# DELIVERABLE

## D5.1 – Connected Vehicles Pilots Design

<table>
<thead>
<tr>
<th><strong>Project Acronym</strong></th>
<th>TT</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Project Title</strong></td>
<td>Transforming Transport</td>
</tr>
<tr>
<td><strong>Grant Agreement number</strong></td>
<td>731932</td>
</tr>
<tr>
<td><strong>Call and topic identifier</strong></td>
<td>ICT-15-2016-2017</td>
</tr>
<tr>
<td><strong>Funding Scheme</strong></td>
<td>Innovation Action (IA)</td>
</tr>
<tr>
<td><strong>Project duration</strong></td>
<td>30 Months [1 January 2017 – 30 June 2019]</td>
</tr>
<tr>
<td><strong>Coordinator</strong></td>
<td>Mr. Rodrigo Castiñeira (INDRA)</td>
</tr>
<tr>
<td><strong>Website</strong></td>
<td><a href="http://www.transformingtransport.eu">www.transformingtransport.eu</a></td>
</tr>
<tr>
<td><strong>Project Acronym</strong></td>
<td>TT</td>
</tr>
</tbody>
</table>

Funded by the European Union’s H2020 GA - 731932
Document fiche

Authors: Pablo Campillo (Answare), David Cobo (Answare), Michael Schygulla (PTV), Contributors

Internal reviewers: Mauro Gil (Indra), Rodrigo Castiñeira (Indra)

Work Package: WP5

Task: T5.1

Nature: R

Dissemination: PU

Document History

<table>
<thead>
<tr>
<th>Version</th>
<th>Date</th>
<th>Contributor(s)</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.1</td>
<td>08/03/2017</td>
<td>Answare, SoFLEET, Autoaid</td>
<td>Connected Vehicles Pilot Design</td>
</tr>
<tr>
<td>0.2</td>
<td>08/03/2017</td>
<td>Answare, SoFLEET, Autoaid</td>
<td>Connected Car Pilot Design</td>
</tr>
<tr>
<td>0.3</td>
<td>15/03/2017</td>
<td>PTV, Fraunhofer IGD</td>
<td>Added contribution for Connected Trucks</td>
</tr>
<tr>
<td>0.4</td>
<td>16/03/2017</td>
<td>TomTom</td>
<td>TomTom contribution added</td>
</tr>
<tr>
<td>0.5</td>
<td>23/03/2017</td>
<td>PTV</td>
<td>New contributions for Connected Trucks</td>
</tr>
<tr>
<td>0.6</td>
<td>28/03/2017</td>
<td>PTV, Answare</td>
<td>Inclusion of modifications after review</td>
</tr>
<tr>
<td>0.7</td>
<td>29/03/2017</td>
<td>Answare</td>
<td>Internal Review</td>
</tr>
<tr>
<td>0.8</td>
<td>30/03/2017</td>
<td>PTV</td>
<td>Created subchapters in section 3.5 to be referenced in 3.6</td>
</tr>
<tr>
<td>1.0</td>
<td>31/03/2017</td>
<td>PTV</td>
<td>Final document for submission to the EC</td>
</tr>
</tbody>
</table>
## Keywords:
Traffic congestions, Accurate travel time, Added value services, Traffic accidents identification, CO2 reduction, Predictive Maintenance, Big Data, Fleet Management, Data Science.

## Abstract (few lines):
This deliverable (D5.1) is outcome of WP5/T5.1 “Pilot Coordination & Experimental Infrastructure” with respect to the use cases, scenarios and the design of the adequate infrastructure and roadmap to enable the evaluation of the “Sustainable connected cars” and “Sustainable connected trucks” pilots in the frame of the Transforming Transport project.

## DISCLAIMER
This document does not represent the opinion of the European Community, and the European Community is not responsible for any use that might be made of its content. This document may contain material, which is the copyright of certain TT consortium parties, and may not be reproduced or copied without permission. All TT consortium parties have agreed to full publication of this document. The commercial use of any information contained in this document may require a license from the proprietor of that information.

Neither the TT consortium as a whole, nor a certain party of the TT consortium warrant that the information contained in this document is capable of use, nor that use of the information is free from risk, and does not accept any liability for loss or damage suffered by any person using this information.

## ACKNOWLEDGEMENT
This document is a deliverable of TT project. This project has received funding from the European Union’s Horizon 2020 research and innovation programme under grant agreement Nº 731932
# Table of Contents

1. **DEFINITIONS, ACRONYMS AND ABBREVIATIONS** ........................................................................................................ 8

2. **EXECUTIVE SUMMARY** ........................................................................................................................................ 10

3. **MOTIVATION AND AMBITION** .............................................................................................................................. 11

4. **DESIGN OF INITIAL PILOT: SUSTAINABLE CONNECTED CARS** ................................................................................. 13

   4.1. **REQUIREMENTS** .................................................................................................................................................. 14

   4.2. **OBJECTIVES** ......................................................................................................................................................... 14

   4.3. **USE CASES AND SCENARIOS** .......................................................................................................................... 17

      4.3.1. **Use cases for stage 1** ...................................................................................................................................... 18

      4.3.2. **Use cases for stage 2** ...................................................................................................................................... 20

      4.3.3. **Use cases for stage 3** ...................................................................................................................................... 23

   4.4. **DATA ASSETS** ...................................................................................................................................................... 24

   4.5. **BIG DATA TECHNOLOGY, TECHNIQUES AND ALGORITHMS** .............................................................................. 25

      4.5.1. **Breakdown estimator system** .......................................................................................................................... 27

      4.5.2. **Traffic jam detector system** ........................................................................................................................... 27

      4.5.3. **Reducing emission system** ........................................................................................................................... 28

   4.6. **POSITIONING OF PILOT SOLUTIONS IN BDVA REFERENCE MODEL** ................................................................. 29

   4.7. **BIG DATA INFRASTRUCTURE** ............................................................................................................................ 30

      4.7.1. **Peculiarities of stage1** ...................................................................................................................................... 31

      4.7.2. **Peculiarities of stage2** ...................................................................................................................................... 31

      4.7.3. **Peculiarities of stage3** ...................................................................................................................................... 32

   4.8. **ROADMAP** ......................................................................................................................................................... 33

5. **DESIGN OF REPLICATION PILOT: CONNECTED TRUCKS** ............................................................................................ 34

   5.1. **REQUIREMENTS** .................................................................................................................................................. 34

   5.2. **OBJECTIVES** ......................................................................................................................................................... 34

   5.3. **USE CASES AND SCENARIOS** .......................................................................................................................... 35

   5.4. **DATA ASSETS** ...................................................................................................................................................... 37

   5.5. **BIG DATA TECHNOLOGY, TECHNIQUES AND ALGORITHMS** .............................................................................. 39

      5.5.1. **Big Data technologies to be used** .................................................................................................................... 39

      5.5.2. **Big Data techniques and processing** ............................................................................................................... 40

      5.5.3. **Analyses and foreseen results** ........................................................................................................................ 41

   5.6. **POSITIONING OF PILOT SOLUTIONS IN BDVA REFERENCE MODEL** ................................................................. 42

   5.7. **BIG DATA INFRASTRUCTURE** ............................................................................................................................ 43

   5.8. **ROADMAP** ......................................................................................................................................................... 44

6. **COMMONALITIES AND REPLICATION** ..................................................................................................................... 45

   6.1. **COMMON REQUIREMENTS AND ASPECTS** ....................................................................................................... 45
Table of Figures

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Figure 1</td>
<td>A company fleet of vehicles.</td>
<td>13</td>
</tr>
<tr>
<td>Figure 2</td>
<td>Initial pilot area.</td>
<td>18</td>
</tr>
<tr>
<td>Figure 3</td>
<td>Use case diagram for stage 1.</td>
<td>19</td>
</tr>
<tr>
<td>Figure 4</td>
<td>Use case for predictive maintenance system.</td>
<td>21</td>
</tr>
<tr>
<td>Figure 5</td>
<td>Use case for Emission Reduction system.</td>
<td>22</td>
</tr>
<tr>
<td>Figure 6</td>
<td>Use case for Traffic Jam Identification System.</td>
<td>23</td>
</tr>
<tr>
<td>Figure 7</td>
<td>OBD II Dongle components.</td>
<td>24</td>
</tr>
<tr>
<td>Figure 8</td>
<td>System Architecture Diagram for Pilot Connected Cars.</td>
<td>26</td>
</tr>
<tr>
<td>Figure 9</td>
<td>Positioning of Cars pilot solutions in BDVA Reference Model.</td>
<td>29</td>
</tr>
<tr>
<td>Figure 10</td>
<td>Heat map based on FCD for European road traffic network.</td>
<td>38</td>
</tr>
<tr>
<td>Figure 11</td>
<td>Positioning of Trucks pilot solutions in BDVA Reference Model.</td>
<td>42</td>
</tr>
<tr>
<td>Figure 12</td>
<td>Big Data infrastructure approach for Connected Trucks Pilot.</td>
<td>43</td>
</tr>
</tbody>
</table>
Table of Tables

Table 1: Data assets to be used in the Connected Cars Pilot ................................................................. 25
Table 2: Roadmap Table for Connected Cars Pilot .................................................................................. 33
Table 3: Data assets to be used in the Connected Trucks Pilot ............................................................... 39
Table 4: Roadmap Table for Connected Trucks Pilot ............................................................................. 44
## Definitions, Acronyms and Abbreviations

<table>
<thead>
<tr>
<th>Acronym</th>
<th>Title</th>
</tr>
</thead>
<tbody>
<tr>
<td>API</td>
<td>Application Program Interface</td>
</tr>
<tr>
<td>CO2</td>
<td>Carbon Dioxide</td>
</tr>
<tr>
<td>CEL</td>
<td>Check Engine Light</td>
</tr>
<tr>
<td>Cx_O_y</td>
<td>Objectives (where x is ‘C’ for cars and ‘T’ for tracks and y is the objective identification number, e.g. CC_O_1.2)</td>
</tr>
<tr>
<td>DATEX</td>
<td>Data Exchange Service</td>
</tr>
<tr>
<td>DTC</td>
<td>Diagnostic Trouble Code</td>
</tr>
<tr>
<td>Dx</td>
<td>Deliverable (where x defines the deliverable identification number e.g. D1.1)</td>
</tr>
<tr>
<td>ECU</td>
<td>Engine Control Unit</td>
</tr>
<tr>
<td>EU</td>
<td>European Union</td>
</tr>
<tr>
<td>ETA</td>
<td>Estimated Time of Arrival</td>
</tr>
<tr>
<td>FCD</td>
<td>Floating Car Data</td>
</tr>
<tr>
<td>GPS</td>
<td>Global Positioning System</td>
</tr>
<tr>
<td>HTTP</td>
<td>Hypertext Transfer Protocol</td>
</tr>
<tr>
<td>JSON</td>
<td>JavaScript Object Notation</td>
</tr>
<tr>
<td>KPI</td>
<td>Key Performance Indicator</td>
</tr>
<tr>
<td>ML</td>
<td>Machine Learning</td>
</tr>
<tr>
<td>MSx</td>
<td>project Milestone (where x defines a project milestone e.g. MS3)</td>
</tr>
<tr>
<td>Mx</td>
<td>Month (where x defines a project month, e.g. M10)</td>
</tr>
<tr>
<td>O</td>
<td>Other</td>
</tr>
<tr>
<td>OBD</td>
<td>On Board Diagnosis</td>
</tr>
<tr>
<td>P</td>
<td>Prototype</td>
</tr>
<tr>
<td>PID</td>
<td>Parameter ID</td>
</tr>
<tr>
<td>PU</td>
<td>Public</td>
</tr>
<tr>
<td>Qx</td>
<td>One-fourth of a year (x identifies the quarter, e.g Q2 is April)</td>
</tr>
<tr>
<td>R</td>
<td>Report</td>
</tr>
<tr>
<td>REST</td>
<td>Representational State Transfer</td>
</tr>
<tr>
<td>Rx</td>
<td>Requirement (where x defines the identification number for the specific requirement, e.g. R2)</td>
</tr>
<tr>
<td>RAM</td>
<td>Random Access Memory</td>
</tr>
<tr>
<td>RPM</td>
<td>Revolutions Per Minute</td>
</tr>
<tr>
<td>-----</td>
<td>------------------------</td>
</tr>
<tr>
<td>TT</td>
<td>Transforming Transport</td>
</tr>
<tr>
<td>UML</td>
<td>Unified Modeling Language</td>
</tr>
<tr>
<td>VIN</td>
<td>Vehicle Identification Number</td>
</tr>
<tr>
<td>WP</td>
<td>Work Package</td>
</tr>
</tbody>
</table>
Executive Summary

This document constitutes the deliverable 5.1 (D5.1) in the frame of the TT project. It covers the designs for the sustainable connected vehicles pilots: Sustainable Connected Cars and Sustainable Connected Trucks. The document specifies the requirements, objectives and scenarios for both pilots and describes the data assets, technology and infrastructure that are going to be used in the Big Data Platform according to the guidelines stated in the deliverable D2.1 Pilot Coordination Handbook.

The aim of the pilots as a whole is to show that Big Data analysis can be used to optimize mobility in terms of emissions, safety, maintenance and traffic routing in their respective sectors.

Big data real-time analytics can greatly contribute to make road transport more sustainable and safe.
1 Motivation and Ambition

Road transport is the second biggest source of greenhouse gas emissions in the EU, after power generation. It contributes about one-fifth of the EU’s total emissions of carbon dioxide (CO2). Road transport is one of the few sectors where emissions have been rising rapidly over the last 20 years. In the period 1990 to 2010 emissions from road transport increased by 22.6%. This increase acted as a brake on the EU’s progress in cutting overall emissions of greenhouse gases, which fell by 15.4%. Passenger cars alone are responsible for around 12% of EU CO2 emissions. From the point of view of security, it also represents a huge challenge to the EU since every year 26,000 people die in the EU in car accidents.

Big data real-time analytics can greatly contribute to make road transport more sustainable and safe and this pilot will demonstrate how this can be possible and which benefits Big Data can bring to these sectors. Specifically, the pilot will focus on:

- Test **high added value services** powered by large scale aggregate analytics such as predictive maintenance, traffic accidents identification, empowering efficient driving and CO2 emissions reduction.
- Optimize the **management of vehicle fleets** through continuous monitoring, vehicle dataset analysis and decision support systems.
- Increase **efficiency and competitiveness** by optimally defining routes based on predictive analysis and mobility patterns.

First pilot is called Sustainable Connected Cars and will be carried out by SoFLEET, Autoaid and Answare. This pilot is focused on cars that belong to different sort of companies. These companies are interested on achieving an efficient management of their fleets. Thanks to devices called OBD dongles installed on the cars and a Big Data architecture, the partners will study techniques and algorithms to offer a decision support system to achieve preventive maintenance of a fleet, monitoring and promote eco-friendly driver behaviours and identification of traffic jams. The dongles gather valuable information from the vehicle such as location, speed, accelerations and fault codes (DTCs). The Big Data is able to deal with this huge amount of information and provide functionalities such as data collection, predictive analysis and visualization.

Second pilot is called Sustainable Connected Trucks and will be carried out by PTV, JDR, TomTom and Fraunhofer IGD. The main objective for this pilot is the enhancement of planning and optimization systems for fleet managers. In order to achieve this goal, it will be necessary the
assessment of traffic flow for trucks journeys and on the detection and analysis of logistic hotspots such as terminals, toll stations and ferry stations. Large amounts of Big Data processing, specifically related to truck fleets all over Europe, will be necessary for this task. Additionally, the use of satellite images as an extra data source on different planning stages in the context of applications for truck fleet managers will be incorporated in this pilot to detect not only the current state of a location but also its changes over time.
2 Design of Initial Pilot: Sustainable Connected Cars

Three partners are involved in this pilot: SoFLEET, Autoaid and Answare.

SoFLEET provides a plug & play solution to connect vehicles and manage company fleets of vehicles (see Figure 1). The solution includes an OBD dongle that work on every vehicle from 2006, a dashboard to analyse eco-driving and other indicators such as fuel consumption, accidents, maintenance book and geolocation, and a mobile application designed for drivers. This application allows our customers (companies that have fleets of vehicles) to reward their drivers according to their eco-driving and the attention they pay to a vehicle that does not belong to them.

![Figure 1: A company fleet of vehicles.](image)

In addition to eco-driving indicators, present fault codes in electronic control units and driver feedback is used to apply breakdown predictions. Fault codes interpretation is hard to achieve and it differs between vehicle brands and models. However, Autoaid has a comprehensive diagnostic data base of these fault codes where they are interpreted and classified by severity. Since 2008, they have been improving it using machine learning algorithms and professional mechanics.

We expect to reduce break downs because of better predictions. The earlier a problem can be detected and a part will be changed before the car breaks down, the less down time will occur within a fleet. The fleets will have less down time as the cars will be better maintained. Costs of
maintenance are less than down time, where costs for repair, staff that is not able to work, replacement car and business damages like late deliveries can involve much higher costs.

Answare is a company with experience in decision support systems that will set up the Big Data infrastructure to collect data, analyze them and improve services already offered by the others partners at a higher scale. Besides, other new services will be addressed such as identification of traffic jams that will allow drivers to be informed about road status and help to decide the best route. In the future, this information would be included in a navigation system as part of the smartphone application of car drivers.

The connected car pilot is a way to develop a new business model for SoFLEET, based on data monitoring we collect from our customers. The benefit for its customers is to have a more accurate diagnostic of breakdowns and potentially a free cost solution. Autoaid aims at improving their diagnostic systems. Answare looks forward to acquire experience in this new domain in order to offer domain-specific solution as services to these companies and/or others. For instance, insurance companies and other maintenance companies such as car workshops have informed about their interest in this kind of solutions.

2.1 Requirements

Within the Connected Cars Pilot the following main requirements have been identified:

- **R1:** Enable car drivers or fleet managers to be notified about needed maintenance tasks of their vehicles. It will decrease breakdowns on the road by better maintenance. So, costs associated with this issue will be reduced.
- **R2:** Enable car drivers or fleet managers to be informed about tasks needed for reducing emissions and, therefore, save fuel consumption.
- **R3:** Enable car drivers to identify traffic congestions in order to avoid them. Car drivers may take this information into account to evaluate some alternate routes. It should avoid journey delays and, therefore, save time and reduce emissions.

2.2 Objectives

To address the requirements, the following solutions are defined:

1. **Breakdown estimator system:** It is the system in charge of predicting breakdowns allowing fleet managers to do preventive maintenance. Predictions are based on fault codes interpreted by Autoaid, sensor data values from cars provided by SoFLEET and
current breakdowns notified by the drivers' smartphone application. The system provides a maintenance report as output that may be updated daily, if any estimation has changed.

2. **Emission reduction system**: This system addresses emission reductions by two ways. On one side, it is based on drivers' behaviour: if it is aggressive or the car average engine RPM is high, the driving is not eco-friendly. On the other hand, a decrease of fuel consumption performance can be caused by other reasons, for example, low tire pressure or poor engine working. This is a recommendation system that will inform the driver about some actions to reduce emissions.

3. **Traffic jam detector system**: This system will try to identify in real-time when there is a traffic jam in an area. The detection is based on geo-tracking data and sensors’ values from every car. Traffic jams alerts will be sent to the driver’s smartphone applications. Drivers will decide by their own, if they avoid the traffic jam and change their route or not.

Additionally, two more modules, that will receive results from the previous systems, are defined:

- **Visualization module**: The main purpose of this module will be to help to analyse data by visualization. It will allow us to check both raw data from car sensors and results.
- **Notification module**: This system will inform drivers about important information generated by the previous systems.

The following objectives have been set to build up the previous systems:

- **Initial objectives**:
  - **CC_O_0.1**: Deployment of a Big Data infrastructure with descriptive and predictive analytics capabilities. It should support both batch and real-time processing. Descriptive analysis is useful to detect the events that are happening based on historical and current data (e.g. traffic jam or poor running engine). Predictive analysis involves complex model-building and analysis in order to predict future events or trends (e.g. breakdowns).
  - **CC_O_0.2**: Development of API services to carry out data injection to the Big Data infrastructure.
    - **CC_O_0.2.1**: Car sensors API by SoFLEET.
- **CC_O_0.2.2**: Car geo-tracking API by SoFLEET.
- **CC_O_0.2.3**: Fault codes severity interpretation API, it will be developed by Autoaid and accessed through SoFLEET.

  - **CC_O_0.3**: Development of a generic visualization tool for data sets analysis:
    - **CC_O_0.3.1**: Visualization of cars' sensor values over time allowing multiple selections based on attributes such as time intervals, model of car, year of production, etc.
    - **CC_O_0.3.2**: Real-time geolocation of vehicles in a map, with the possibility of showing the current journey path already covered by a given vehicle.
    - **CC_O_0.3.3**: Visualization of previous journeys of each vehicle.
    - **CC_O_0.3.4**: For every journey (both real-time and historical), it will be possible to select the kind of event to be displayed. They will be displayed over the map according to their location and timestamp.
    - **CC_O_0.3.5**: For every historical journey, it will be possible to replay the path of the vehicle in the map according to geolocation, timestamp and other values such as speed. To allow a soft replay, intermediate values will be interpolated.

  - **Breakdown estimator system objectives**:
    - **CC_O_1.1**: This objective extends the SoFLEET smartphone application to allow drivers to:
      - **CC_O_1.1.1**: notify information about a current car breakdown.
      - **CC_O_1.1.2**: be notified with a maintenance report.
    - **CC_O_1.2**: Application of machine learning techniques to carry out predictive analysis of car breakdowns. Several models will be trained and evaluated based on values of sensors' cars, including DTCs, and breakdown notifications received from the driver of the mobile application.
    - **CC_O_1.3**: Visualization of breakdown report for every car.
- **CC_O_1.4**: Building up the service that will send the cars' breakdown reports to the drivers' smartphone application.

- **Emissions reduction system objectives:**
  - **CC_O_2.1**: Monitoring of fuel consumption of every car to identify situations when their consumption has increased.
  - **CC_O_2.2**: Monitoring of drivers' behaviours to identify non-eco-friendly ones.
  - **CC_O_2.3**: Extension of SoFLEET smartphone application to allow drivers to be notified with recommendations in order to reduce emission.
  - **CC_O_2.4**: Building up the service that will send the emission reports to the drivers' smartphone application.
  - **CC_O_2.5**: Visualization of historical emission evolution for every car.

- **Traffic jam detector system objectives:**
  - **CC_O_3.1**: Applying algorithms to carry out descriptive analysis to identify situations where a car is in a traffic jam and evaluation of the traffic congestion in real-time.
  - **CC_O_3.2**: Extension of SoFLEET smartphone application to allow drivers to be notified about nearby traffic jams.
  - **CC_O_3.3**: Building up the service that will send detected traffic jams to the drivers' smartphone application.
  - **CC_O_3.4**: Real-time visualization of traffic jams in a map.
  - **CC_O_3.5**: Heatmap visualization of historical traffic jams by hour.

### 2.3 Use cases and Scenarios

The **stakeholders of the pilot** are the partners (SoFLEET, Autoaid and Answare) and the final users. These final users are the owners of fleet cars. They will be companies that will take advantage of the TT Big Data solutions implemented in the pilot. Every car driver has a smartphone where a SoFLEET application is installed. This application will be enhanced to achieve the pilot requirements: preventive car maintenance, CO2 emission reduction and traffic jam
detection. Following subsections describes each use case. For every use case, a figure has been included using UML Use Case diagram as notation.

2.3.1 Use cases for stage 1
In the first stage, the Big Data infrastructure will be deployed and initial data from cars sensors will be injected, from the early 5000 cars. The owners of these cars will be companies set in southwest of France, see Figure 2. Besides, a first version of the generic visualization tool for data sets analysis will be developed allowing the data scientist to prepare and analyse the data.

![Figure 2: Initial pilot area.](image)

The roles involved in the first stage use cases are (see Figure 3):

- **OBD II Dongle** is the device installed in the clients’ cars which sends data from sensors to the Big Data infrastructure.
- **Autoaid DTC Service** offers access to the comprehensive diagnostic data base of fault codes and provides classification for the fault codes severity.
- **Fleet manager** is the operator in charge of orchestrating and monitoring the fleet of cars.
- **Referential data service** offers important information related with the car such as fuel type, brand, model and number of cylinders.
- **Data visualization service** takes data both from historical and real-time database to display relevant information for its analysis.

- **Data scientist** takes advantage of the Big Data infrastructure to do data analysis and get insights and value from data in order to develop effective services to achieve the stated requirements.

The main target of this stage is to start to store and display data from sensors' cars. The initial 5000 cars (those with an OBD II dongle installed) will begin to send data, including location, from their sensors every 1 minute approximately to the Big Data infrastructure. For every new vehicle, important data related with its features will be requested by the referential data service. This process will be necessary just once for each car, since this information does not change. Additionally, the Autoaid DTC service will send fault code (DTCs) severity estimations for every car once a day.

All this information will be stored at the Big Data infrastructure and used, among other services, by the data visualization service. This service will offer a dashboard where the data scientist will analyse the data thanks to several functionalities:

- **Plot sensor values over time.** A generic data visualization tool will allow the data scientist to select different fields and generate plots about how these data fields evolve over time. This tool will be useful for the data scientist for a first understanding of the data.
o Display historical cars’ journeys. It will allow both the data scientist and the fleet manager to check and analyse the journeys of the cars. They will be able to display historical journeys of a given range of hours and dates.

o Display events along journeys. This functionality will allow both the data scientist and the fleet manager to analyse a given journey of a car by focusing on a given set of events, such as driver behaviour and accelerations in each axis. This information can help the data scientist to understand data and select them to carry out a descriptive or predictive model-building. The fleet manager could take advantage of this feature to check what happened during a given journey that did not end in time.

o Display real-time cars’ journeys. This feature is especially interesting for the fleet manager, since it will allow him to check the location of every car of the fleet and decide quickly which vehicle should attend a given service, saving time and CO2 emissions.

2.3.2 Use cases for stage 2
At the end of the second stage, a first version of all services should be deployed and validated against 15000 cars. So, some new roles will be included from this stage:

o **Preventive maintenance service** aims at avoiding unexpected breakdowns of cars. If a car is broken during a journey it implies a significant delay and waste of money for the company. To avoid these situations, it is worth to predict future car breakdowns and do preventive maintenance.

o **Reduction emission service** aims at reducing CO2 emission by monitoring fuel consumption. If it is higher than usual, then the system will try to find out the reason. For example, an aggressive behaviour of the car driver.

o **Traffic jam identification service** finds out situations where cars follow a given path slower than usual. These situations will be notified to car drivers to try to avoid them and save time and fuel (CO2 emission reduction). In the end, this system will be implemented in stage 3.

o **Notification service** is in charge of sending alarms and information detected by the previous services, such as breakdowns reports and traffic jam identifications.

o **Smartphone app** is the mobile application developed by SoFLEET and it is used by the car drivers.
Preventive maintenance system

Preventive maintenance system will use every data gathered from cars plus these from Autoaid DTC service and current car breakdowns notified by smartphone applications. Every time a car is broken, this fact should be notified through the smartphone app to the system, see Figure 4. The app should facilitate the notification process, so it will offer a set of car breakdown types. The driver will select one of them once a mechanic has evaluated the breakdown. It will be very useful to build classification models that allow the system to predict when a given breakdown will probably happen.

![Figure 4: Use case for predictive maintenance system](image)

The preventive maintenance service will carry out two tasks every day. Firstly, given the new data gathered, it will update their predictive models about car breakdowns. Secondly, a maintenance report will be generated or updated for every car. It will contain which breakdowns are more probable and their severity. These reports can be sent to the smartphone application in order to the driver takes it into account or displayed in a dashboard. The last option would be more appropriate to allow fleet manager to monitoring state of the cars and schedule preventive maintenance. It is expected that the number of breakdowns on the road will be reduced getting a better operational performance of the fleet by avoiding unexpected losing of availability of a car.
**Emission reduction system**

Given data values from cars' sensors, the emission reduction service will track the fuel consumption of the car and detect these situations where the consumption is higher than expected. To achieve this goal, the service will build a descriptive model that will identify these situations and try to find out causes of this raise. The service will focus on recognizing driver behaviour patterns that are not eco-friendly.

So, this service will supervise driver behaviour and generate alarms when it was not appropriate for reducing emissions, see Figure 5. These alarms will be sent to the driver in order to he or she takes it into account and modifies his/her behaviour having a softer driving. Moreover, these alarms will be available through a dashboard to allow fleet manager to check the drivers' behaviour and their evolution. If a driver does not change his/her behaviour, the manager could warn to him about this. If driver behaviour is not the origin of the consumption rise, then it will be registered in order to follow a set of checks such as tire pressure and over speeding. Allowing companies to be informed about car consumption performance, actions to increase this performance and evaluate how the suggested actions affect to cars' consumption should motivate them to reach a higher performance of the consumption of their car fleet.

![Figure 5: Use case for Emission Reduction system](image-url)
Traffic jam identification system

Traffic jam identification system will try to detect traffic jams in real time based on data gathered from vehicles, see Figure 6. New models will be built, deployed and evaluated to detect these situations. When the traffic jam identification service detects one, then it launches an alarm with information related with it (position and congestion level). This alarm is immediately sent to the driver smartphone application if the distance between them is lower than a stated one in the settings. Once the alarm arrives and is displayed in a map, the driver can evaluate an alternative route or go on with the current one.

Additionally, this information can be exploited by the fleet manager through the dashboard. The manager would be available to display the current state of the traffic plus location of the cars in real time and make decision based on it. For example, assign a car to attend a given services since it will not have to cross a traffic jam. Besides, a heat map of historical traffic jams would allow a fleet manager to know how the state of the roads usually is at a given hour and schedule journeys more efficiently.

Figure 6: Use case for Traffic Jam Identification System.

2.3.3 Use cases for stage 3
The number and type of use cases and scenarios will be the same than the previous stage. However, in this case, the number of vehicles will be 25,000 and the learning models of the
services will be improved continuously. Besides, a new service will be deployed to KPI monitoring. It will allow us to evaluate the performance of the TT solutions deployed in the pilot. The use cases for this new service will be defined as soon as the corresponding KPIs do.

2.4 Data Assets

Most data come from OBD-II dongle installed on vehicles. The OBD-II is a standard that specifies the type of diagnostic connector and its pinout, the electrical signalling protocols available, and the messaging format. Concretely, the devices chosen by SoFLEET are Munic Boxes which are equipped with GPS, 3-axes accelerometer, 2G quad modem and processing capabilities to run scripts and recognise events such as driving behaviour patterns, car crashes and sudden movements. All these data are available in the cloud using http protocol, see Figure 7.

Data gathered by OBD devices make up both "car sensors" and "geotracking" data assets. Besides, a "Referential" data asset will be used to retrieve data related with vehicles features such as brand, model, production year and engine power. It is highly important to take this information into account to carry out both descriptive and predictive analysis properly. Finally, "Autoaid API & Enhanced DTCs" data asset will be key to achieve preventive maintenance, since DTCs interpretation for every vehicle is a complex task. Data assets are identified and described in Table 1.

<table>
<thead>
<tr>
<th>Name of Data Asset</th>
<th>Short Description</th>
<th>Initial Availability Date</th>
<th>Data Type</th>
<th>Link to Data ID Card</th>
</tr>
</thead>
</table>

![Figure 7: OBD II Dongle components.](image-url)
| **Referential** | To have more precision on each vehicle, we use a service (SIV) which is available for French cars only. Another part of the data can be filled by users. | 30/07/2017 | JSON | https://3.basecamp.com/3320520/buckets/2480237/uploads/399832773 |
| **Autoaid API & Enhanced DTCs** | Autoaid provides the infrastructure to access telematics web services to various functionalities required to read data like identifier, sensor values or fault codes from electronic control units (ECU) built into a vehicle and translates their output for users. Fault codes (DTCs) and their descriptions are categorized by severity regarding vehicle maintenance by mechanics. | 03/04/2017 | JSON | https://3.basecamp.com/3320520/buckets/2480237/uploads/427885446/download/Data_ID_Car_d_AA_API.xls?attachment=true |

**Table 1: Data assets to be used in the Connected Cars Pilot**

### 2.5 Big Data Technology, Techniques and Algorithms

The Big Data solution of Connected Car Pilot follows a **microservice architecture** where both batch and streaming processing are employed. The microservices approach has become more and more popular, at least partly because companies who adopt this style of architecture move faster and compete better. The microservices idea is simple: larger systems should be built by decomposing their functions into relatively simple, single-purpose services that communicate via lightweight and simple techniques such as REST or a messaging system (Apache Kafka, MapR streams or Amazon Kinesis). Such services can be built and maintained by small, very efficient teams. Services could address different tasks such as data collection, preparation, model-building, data visualization and go on. It will allow us to work on the different systems of the pilot at the same time. These microservices are grouped to make up the systems defined in this pilot.
They communicate and coordinate each other using long-term and real-time data storage. Depending on the type of data storage used, they will carry out batch or real-time processing, respectively.

Figure 8 shows a diagram of the microservices architecture proposal for Connected Cars Pilot. At the bottom of the figure, we can see the legend. Horizontal cylinders represent real-time data storage (e.g. Kafka and MongoDB) and vertical ones correspond to long-term data storage (e.g. HDFS). Rectangles represent processing units. Clouds are used to represent external entities such as SoFLEET and Autoaid. The systems defined for this pilot (section 2.2) are identified by coloured boxes that group storage and processing units. All these systems are fed by the data assets defined in section 2.4 that are stored in real-time. Technology more probably used in this case
will be a message system such as **Kafka** or **Sofia2 Broker**. They are open source projects that offer fast message throughputs using subscription-notification of records paradigm. These records are called topics. Every microservice can subscribe to these topics and consume them at different speeds and in an asynchronous way depending on the requirements.

In the sections below, every data and processing for each main system will be described including technologies and algorithms that will be employed in this pilot.

### 2.5.1 Breakdown estimator system

It is the system depicted to predict breakdowns and allow fleet manager to do preventive maintenance. The system provides a maintenance report that may be updated weekly, if any estimation has changed. This system does batch data processing. A first stage collects, prepares and stores the data in a proper way for their analysis. This will be carried out by a microservice that will be subscribed to every topic, and its output will be stored in a long-term storage such as HDFS. The analysis consists on model-building for car breakdown detection. This predictive analysis will consider whole data gathered. Algorithms considered to be employed for model-building include supervised machine learning ones since car drivers will notify current car breakdowns. We are going to focus on decision trees such as Boosted Trees, Bagging, Random Forest and Rotation forest, among others. The frequency of this analysis must be stated. It could be a fixed periodicity or dependent on number of breakdowns received by car drivers.

Once models are built, they are deployed as microservices and executed in batch mode at a given frequency, for example every week. Their outputs are used by the maintenance report generation microservice to generate the maintenance reports. These reports are stored in both real-time storage and long-term storage. Reports are consumed by the microservices deployed in the visualization and notification systems. So, reports can be displayed and sent to the driver application as soon as they have been generated.

### 2.5.2 Traffic jam detector system

Like the previous system, it has also two stages, one for descriptive analysis and other for detection. Model building will be performed using batch processing. So, a microservice will prepare the data in order to facilitate later the building of the model. Unlike maintenance system, the built models for traffic congestion are going to be deployed to process data in real-time and identify traffic jams as soon as possible depending on the information from the connected cars at that time.
Traffic jams detections will be sent both to the visualization and the notification systems. In the visualization system, a microservice called “Car in a traffic jam” will implement the gadget to display in a map the identified traffic jams. However, the microservice in the notification system (“Traffic jam notifier”) will send notifications to car drivers who are in the vicinity of traffic congestion zones.

The algorithms will be based on the detection of velocity changes in nearby locations (we are evaluating the use of geohash geocoding for locations) throughout time and crowd sourcing techniques.

2.5.3 Reducing emission system
This system will monitor fuel consumption of every vehicle according to features such as power engine, fuel type and year of manufacturing, among others. An alarm will be triggered if a car deviates from its normal range of consumption. At the same time, drivers' behaviours are also monitored and, if they are not soft and eco-friendly, another alarm is triggered. Fuel consumption may increase because of a bad driving behaviour. If the behaviour is not appropriate, the system will suggest the driver to follow a set of actions that could solve the situation of fuel increased consumption.

Evolution of both driving behaviours and fuel consumption will be available in a dashboard. Two microservices (“Driver behaviour” and “Consumption”) will implement the gadgets which will be subscribed to the corresponding topics. At the same time, two microservices (“Consumption Notifier” and “Behaviour mode notifier”) in the notification system will send the generated alarms to the corresponding car driver's application.

Driving behaviour identification is performed locally at every vehicle by the OBD II dongle. The OBD II dongle will send driver behaviour changes as events to the Big Data infrastructure. The microservice just checks every event received and forwards it to the other microservices. However, clustering and statistical algorithms will be needed to classified and identified fuel consumption variations from the normal range.

This system will detect alterations in the way drivers use their cars that can lead to higher emissions. It will also detect fuel consumption increments higher than normal.
## 2.6 Positioning of Pilot Solutions in BDVA Reference Model

In this section, solutions defined in the previous subsections 2.5.1, 2.5.2 and 2.5.3 are positioned in the BDVA Reference Model according to the techniques they use, see Figure 9.

![Figure 9: Positioning of Cars pilot solutions in BDVA Reference Model.](image-url)
2.7 Big Data Infrastructure

Regarding the Big Data infrastructure, the identified components that we are going to use are (these components will be present at all stages unless otherwise indicated in the component description):

- **Hadoop**: Hadoop is the core for most Big Data infrastructures. It comprises of:
  - **HDFS**: the Hadoop distributed file system, which was designed for storing very large files on even inexpensive hardware organized in clusters and optimized for write once read many patterns.
  - **Yarn**: Yarn is a resource management system for job scheduling and monitoring of distributed programs (not only Map-Reduce jobs) on Hadoop clusters. It is installed with the same package for Hadoop installation from Hadoop 2.0.
  - **MapReduce**: Provides a parallel processing system for data sets.

- **Spark**: Spark is a general processing engine designed to perform batch and streaming processing. It also can be used for interactive queries and machine learning. Spark is planned to be used from stage 2. Other ML oriented components such as Mahout and H2O can be used.

- **Kafka**: Kafka is aimed for distributed streaming platforms allowing the publication and subscription to streams of records.

- **Pig**: Pig is a high-level platform for creating programs on Hadoop that abstracts the programming from the java MapReduce language (Pig can be used also for spark jobs).

- **Real Time Database**: We will need a real-time database to collect the information for car sensors. Among the possible options are MongoDB and the time series database InfluxDB.

- **General purpose database**: A general purpose DBMS (like Cassandra or ElasticSearch) could also be necessary to store Autoaid processed data, configuration parameters, results from analysis and other parameters as weights deduced in the machine learning processes. For historical database, HDFS through Hive could be use.

The operating system of choice will be a free distribution (Ubuntu, CentOS or openSUSE). We will need a stable system, of course, but at the same time a flexible system because the Hadoop
ecosystem is changing continually and new versions and services are issued very frequently. For this reason, we will probably choose Ubuntu Server 14.04 LTS 64 bits.

Concerning security, SoFLEET will remove from its web services personal data to guarantee that all the information from vehicles that are going to be stored in the Big Data databases is anonymous. Communication with web services will be done through https using hashed passwords.

The following sections summarize the specifics for each stage regarding resources consumption and scalability.

2.7.1 Peculiarities of stage 1
As Big Data infrastructure, we will use only one machine for stage 1. The foreseen amount of data at that phase makes this scenario more suitable for us. Results of that phase will give us a more insightful knowledge of whether we will scale horizontally the solution through a cluster or not.

For stage 1, we will configure the Single Node Hadoop cluster in pseudo distributed mode, meaning that all master as well as slave daemons will run on the same machine.

Based on the expected workload for Stage 1 we could work with a node with 16 GBs of RAM and 4 virtual CPUs with around 2.3 GHz for speed capacity.

The estimated quantity of data that SoFLEET server receives each day varies between 800MB and 1 GB for 5000 vehicles. For Stage 1, the maximum amount of raw data will be:

\[ 3 \text{ months} \times 1 \text{GB/day} \times 31 \text{ days/month} = 93 \text{ GB} \]

We will round up this capacity to 200 GBs (apart from SoFLEET information we will have Hadoop access control information, Autoaid data, etc.).

2.7.2 Peculiarities of stage 2
For stages 2 and 3 we will have to scale to obtain more resources. As machine learning algorithms, will be involved in these phases, processing capacity and RAM memory should be incremented (we estimate up to 24 cores of CPU and 128 GB RAM for stage 3, but this could vary depending on our findings). Another important aspect to consider is capacity. In state 2, the maximum raw data capacity will be roughly:

\[ 93 \text{ GB (for stage 1)} + 1 \text{GB/day} \times 31 \text{ days/month} \times 6 \text{ months} \times 15000 \text{ cars/5000 cars} = 744 \text{ GB} \]
Again, these numbers should be updated (at least doubled) to reflect the increment due to control information, new generated data coming from ML models, Autoaid information, audit history, possible new information sources like weather conditions, possible traffic information through Datex 2, etc.

2.7.3 Peculiarities of stage 3
For stage 3, we envisage a maximum amount of injected data coming from cars given by the next calculation:

\[
744 \text{ GB (for stage 2)} + 1\text{GB/day} \times 31\text{ days/month} \times 12\text{ months} \times \frac{25000\text{ cars}}{5000\text{ cars}} = 2.6\text{ TB}
\]

The same remarks made in stage 2 concerning additional storage capacities, due to possible new data sources and generated data, are also valid for stage 3.

At stage 3, we can even consider build up a Hadoop cluster with several nodes to see if the processes are executed more rapidly and test additionally the horizontal scalability. This cluster would consist of:

- One active Namenode for metadata of the data being stored.
- One secondary Namenode (if standby Namenode is not used): A secondary Namenode takes the responsibility of merging fsimage (the snapshot of the Namenode’s filesystem after start up) with edit logs (sequence of changes after the Namenode started).
- One standby Namenode: With the introduction of High Availability since Hadoop 2.0, the standby Namenode provides automatic failover in case of The Active Namenode fails. That means that the secondary Namenode is not needed.
- x Datanodes: The Datanodes will store the actual data (dimensioned according to the data volume, throughput and replication policies).
2.8 Roadmap

The planning of the development of the project to achieve the proposed objectives is summarized in Table 2. For its elaboration, the development time and the amount of data available at each stage have been taken into account.

<table>
<thead>
<tr>
<th>Stage</th>
<th>Delivery Date (Project month)</th>
<th>Features / Objectives Addressed</th>
<th>Embedding in Productive Environment</th>
<th>Big Data Infrastructure Used</th>
<th>Scale of Data</th>
</tr>
</thead>
<tbody>
<tr>
<td>S1: Technology Validation</td>
<td>M6</td>
<td>CC_O_0.1 CC_O_0.2 CC_O_0.3</td>
<td>Big data infrastructure (at least the modules used for the objectives). Data ingestion. Visualization module.</td>
<td>1 machine (200 GBs) with Hadoop, RTDB, General Purpose DB, Kafka, Visualization (depending on DB e.g. Kibana for ElasticSearch)</td>
<td>5000 cars</td>
</tr>
<tr>
<td></td>
<td></td>
<td>CC_O_1.1 CC_O_1.3 CC_O_2.x (except CC_O_2.3 and CC_O_2.4)</td>
<td>S1 + implementations of breakdowns detector system + emissions detector system + visualizations for emissions and breakdowns.</td>
<td>S1 state + implementation of algorithms for breakdown detection and emission reduction (including Spark/Mahout algorithms for ML) + visualizations for breakdowns and emissions.</td>
<td>vertical scalability (RAM/CPU/disk) to support up to 15000 cars</td>
</tr>
<tr>
<td>S2: Large-scale experimentation and demonstration</td>
<td>M12</td>
<td>CC_O_0.1 CC_O_0.3 CC_O_1.1 CC_O_1.3 CC_O_2.x (except CC_O_2.3 and CC_O_2.4)</td>
<td>S1 + implementations of breakdowns detector system + emissions detector system + visualizations for emissions and breakdowns.</td>
<td>S1 state + implementation of algorithms for breakdown detection and emission reduction (including Spark/Mahout algorithms for ML) + visualizations for breakdowns and emissions.</td>
<td>vertical scalability (RAM/CPU/disk) to support up to 15000 cars</td>
</tr>
<tr>
<td>S3: In-situ trials</td>
<td>M24</td>
<td>CC_O_0.x CC_O_1.x CC_O_2.x CC_O_3.x</td>
<td>S2 + implementations of notifications of breakdowns and emissions + implementation of traffic jam detector system objectives.</td>
<td>S2 state + improvements</td>
<td>25000 cars + scalability (maybe horizontal)</td>
</tr>
</tbody>
</table>

Table 2: Roadmap Table for Connected Cars Pilot
3 Design of Replication Pilot: Connected Trucks

Logistic service providers, fleet operators and truck drivers have to cope with various challenges in their day to day business. Customers expect 100% reliable and accurate transport execution from pick up to delivery processes. The traffic network infrastructure however often reaches its capacity limit especially in certain urban areas, at international seaports, cross-border areas or other logistic hot spots. Life traffic information and traffic predictions are widely available via different navigation systems or other traffic apps. Nevertheless, truck specific traffic information or real time information for Heavy trucks are often not available or at least not in the required quality or format for truck routing or logistic planning applications. Lane specific traffic information for trucks for instance about truck congestions at border stations are not available (only TMC messages via radio broadcast can be received on certain traffic events at borders). To overcome this lacking the Sustainable Connected Trucks pilot is being set up applying different Big Data approaches for added value traffic information and precise routing and planning applications specifically designed for heavy truck fleets.

3.1 Requirements

Within the connected trucks pilot 3 main requirements have been identified:

- **R1**: Enable truck drivers / fleet managers to reduce buffer times and delay times often at ramps/ports.
- **R2**: Enable planners to adapt plans according to realistic travel times resp. ETA-values.
- **R3**: Enable technical systems to provide more reliable routing results and traffic predictions.

These requirements cover 3 different perspectives. At first the end-user perspective, especially the truck driver who has the need in day to day business to be on time at the customer/logistic facility, this is more the real-time aspect of the pilot. Secondly, the planners/dispatcher perspective who has to plan the operation of the entire vehicle fleet in a realistic and predictable manner and thirdly the technical point of view where systems have to be enabled to provide more reliable and exact routing and planning results to the end-user for different logistic use cases.

3.2 Objectives

There are different approaches to address the requirements for a sustainable connected trucks fleet.
Main objective within this pilot is to process large amounts of big data specifically related to truck fleets all over Europe. The approach is to generate results out of this with specific added value for logistics application such as routing and planning.

Especially at terminal, toll stations, ferry stations or other logistic hot spots a detailed and enhanced detection of traffic flow and situation for trucks is the basis for enhanced planning and optimization systems for fleet managers.

Main challenge is to extract very detailed and specific traffic infrastructure related information, which is of high relevance for heavy trucks. The following five objectives within this pilot can be summarized:

- **CT_O_1**: Using Big Data approach on traffic and movement data specifically related to truck fleets.
- **CT_O_2**: Generation of truck fleet specific traffic patterns.
- **CT_O_3**: Provision of truck specific traffic information (actual and predictive).
- **CT_O_4**: Event related information on specific POIs (ramps, ports, toll stations).
- **CT_O_5**: Analyses on infrastructure, events, lane-specific traffic conditions for trucks (satellite images).

### 3.3 Use cases and Scenarios

For modern tour-planning and optimization software, detailed data of network conditions is crucial. For an accurate trip planning as well as for prognosis models exact travel times on the entire network are indispensable. Three main areas for innovation potential have been identified:

- Speed patterns
- City logistics
- Traffic modelling

**Speed patterns**

Speed patterns describe the actual routing (driving) speed on a road element for a defined time cluster over a day. Speed patterns give a forecast of speeds every quarter of an hour (or a similar raster of another granularity) for every day of the week.

At the moment, PTVs speed patterns production is based on historic speed patterns delivered by a Big Data provider (e.g. TomTom). Within TT pilot activities specific speed patterns can be created, e.g. for specific vehicle types, entire company fleets, specific regions or very detailed for...
urban logistics, by using specific input data. Speed Patterns can be used in software solutions that needs exact routing speeds for exact planning and optimization results.

At PTV these logistic standard products are PTV Smartour, Map & Guide or the PTV xServer. Furthermore, for Drive & Arrive and the arrival board exact travel times are a crucial part to calculate the time of arrival of the customer and to organize the ramps.

**City Logistics**

Optimisation of logistic traffic in urban areas becomes an increasing value in PTV planning tasks. Especially in congestion areas and big cities, detailed information about road conditions and travel times are important for a reliably customer supply. Very detailed speed patterns are one column for an exact calculation of ETA.

**Traffic modelling**

Furthermore, Speed patterns are one input for an automatic OPTIMA model calibration.

The uses cases/ scenarios and can be depicted as follows:

- **Use Case 1 – Reliable Routing**
  - Big Data analyses of truck specific traffic data in order to provide reliable routing results with high probability.

- **Use Case 2 – Actual and predictive traffic information for trucks**
  - Generation of traffic information with prediction for heavy trucks.

- **Use Case 3 – Analytics of event based traffic situations at certain POIs**
  - Using historical data (congestion, weather incidents at ports, terminals) to derive typical traffic situations in future.

- **Use Case 4 – Infrastructure related analyses of traffic network conditions on corridors, logistic relevant areas**
  - Analyses of satellite images regarding traffic network conditions for heavy trucks (e.g. lane specific at borders).

The use cases will not cover a certain area, traffic/ movement data will cover (in different level of details) the whole European area. The truck fleet from the end user JDR (1000 trucks per day) are operating all over Europe.
3.4 Data Assets

Satellites record our planet 24 hours a day and deliver current data from every place on earth. Different techniques like active or passive remote sensing deliver data for multiple applications and analysis methods to get information on the earth’s surface. The information of satellite images is not only available for a specific time, but is continuously collected worldwide. The repetition rate of information about a specific place of interest depends on the orbital period of satellites and the size of the recorded area. It is possible to detect the current state of a location, but also to detect changes over time. Additionally, satellite images cover a large area and are not limited by geographic or political boundaries. Nevertheless, the influence of atmosphere and weather on the usability of images should not be underestimated and severely influences the number of usable images for analysis.

Within the Transforming Transport project, the Fraunhofer IGD validates the effort of remote sensing satellite images in context of applications for truck fleet managers. Long-term and short-term data are currently used for the planning of truck fleets. The task is to investigate how far satellite images (which cover both temporal aspects) can extend the existing data base for planning. Therefore, demands are defined and documented on different topics like quality, spatial and spectral resolution or repeat cycle to deduct answers to usability, availability or reliability for the requirements of connected trucks.

The examination is primarily focused on two remote sensing data sources:

- **Copernicus** as the European Program for the establishment of a European capacity for earth observation. The information is freely and openly accessible. Copernicus comprises a range of technologies, such as radar and multi-spectral imaging instruments with different spatial resolution for land, ocean and atmospheric monitoring. A group of satellite missions continuously records images in predefined segments. (For further information: [http://www.copernicus.eu/](http://www.copernicus.eu/))

- **European Space Imaging (EUSI)** as a leading supplier of commercial very-high resolution satellite images. It derives services and multi-spectral images from the missions WorldView-1, WorldView-2, WorldView-3, GeoEye-1, QuickBird and IKONOS (archive). Images of points of interest are delivered on request. (For further information: [http://www.euspaceimaging.com/](http://www.euspaceimaging.com/))

The focus gives the opportunity to examine these sources in depth and draw a conclusion for productive use of satellite images as an additional data source on different planning stages in
general. Together with the other partners the integration of this information to existing data sources will be validated considering added values and idle capacities in truck fleet management.

TomTom is collecting Floating Car Data since 2008 for the measurements of speeds in the road network. While starting with a small fleet and a low coverage at the beginning this has grown enormously covering nearly the whole existing road network, which is displayed on a heat map based on FCD, see Figure 10.

![Figure 10: Heat map based on FCD for European road traffic network](image)

This strong increase is on the one hand based on the number of installed TomTom Navigation Devices and Navigation Apps for Mobiles. On the other hand, due to the fact that a lot of Automotive Manufacturers (Volkswagen, Daimler, Audi, BMW ...) are using TomTom traffic products for in car usage, and are also deliver Floating Car data in the back channel. These data are all stored in a huge in-house data structure and used in the data fusion for live evaluation on travel times and traffic events like jams and road works, but also for historical analysis like jam probabilities and time based speed profiles. This database contains right now trillions of anonymous speed measurements, and is growing daily by billions of new data points.
Name of Data Asset | Short Description | Initial Availability Date | Data Type | Link to Data ID Card (in basecamp)
---|---|---|---|---
Movement data | FCD, historical truck fleets | Q2 17 | GPS traces | Will follow
Traffic data | Mixed fleet traffic information | Q3 17 | Traffic information, patterns | See above
Satellite images | Copernicus or EUSI satellite images from different locations corridors | Q4 17 | Images different picture formats | See above

Table 3: Data assets to be used in the Connected Trucks Pilot

3.5 Big Data Technology, Techniques and Algorithms

3.5.1 Big Data technologies to be used

Network data are usually needed in an already prepared format of PTV. This means that the network, e.g. a TomTom Europe street graph is available at a certain time. After a short time, further preparations which need a referencing to this network should be processed.

For batch processing or data analytics use cases a fast access to all data must be established. For different processing steps, e.g. different batch processing, data analytics steps or even a data selection, there may be the need to store the data multiple times in different formats to guarantee a performant access. Last but not least, a huge amount of data will be collected and stored.

It is required to use cloud technologies for being able to scale in a wide horizontal way without having the initial investment costs for a huge hardware stack.

- First experiences using cloud technologies for data preparation have been made during developments for geo referencing and data aggregation. During the development, further cloud technologies must be evaluated and Know How must be established.
- Concrete cloud solution could be for instance Microsoft Azure.
3.5.2 Big Data techniques and processing

The next step could be the generation of speed patterns already realized as a prototype. For the generation of time dependent traffic patterns, it is also necessary to reference the FC data to the network. Different to the Turn ratios use case, it can be decided whether map matching or routing between consecutive FCD-positions is used or if a simpler point to network reference is determined. Depending on that, the coverage and quality of the reference result and the generated speed patterns may vary. If the referencing is implemented within the turn ratios use case, the results can be reused for further processing in both use cases.

Following to the referencing an aggregation of speeds for all road network segments will be implemented. Therefore, the granularity of time slots can be configured for which data should be aggregated. Furthermore, optional filter strategies can be discussed, to leave out unrealistic speed values contained FCD due to technical errors. The results then can be exported to the already defined speed pattern Feature Layer format which can already be used by the PTV products.

The processing of data for reliable routing information is the challenging aspect. First of all, extracting usable probabilistic speed profiles from TomTom probe data will expand drastically on the data volume of routing-relevant information. This is because of instead the single, average flow speed per road stretch, we aim to provide a more thorough description of the flow speed distribution. Thus, Big Data approaches need to be considered for managing the resulting amount of data.

Secondly, actually computing a reliable route based on this data set does not only pose high demands on computational resources and suitable algorithms but also on accessibility of this data set, so this aspect will pose another challenge.

In this project, existing analysis methods from remote sensing or image processing will be examined in order to derive relevant information for fleet planning. Results can be the detection and classification of vehicles with their individual speed; or the utilization of routes or lanes at certain days or times a week. The focus of investigation can vary on different obstructions to traffic. Therefore, the investigation can include terrain analysis for landslides or high-resolution analysis for individual closed lanes. The question of an automated evaluation is checked. In particular, the analysis of transport hubs or points of interest like airports can give additional information on critical areas.

The analysis of an example area such as the Rhine-Main metropolitan region is intended to investigate various questions in a first step. Based on these results, an analysis of a practical
approach can be carried out. The availability of historical and latest images for a specific region is tested and transport hubs will be analysed.

3.5.3 Analyses and foreseen results

The results are used to validate the added value of individual analytical methods or the use of satellite images in general for the transport planning of trucks. The goal is to examine if the use of satellite images can generate an additional value to the existing long and short term analysis methods.

TomTom will analyse its historic database of GPS traces to provide a novel form of travel time information that allows for reliable trip predictions. Instead of aggregating observed travel times to single estimates per links, aggregated probability distributions on travel times will be provided. This data can be used to compute reliable ETAs of given trips, that is, not only the expected total travel time, but also upper bounds for given confidence so that a trip planner can rely on reliable trip execution even in multi-legged itineraries. Secondly, TomTom will evaluate the use of such distributions in reliable route computations, prototyping a route service that no longer minimizes the expected travel time (“fastest route”), but rather the upper bound for given confidence.
### 3.6 Positioning of Pilot Solutions in BDVA Reference Model

The approach and structure within the connected trucks pilot can be positioned within the BDVA Reference model as Figure 11 shows. Numbers 1, 2 and 3 correspond to subsections 3.5.1, 3.5.2 and 3.5.3, respectively.

![Diagram](image)

**Figure 11: Positioning of Trucks pilot solutions in BDVA Reference Model.**
3.7 Big Data Infrastructure

The Big Data infrastructure to be set up and used within Transforming Transport will be based on existing components whereas the results out of the Big Data use cases will be deployed in applications which can be demonstrated within different end user scenarios. A first draft of the architecture is shown in Figure 12.

![Figure 12: Big Data infrastructure approach for Connected Trucks Pilot](image)
### 3.8 Roadmap

The initial Roadmap for the Replication Pilot Sustainable connected trucks with respect to the three stages is drafted in Table 4.

<table>
<thead>
<tr>
<th>Stage</th>
<th>Delivery Date (Project Month – started/ completed)</th>
<th>Features / Objectives Addressed</th>
<th>Embedding in Productive Environment</th>
<th>Big Data Infrastructure Used</th>
<th>Scale of Data</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>S1: Technology Validation</strong></td>
<td>M06/M12</td>
<td>Setup and implementation of infrastructure, analyses of technology and historical data, CT_O_1.</td>
<td>TomTom, PTV, backend systems</td>
<td>TomTom, PTV, FhG</td>
<td>large</td>
</tr>
<tr>
<td><strong>S2: Large-scale experimentation and demonstration</strong></td>
<td>M13/M18</td>
<td>Provision of routing results and traffic information, CT_O_2, CT_O_3.</td>
<td>TomTom routing; PTV xServer; JDR Planning System</td>
<td>TomTom, PTV, FhG</td>
<td>large</td>
</tr>
<tr>
<td><strong>S3: In-situ trials</strong></td>
<td>M24/M30</td>
<td>Provision of infrastructure related information and traffic/ routing information, CT_O_4, CT_O_5.</td>
<td>see above</td>
<td>see above</td>
<td></td>
</tr>
</tbody>
</table>

*Table 4: Roadmap Table for Connected Trucks Pilot*
4 Commonalities and Replication

4.1 Common Requirements and Aspects

The main goal for the Sustainable Connected Trucks pilot is to optimize the management of vehicle fleets in order to predict and improve ETAs. In doing so, analysis of satellite images will be used in the process to detect possible situations that could affect schedules. On the other hand, the Sustainable Connected Cars pilot aims to detect and prevent DTCs and spare CO2 emissions. In this sense, both pilots complement each other, but we cannot consider the trucks pilot as a replica of the cars pilot. Trucks will not take advantage of the algorithms and processes used in the car pilots to detect and prevent breakdowns and make the driving eco-friendlier. Similarly, cars will not use the routing resources and image analysis from the trucks pilots. Devices installed in cars are different that the ones installed in trucks and APIs to access data are different and won’t be shared between the two pilots.

Nevertheless, as a result of both pilots, CO2 emissions should be reduced, though using different means. Another commonality could be the detection of traffic jam situations. In this sense, it must be take into account that GPS systems for trucks are different from GPS systems for cars, as trucks are not allowed use roads that are prohibited to them.

4.2 Replication

In both pilots, every vehicle will contribute as a sensor in the road network. Taking advantage of this fact, one of the objectives of the cars pilot will be to detect traffic jams while the trucks pilot will additionally take information from traffic jams to be used in the calculation of speed patterns that will be used in the routing algorithms. In this context, the determination of traffic congestions, the trucks pilot will add much more complexity to the foreseen processing in the initial pilot. Traffic events can also be used in the determination of traffic jams (for this Datex 2 could be a good source of information) and will be used in the connected trucks pilot.
5 Conclusions

In this document, the design for the Sustainable Connected Vehicles Pilots within the TT project has been described. The design comprises the specification of the requirements and objectives, the enumeration of the data assets that will be used, the description of use cases and scenarios and the Big Data infrastructures and components that will be deployed for the three stages. Both pilots will try to help fleet managers and drivers reduce costs and achieve a more efficient planning. The Sustainable Connected Cars Pilot will be focused on preventive maintenance, reduction of polluting emissions to the environment by means of decreasing fuel consumption and detection of traffic jams. The Sustainable Connected Trucks Pilot will reduce times and delays by the determination of reliable routing and traffic predictions, enabling also fleet managers to adapt plans according to realistic travel times. Both pilots complement each other and will benefit from the Big Data technologies in terms of scalability, data versatility and real time and batch Big Data processing, among other benefits.