Date: **20 February 2017**
Subject: **Data from Vehicles**

"Data from Vehicles"

What are we talking about?

Workgroup:

Leo Bingen (RAI Association)
Wouter van Haaften (University of Amsterdam)
Chris Huijboom (HAN, University of Applied Sciences, Arnhem & Nijmegen)
Mike Pinckaers (ANWB, Royal Dutch Touring Club)
Pierre van der Stokker (Beijer Automotive)

Client:

Joelle van den Broek (DITCM, Dutch Integrated Testsite Cooperative Mobility)

1. Introduction

Nowadays, systems in modern vehicles produce enormous amounts of data. This is true both for the passenger car and the commercial vehicle as well as for the motorcycle and even the bicycle. These data offer a foundation for information about the condition and the use of the vehicle. Furthermore, in case of data communication with the surroundings, these data offer considerable possibilities for traffic management, traffic information and incident management, among other purposes – both national as well as international (European). An increasing number of (commercial) services based on these vehicle data are becoming available from the private sector (including the automotive branch and service providers in the field of mobility) concerning the use and condition of the vehicle, such as traffic information and reminders for servicing and repairs, and driving-task support services in the aftermarket. The public sector (EU, national and local governments) is very interested in the data from vehicles to modernise (read: digitise) and improve its information provision to the road user. But there are many more possibilities; certainly when the vehicle obtains an online connection with the internet.

In practice, it appears that there is still much uncertainty regarding the question: "*Which data are collected in a vehicle, who can access those data, and who can/may process them and under which conditions?*"

This is true in a broad sense for interested parties such as governments, insurers, new parties in the market for mobility services, and the mobility sector itself with regards to applications for safety, comfort, vehicle condition, traffic information, accident analysis and other factors.

This document seeks to provide some clarity about these issues, so that parties (public and/or private) know what they are talking about and can actually take advantage of identified opportunities.

This document is limited to the *passenger car*, although most of what is discussed is also applicable to (light) commercial vehicles. Furthermore, this document is limited to the data which is present in the *internal* network of the car and data which may be generated by fixed provisions installed in the vehicle aftermarket – otherwise known as *vehicle-mounted* systems. At this stage, *personal* (mobile) systems, such as smartphones and tablets and the data produced by these systems, are disregarded for the time being.

2. Data from vehicles

With regards to data from a vehicle, various aspects can be distinguished which are relevant in the context of this document. These aspects are:

2.1. Vehicle data
2.2. Availability
2.3. Legal aspects

This distinction is applicable to both the *internal* (branded) systems of a car as well as to any (fixed) systems installed *aftermarket* (a 'box').

2.1. Vehicle data

By far, the greatest ambiguity appears to concern vehicle data. For many, there is insufficient knowledge of which *system* produces which data within the vehicle and for which purpose.

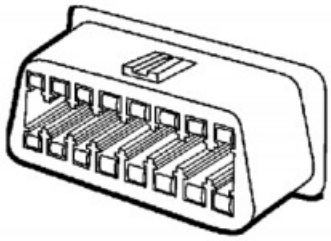
Moreover, it is often not clear how these data are accessible for which applications and for which authorities. Various types of vehicle data are further inventoried in this section. Doing so, we limit ourselves to data relating to the primary automobile functions. Service and entertainment data are not considered here.

2.1.1 Internal vehicle systems

The data sources (vehicle sensor data) are available in the internal network of the vehicle. The reading out of vehicle data from this internal vehicle network by telematics systems via the so-called Controller Area Network (CAN) bus in the vehicle can occur in various ways. A physical connection can be created on the diagnostic bus of the standard diagnostic connector (On Board Diagnostics, or OBD) which one will find in every modern car. This allows access to a selection of the data present in the vehicle.
In addition, a direct physical connection can be made to the CAN bus network that is also present in every modern automobile.

2.1.1.1 Diagnostic data, generally accessible via EOBD / OBD II

Each modern car has a factory-installed diagnostic connector, also called a Data Link Connector (DLC) or an On Board Diagnostics (OBD) connector. OBD is an American emissions standard. The current version is OBD II, which is identical to the European variation EOBD, the emissions standard valid for cars sold in Europe.



As of 2001 (for petrol engines) and 2003 (for diesel engines), each new car imported into the EU must be provided with a standardised diagnostic connector.
This standard must enable non-branded car companies to be able to make a diagnosis in the case of a discrepancy in the mixture control (and thus in the emissions). If a malfunction or discrepancy occurs in the mixture control, a fault code is stored in the electronics of the car and the driver is informed via a notification on the dashboard. These fault codes and/or data can be read via fixed protocols by means of a diagnostic device. This diagnostic device is connected to the diagnostic connector and transmits (queries) a request for a so-called Parameter ID (PID) to the electronics system in the car, which then answers with the relevant fault code. This can be done, for example, via certain network protocols in the car such as PWM, ISO or KWP, but is nowadays done primarily via the CAN bus. Each car manufacturer does this in its own manner.

In addition to the diagnostic purpose for which the diagnostic bus was initially intended, it is possible to read out data from the vehicle during a trip in order to determine emissions. Moreover, this concerns only a *limited* number of parameters, such as speed, RPM, motor load and coolant temperature, among others.

As a rule, car manufacturers are very cautious about the use of the OBD-II port for getting access to CAN bus data by third parties. This is partly related to the fact that in order to acquire certain CAN data via the OBD-II port, a message must first be 'written' on the CAN bus in order to then obtain the desired data as an *answer*. If this is not performed with the right equipment and the right expertise, risks may occur for the correct functioning of the internal systems of the car. Nevertheless, legally authotorised parties must always be able to survey the car with regards to the (traffic) safety of the car driver and other road users.

2.1.1.2 Proprietary data, private sensor data via CAN bus

Another method of reading vehicle data is from the CAN bus network. Since 2003, practically every new car is provided with a network of one or more CAN buses. Via this CAN bus network, the control units and sensors are connected to each other and they exchange their information. In many car models, these are 'closed' CAN buses with no readout plug but there are also cars in which the selection of data from one or more CAN buses ends up in the diagnostic connector, with or without an integrated diagnostic bus.

Unlike the limited, *generally accessible* data on a diagnostic bus, the CAN bus network contains a wealth of data which only can be distilled with the proper expertise. These data include parameters such as speed, mileage, brake use, braking power, use of windscreen wipers, use of fog lights, outside temperature, seat belt use, seat use, fuel level and the like, also known as **Probe Vehicle Data**. However, these data are, as a rule, 'proprietary': in other words, not generally accessible and mostly encrypted by the car manufacturer so that they are usually not plainly understood and useful for more general purposes.

The reluctance of car manufacturers also applies to this mere 'reading of and/or listening to' data directly on the CAN bus. Therefore encryption, among other protections, is used although the risks with the use of the proper equipment and expertise are limited here. However, there are currently many systems available on the market for which there is insufficient certainty of the quality and the possible risks to the integrity of the in-car systems and the associated risks for liability, safety, warranty, etc.

2.1.2 Aftermarket systems

In addition to the previously discussed internal network of a car that is factory-installed, it is possible to install systems in a vehicle after production (aftermarket) for the purpose of reading vehicle sensor data and potentially generating data itself. These can by systems which are:

1. fully stand-alone and thus not connected to the internal (CAN bus) network of the car
2. systems that are connected on the OBD-II port
3. systems that are directly connected to the CAN bus afterwards

The systems included under options 2 and 3 concern boxes or dongles from a variety of suppliers and service providers. These systems are, insofar as they are not fitted by the dealer of the vehicle brand, installed in the so-called 'aftermarket' (retrofit) and therefore fall outside the field of vision of the car manufacturer and its brand organisation. As mentioned, car manufacturers are generally hesitant to allow the connection of aftermarket systems to the OBD-II port. On the other hand, many of these systems have been implemented without a problem, thus providing the quality of the system itself as well as of the data being apparently sufficiently up to par. Recently, the Original Equipment Manufacturers (OEMs) themselves are also using dongles in their service provision.

2.1.2.1 Fully stand-alone

Examples of stand-alone systems include: navigation systems, vehicle tracking systems (Track & Trace), fleet management systems, trip registration systems, etc. These systems contain dedicated sensors, such as a GPS module and sensors for speed, acceleration/delay, inclines, etc., which are necessary for the intended applications and for communication with a back office through a SIM card, for example. Usually, no link with the internal (CAN bus) network of the car is required for these 'simple' systems.

2.1.2.2 Connected to the OBD-II port

There are systems available on the market which can be connected to the OBD-II (diagnostic) port with a plug. The most common form of these are the so-called 'dongles'. These are small boxes with a built-in plug which can be directly inserted into the OBD-II port. Although some dongle suppliers indicate that the dongle can be used to extract practically all of the data from the car's internal network, this is generally not true. As a rule, it concerns only that data which, as previously described, is *generally accessible* and limited in scope. If a dongle or another system which is connected to the OBD-II port on the CAN bus is nevertheless able to extract data from the CAN bus using a *query*, the previously mentioned risks ensue for the manufacturer, certainly if the queries occur while driving.

The easy installation also appears to be a great advantage of the use of dongles. Specifically, each driver can look up the diagnostic connection in his or her car and plug the dongle into it. This ease of use is very appreciated by the users. This is evident from a pilot of the Royal Dutch Touring Association (ANWB), among other sources, where only 2% of the users found it difficult to install the dongle. In practice, however, it appears that dongles don't always function well due to the physical differences in the design of the diagnostic connection among the many brands and models. In addition, there is a great variety of quality among the dongles on the market.

Especially with regards to cybersecurity, caution is urged in the use of the diagnostic (OBD-II) port for linking with aftermarket systems. The establishment of minimum requirements for the use of aftermarket systems such as dongles but also fixed, fitted systems may eliminate the objections mentioned.

2.1.2.3 Aftermarket installation directly on the CAN bus

In principle, all electronics of a vehicle can be *read* via a *direct* CAN bus link, albeit via a *specific protocol* which is dependent on the relevant vehicle manufacturer, model and year of production. With such a connection on the CAN bus network, a universal provision exists with which all vehicle data present and useful in the vehicle can be made available flexibly for multiple services. All this, only by 'listening' to the CAN bus.

However, this does require smart technologies and especially the right expertise to not only be able to listen to these CAN data but also to be able to understand the data, without needing queries to the CAN bus to disclose these data. In addition, the encrypted (proprietary) data must be converted into usable data which can be universally used for information.

Aftermarket solutions already exist for interpreting/converting these brand and vehicle-specific protocols for all common vehicles into universally usable vehicle information and sensor data. Although this method does *not* require *querying* the CAN bus to obtain the desired data, nevertheless an 'opening' arises for inexperienced and/or malicious persons to break in. To actually prevent such a break-in, professionally installed CAN bus connections usually have no built-in writing possibility.

2.1.2.4. Use and services

Despite the reservations mentioned here concerning systems connected aftermarket directly on the OBD-II port (such as a dongle), such systems can nevertheless have added value for especially the private consumer, who requires only a limited functionality. The option to be able to easily add such systems and conveniently transfer them to another vehicle, coupled with the relatively low purchase price[1], may play a role here.

The establishment of 'conditions' for proper systems, such as those mentioned in 2.1.2.2, can make an important contribution towards more and better insight into the quality and reliability of such systems as well as the possible risks. Both the data generated by the internal system of the vehicle, as well as that generated by an aftermarket system, can form the foundation for the offering of (commercial) services to the driver/user of the vehicle. The extent to which the user himself or herself can determine which data for which purpose goes to which party influences to a great degree the user's selection of certain services.

A special position is reserved here for the so-called Event Data Recorder (EDR). This device, which is already mandatory in the United States, registers several relevant data in the last seconds before an accident. These data can support the analysis of the circumstances of the accident. Road users, insurers, police and justice staff may have an interest in access to these data. The EDR is also often supplied standard in new cars in the Netherlands, sometimes even without the car buyer knowing it. Partly for this reason, there is still much discussion about who has access to the data from the EDR and under which circumstances, what they may be processed for, and what the possible consequences may be regarding liability and data protection.

[1] In addition to the (possibly relatively low) purchase price of a system, the periodic recurring costs, such as those for data communication and services supplied, must be taken into account.

2.2. Availability

Finally, a distinction in terms of data from vehicles can also be made regarding:

- Data which is read out retrospectively.
- Data which is received in near real-time from the riding vehicle.

2.2.1 Reading out retrospectively

Offline data concerns data which is *periodically/incidentally* extracted from the vehicle. This may occur in case of the extraction of diagnostic data from the vehicle in the car company's workshop using a special test device whilst conducting servicing or repairing a malfunction. To do this, a plug of the test equipment is connected to the previously described OBD-II port of the vehicle. The memory of the vehicle's internal system is then read, including any fault codes. This concerns both the *generally* accessible data as well as the *proprietary* data, although limitations concerning this last type apply (also see 2.1.1.2). In addition, the internal systems can be reset in this manner by the car company, if necessary and with the proper authorisation.

2.2.2. 'Real-time' receipt of data

Increasing numbers of new vehicles fresh from the factory are able to transmit data to the cloud or to a private (connected) back office in (near) *real-time* or at least periodically. But even older cars can (still) be connected using an aftermarket system. This can involve both the data of their internal (CAN bus) system as well as data collected and transmitted via an aftermarket system. This generally concerns data of which a (near) real-time transmission (and collection by the recipient) is important for the application based on this.

With a trip registration system for tax purposes, for example, it is important that the position of a vehicle at any given moment is frequently recorded in a back office. Such a system should thereby be clearly linked to the vehicle and/or further measures must be taken to guarantee the continuity of the data stream. Furthermore, a (near) real-time connection is necessary to achieve the required timeliness of the customised provision of traffic information.

Currently, most 'real-time' connections (still) run via the GSM network. This can involve considerable costs for high-frequency data transmission.

However, developments regarding data communication – both speeds and costs – are proceeding quickly. After 3G and 4G, now WiFi-P and 5G communication opportunities are in sight. The (near) real-time data which are transmitted through vehicle data connections are currently still limited in scope. And the recipients of the data streams are still quite narrowly defined: predominantly just the car manufacturer or a service provider contracted by it or linked to the aftermarket equipment.

Nevertheless, the expectation is that this will change in the (very) short term as more cars will be connected, both from the factory as well as using aftermarket systems.

Legal aspects

2.3.1 Data protection

A share of the vehicle data is linked to the driver/owner/holder of the vehicle. With regards to data from a vehicle, therefore, it is important to gain insight into whether, and if so to what extent, we can speak of personal data to which data protection legislation applies. This concerns data which are identifying, such as name, address, registration number and Vehicle Identification Number (VIN), but also data which are at first non-identifying but in combination with other data can be traced to a person. The question of whether the processor of the data has access to such other data is especially important here.

2.3.1.1 Data which are not personal data

In this memorandum, data are not considered to be personal data if they are solely related to the functioning of the internal systems in the vehicle and thus can offer the basis for information about the actual condition and use of the vehicle and all associated systems and sensors[2]. This definition applies to both the internal network of the car (CAN bus) as well as to data originating from aftermarket systems. These data can, in turn, be separated into data used exclusively for diagnosis of the vehicle and data for other (brand) purposes. One condition is that these data do not contain data which can lead to the identification of a person. Think here of vehicle identification data and location data of the vehicle.

2.3.1.2 Personal data

If the vehicle, or the aftermarket systems therein, produce(s) personal data – including data traceable to a person – then the Netherlands Data Protection Act (DPA)[3] is applicable. This means that with regards to these data and in accordance with the Act, great care must be taken regarding the collection, storage and use of such data. The OEM or service provider must also facilitate the right of the user to view and correct personal data. A detailed description of the data protection aspects falls outside the scope of this document.

2.3.1.2.1 Internal network (CAN bus) of the car

Data originating from the internal (CAN bus) vehicle network generally contains no direct identifying data but there may be data 'traceable to an identifiable person', especially when a combination of these data with other data with regards to the vehicle is made. This could be location data of the vehicle combined with the registration number or VIN. In this way, the driver/holder of the vehicle could in principle be detected and it could be determined where this person has been and when. In general it can be said that all vehicle data can be traced to a person in combination with, for example, registration number and/or VIN. To the extent that such a link does not take place, the question of whether it concerns personal data must be determined based on the actual processing. Vehicle location data, and the degree to which the information is 'visible' outside the vehicle, play an important role. Especially the discussion of the degree

[2] There is still no consensus regarding what must be considered personal data. German DPA: Vehicle data is not personal data if the data do not leave the vehicle. French DPA: All data in the vehicle are personal data.

[3] This law will be replaced in May 2018 by the Regulation EU nr. 2016/679.

to which Cooperative Awareness Messages (CAM) –transmitted by the vehicle for cooperative driving applications – must be considered personal data is not yet finalised. There is reason to assume that, based on the new General Regulation on Personal Data (AVG in Dutch), an important share of the data which will be transmitted by the vehicle, and therefore can be processed by third parties, will be considered to be personal data.

2.3.1.2.2 Aftermarket systems

Because of the nature of the application of some aftermarket systems, data from these systems may directly contain personal data or data traceable to a person.

With aftermarket systems, a distinction can also be made between data which is specifically collected for vehicle diagnosis purposes (as part of the application of the system) and specific sensor data needed for a certain specific application of the aftermarket system. Think of, for example, Track & Trace systems or Trip Registration systems. With these systems, there can be a direct link with the personal data of the driver/holder of the vehicle as a result of the nature of the system application, so the data protection act should be taken into account.

2.3.2 Data control

When it comes to control over the vehicle data, it should first be establish that this control basically is a subject toan agreement between the parties involved. For example, the car buyer can be asked by the OEM to give consent for the use of vehicle data. With the signature on the agreement, the control over the data is initially arranged between the buyer of the vehicle and the OEM. The question is, however, what the consequences are of the inclusion of such a control clause in the purchase agreement of the vehicle. Specifically, the manufacturer may not, in advance, make the car warranty dependent on the consent. The question thus is whether the OEM can claim exclusive use or whether the vehicle owner can also assert user rights to the data? In addition, different situations apply. First of all, there is the difference between data accessible via the OBD-II port and the other vehicle data, which is not initially accessible. This includes the special position of the data which are important for warranty activities, which also (legally) may be conducted by parties other than the (representative of) the vehicle manufacturer.

2.3.2.1 Diagnostic data

Diagnostic data are limited to information which is necessary for the servicing of the vehicle. Through limited query rights to the CAN bus, the data are made available by the OEM to brand dealers and also to non-branded servicing companies. For the latter a legal obligation[4] has been established. Querying consists of requests being made to the system.
The system provides answers to these questions in the form of data. This way all companies are granted a limited users right in the context of servicing of the vehicle. For a reimbursement, the servicing company thus receives a limited users right with regards to the data for the specific servicing and repairs objective.

2.3.2.2 Other vehicle data

Other data besides diagnostic data are, in principle, not freely available. Nonetheless, some companies are able to obtain rough data from the CAN bus, so that more information becomes available than that which is provided by the OEM. These data, however, must undergo a drastic process to lead to accessible information. Regardless of the fact that these data can also contain personal data, the question exists whether the control over these data shouldn't also lie (partly) with the owner/user of the car.[5] Currently a study on behalf of the EU Commission (DG MOVE) is being conducted in which various models for the control of vehicle data are investigated. Especially the difference in interests between OEMs on the one hand and consumers and aftermarket parties on the other hand will be further examined in response to the discussion on this issue between the European Automobile Manufacturers' Association (ACEA) and the International Automobile Federation (FIA), which is more representative of the users.[6] The results of the survey, in which several models are proposed, can facilitate a further development of the data control of the vehicle.

2.3.3 Liability

Another aspect which plays a large role in the protection of vehicle data by the OEMs is the risk of third parties querying the CAN bus. In this respect, it has already been indicated that OEMs want to limit external influences on the CAN bus as much as possible, in connection with product liability. The CAN bus data contains data of many vital functions of the vehicle. The CAN bus data is indispensable for driving these functions. Because the OEM guarantees the integrity of the vehicle data, and an error in the data on the CAN bus can lead to the modification and failure of vehicle functions, there are few or only limited querying rights granted for the CAN bus. Querying is only possible on the secure and protected OBD-II port and within the framework such as that described in 2.1.1.1. However, the CAN bus can be listened to, whereby the integrity of the data is not jeopardised. The data flow to be listened to still requires much processing, though. Some specialised aftermarket companies are pursuing this (see also 2.3.2.2).

[4] European regulations are applicable here; the so-called EURO V/VI regulation with respect to Repair & Maintenance Information.
[5] Volvo has recently commented on this. The German auto industry does not seem to be ready yet.
[6] Research by TRL 'In-vehicle data and resources survey' 2016-2017.

3 Preliminary conclusions

What does the preceding inventory tell us, specifically? What is the expectation for the short term? The picture within the workgroup is that the quantity of data in vehicles is large and growing. For now, the OEMs appear to be unprepared to go further than the free issuance of data which can be obtained from the vehicle via the OBD-II port. However, this required data stream is too limited for many intelligent transportation systems (ITS) applications.

In the short term, the CAN bus vehicle data appear to be more broadly accessible only through aftermarket companies who collect and decode the data from the CAN bus. Currently, numerous vehicle-related services are being introduced on the market using this data. The next question is whether, and to what extent, this can be scaled up? The workgroup has observed that the available aftermarket provisions can be used to scale up data, provided that this occurs in small steps and especially with a limited number of data types, as determined from the market demand, and with qualitatively reliable systems. Often a valuable service can be created with, for example, only three data elements. Moreover, there must be a business case. To move from supply-driven (100 data elements available) to demand-driven (three data elements required), it appears that a bottom-up approach is still the appropriate way, since not everything can be arranged at once ahead of time.

What are the obstacles, and should we invest in removing them? According to the workgroup, it is especially the focus on large steps that can become an obstacle. In fact, the obtainment and use can only be launched with appealing use cases. The workgroup expects that the most successful applications will start with a limited data set and a limited users group, which, for that matter, can also quickly grow. The possible lack of willingness on the part of the OEMs to make data available, and the lack of ability to organise the obtainment of data in other ways, is still a limitation. At this moment, it seems that data use within ITS will be primarily taken up by the aftermarket, with CAN bus data or via other devices such as On Board Units (OBUs) and smartphones. These applications can perhaps serve as the driving force for Smart Mobility, as soon as it is clear that new applications actually work. Problems may also be expected in the battle for data ownership. The TRL survey currently underway for the EU should offer building blocks for this discussion. Moreover, a share of the data can be designated to be personal data, for example, because they are associated with a vehicle identification. This can serously hamper the usability of the data for other applications.

Likewise in the European context, the issue of data from vehicles with regards to who can access them and what can/may be done with them is a very 'hot' subject among the private (sector) parties. From the point of view of the automobilists, ownership is claimed which must be granted by the OEMs. Within the automotive sector, the discussion mainly takes place between the European umbrella organisations ACEA (for OEMs), CLEPA (European Association of Automotive Suppliers), the FIA (Federation Internationale de l'Automobile, supporting automobilists/consumers) and the FIGEFA (Automotive Aftermarket Distributors, on behalf of aftermarket parties), as well as parties associated with these organisations. The solutions supported by parties range from the creation of interim agreements (for example, a neutral server with a selection of vehicle data for third parties) to the desired legislation and regulations.

Finally, a point of attention is the quality of the available data. If the data are of insufficient quality, it will be more difficult to create a successful application. The project recently launched by the University of Applied Sciences at Arnhem/Nijmegen (HAN) called 'VIA-NOVA' aims to make a significant contribution in this regard. Data management can also become an issue in the sense that the managing party must be sufficiently reliable in order to successfully take on the management role. Indeed, under the current circumstances it seems that any hint of bias can bring the data streams to an immidiate standstill.

-0-0-0-0-0-0-0-