "2021-12-30T11:44:44.148Z"
  }
]
```


Code	Description	Links
404	No trip found in back office Media type **application/json** **Example Value** Schema	*No links*


```
{
  "message": "string"
}
```


Trips Data

Endpoints for session of trips

GET `/trips/{tripid}/data` Find trip data by trip id

Find trip data by trip id from the backoffice

Parameters

Try it out

Name	Description
tripid * required `string` *(path)*	ID of trip associate to session tripid - ID of trip associate to session

Responses

Code	Description	Links
200	successful operation Media type **application/json** Controls `Accept` header. **Example Value** Schema `{}`	*No links*
404	No session found for the trip in back office Media type **application/json** **Example Value** Schema `{ "message": "string" }`	*No links*

GET `/trips/{trip_uuid}/data/{dtype}` Find specific trip data dtype by trip id

Find dtype from a trip data by trip id from the backoffice

Parameters

Try it out

Name	Description
trip_uuid * required string *(path)*	ID of trip associate to trip data trip_uuid - ID of trip associate to trip data
dtype * required string *(path)*	dtype from trip data dtype - dtype from trip data

Responses

Code	Description	Links
200	successful operation Media type application/json Controls `Accept` header. Example Value Schema `{}`	*No links*
404	No session found for the trip in back office Media type application/json Example Value Schema `{ "message": "string" }`	*No links*

Simulations

Endpoints for simulation data

POST `/data/simulation/` Upload simulation data. Requires read_write access

Parameters

Try it out

No parameters

Request body required

multipart/form-data:

Needs to be a CSV (.csv) file. The CSV File name should have this syntax- <p_id>*<DD-MM-YYYY-HH-MM-SS>*<country_code>_<mode(singular)>.csv (e.g. AfdgTgZDuY4UzsdP8cEjQ_distraction_A_01-01-2021-23-59-59_DE_car.csv). CSV file should contain the simulation results.

sim_file
`string($binary)`

Responses

Code	Description	Links
200	successful operation Media type application/json Controls `Accept` header. **Example Value** Schema	*No links*


```
{
  "message": "string"
}
```

Code	Description	Links
500	Upload failed Media type **application/json** **Example Value** Schema `{ "message": "string" }`	*No links*

Surveys Endpoints survey data management

POST `/data/survey/` Upload survey data. Requires read_write access

Parameters

Try it out

No parameters

Request body required

multipart/form-data:

File containing survey results. Should be CSV file. The CSV File name should have this syntax- <p_id><*country_code*><mode(singular)>.csv (e.g. AxUTgZDuYzrJHP8cEjQ_DE_car.csv)

survey_file
`string($binary)`

Responses

Code	Description	Links

Code	Description	Links
200	successful operation Media type **application/json** Controls `Accept` header. **Example Value** Schema `{ "message": "string" }`	*No links*
500	Upload failed Media type **application/json** **Example Value** Schema `{ "message": "string" }`	*No links*

ApiUsers Endpoints for managing users

PUT `/auth/user` Create user

Creating user for api usage with different right. Required superadmin access.

Parameters

Try it out

No parameters

Request body required

application/json

User object to register

Example Value Schema

```
{
  "_id": "string",
```

```
  "username": "string",
  "password": "string",
  "who": "string",
  "grants": "string"
}
```


Responses

Code	Description	Links
201	User created Media type **application/json** Controls `Accept` header. **Example Value** Schema `{}`	*No links*
500	Server error Media type **application/json** **Example Value** Schema `{ "message": "string" }`	*No links*
default	successful operation Media type	*No links*

POST `/auth/login` Logs user into the system

Parameters Try it out

No parameters

Request body required

application/json

Login credentials: username and password

Example Value Schema

```
{
  "username": "string",
  "password": "string"
}
```

Responses

Code	Description	Links
200	successful operation Media type application/json Controls `Accept` header. **Example Value** Schema `{}`	*No links*
400	Invalid username/password supplied Media type application/json **Example Value** Schema `{}`	*No links*

Sync

GET `/sync/` Request to see sync status.

Parameters Try it out

No parameters

Responses

Code	Description	Links
200	successful operation Media type **application/json** Controls `Accept` header. **Example Value** Schema	*No links*


```
{
  "message": "string",
  "status": "string",
  "requested_items": [
    "string"
  ],
  "last_sync_attempted_at": "string",
  "last_sync_finished_at": "string"
}
```


GET `/sync/trigger` Request to sync data from external servers. Requires read_write access

Parameters

Try it out

No parameters

Responses

Code	Description	Links

Code	Description	Links
200	successful operation	*No links*

Media type

application/json

Controls `Accept` header.

Example Value Schema


```
{
  "message": "string",
  "status": "string",
  "requested_items": [
    "string"
  ],
  "last_sync_attempted_at": "string",
  "last_sync_finished_at": "string"
}
```


GET `/sync/trigger/drivers` Request to sync drivers from external servers. Requires read_write access

Parameters

Try it out

No parameters

Responses

Code	Description	Links
200	successful operation	*No links*

Media type

application/json

Controls `Accept` header.

Example Value Schema


```
{
  "message": "string",
  "status": "string",
  "requested_items": [
    "string"
  ],
  "last_sync_attempted_at": "string",
  "last_sync_finished_at": "string"
}
```

GET /sync/trigger/vehicles Request to sync vehicles from external servers. Requires read_write access

Parameters

Try it out

No parameters

Responses

Code	Description	Links
200	successful operation	*No links*

Media type

application/json

Controls `Accept` header.

Example Value Schema


```
{
  "message": "string",
  "status": "string",
  "requested_items": [
    "string"
  ],
  "last_sync_attempted_at": "string",
  "last_sync_finished_at": "string"
}
```


GET /sync/trigger/trips Request to sync trips from external servers. Requires read_write access

Parameters

Try it out

No parameters

Responses

Code	Description	Links

Code	Description	Links
200	successful operation Media type **application/json** Controls `Accept` header. **Example Value** Schema	*No links*


```
{
  "message": "string",
  "status": "string",
  "requested_items": [
    "string"
  ],
  "last_sync_attempted_at": "string",
  "last_sync_finished_at": "string"
}
```


Schemas

```
Vehicle {
    _id                     string
    uuid                    string
    gw_uuid                 string
    name                    string
    class                   string
    steering                string
    received_at             string($date-time)
    owner                       {...}
}
```

```
Trip {
    _id                     string
    uuid                    string
    trip_start              string($date-time)
    trip_end                string($date-time)
    distance                number
    duration                number
    dtypes                  string
    vehicle_uuid            string
    project_uuid            string
    trip_data_collected     boolean
    trip_score_collected    boolean
    trip_video_collected    boolean
    safety                  Safety {...}
    received_at             string($date-time)
}
```

Safety {

```
    health                          {...}
    vehicle_control                 {...}
    road_sharing                    {...}
    speed_management                {...}
    safety_devices                  {...}
}
```

Tripdata {

```
    _id                 string
    data                            [...]
    trip_uuid           string
    nb_drivers          number
    config                          {...}
    oseven                          {...}
}
```

Driver {

```
    uuid                string
    idreams_id          string
    name                string
    gender              string
    age_30              boolean
    pro_driver          boolean
    owner                           {...}
    vehicles                        [...]
}
```

VideoMetadata {

```
    trip_uuid           string
    vid_start           string
    vid_end             string
    duration            string
    url                 string
    events                          [...]
}
```

VideoEvent {

```
    type                string
    evt_start           string
    evt_end             string
    duration            number
    delta               number
}
```

PIF

ApiUser

```
ApiUser {
    _id          string
    username     string
    password     string
    who          string
    grants       string
}
```

TokenResponse

```
TokenResponse {
    access-token     string
    refresh-token    string
}
```

ApiResponse

```
ApiResponse {
    type         string
    message      string
    content          {...}
}
```

Ts

```
Ts [string($date-time)]
```

Lat

```
Lat [number]
```

Lon

```
Lon [number]
```

SPD

```
SPD [number]
```

Diagnostic

```
Diagnostic [string]
```

Value

```
Value [number]
```

Rule `[string]`

Distance `[number]`

Duration `[number]`

HW_measurement `[number]`

Geotab_GPS `{`

`ts`	**Ts** `[...]`
`lat`	**Lat** `[...]`
`lon`	**Lon** `[...]`
`spd`	**SPD** `[...]`

`}`

Geotab_Data `{`

`ts`	**Ts** `[...]`
`diagnostic`	**Diagnostic** `[...]`
`value`	**Value** `[...]`

`}`

Geotab_Exceptions `{`

`ts`	**Ts** `[...]`
`diagnostic`	**Diagnostic** `[...]`
`rule`	**Rule** `[...]`
`distance`	**Rule** `[...]`
`duration`	**Rule** `[...]`

`}`

```
ME_FCW_Map {
    ts                  Ts [...]
    lat                 Lat [...]
    lon                 Lon [...]
}

ME_HMW_Map {
    ts                  Ts [...]
    lat                 Lat [...]
    lon                 Lon [...]
}

Geotab_ME_HW {
    ts                  Ts [...]
    hw_measurement      HW_measurement [...]
}
```