FESTA

D2.2: Data Requirements for FOT Methodology

21 May 2008

Grant agreement no.: 214853
Workpackage: WP2.2 Data Requirements
Tasks: T2.2.1 Data Acquisition System Requirements
T2.2.2 Data Storage Requirements
T2.2.3 Data Analysis Requirements
Deliverable no.: D2.2
Document title: Data requirements for FOT Methodology
Deliverable nature: Public
Deliverable status: Final to consortium leader

Consortium:

Centro Ricerche Fiat, University of Leeds, BMW Forschung und Technik GmbH, Daimler AG, Gie Recherches et etudes PSA Renault, Volvo Car Corporation, Volvo Technology Corporation, Robert Bosch GmbH, A.D.C. Automotive Distance Control Systems GmbH, Delphi France SAS, Loughborough University, Chalmers University of Technology, Institut National de Recherche sur les Transports et leur Sécurité INRETS, Statens Väg- och Transportforskningsinstitut VTI, Nederlandse Organisatie voor toegepast-natuurwetenschappelijk onderzoek TNO, Bundesanstalt fuer Strassenwesen BASt, Valtion Teknillinen Tutkimuskeskus VTT, INFOBLU SPA, Orange France, European Road Transport Telematics Implementation Coordination Organisation ERTICO, Universitaet zu Koeln

Disclaimer:

The FESTA Support Action has been funded by the European Commission DG Information Society and Media in the 7th Framework Programme. The content of this publication is the sole responsibility of the project partners listed herein and does not necessarily represent the view of the European Commission or its services.

List of abbreviations

AC	Alternating current
ACC	Adaptive cruise control
AD	Aftermarket device
BLOB	Binary large object
CAN	Controller area network (Network protocol and bus standard used mainly in automotive applications.)
DAQ	Data acquisition
DAS	Data acquisition system
DC	Direct current
DGPS	Differential global positioning system
ECU	Electronic control unit
EDR	Event data recorder
EMC	Electromagnetic compatibility
FCW	Forward collision warning
FlexRay	Automotive network communications protocol and bus standard, high speed.
FOT	Field operation test
GIS	Geographical information system
GPS	Global positioning system
GUI	Graphical user interface
JPEG	Joint photographic experts group (also image compression standard)
LDW	Lane departure warning
LED	Light emitting diode
LIDAR	Light detection and rating (optical remote sensing technology)
LIN	Local interconnect network (Network protocol and bus standard used mainly in automotive applications.)
MJPEG	Motion JPEG
MOST	Media oriented systems transport (Network protocol and bus standard used mainly for media components in automotive applications.)
MPEG	Moving picture experts group
MPEG-4	MPEG standard, version 4. (Set of methods and standards for audio and video compression.)
NAS	Network attached storage
ND	Nomadic device
NTP	Network time protocol
OEM	Original equipment manufacturer (usually vehicle manufacturer in this context)
RAID	Redundant arrays of independent disks
SAN	Storage area network
SMS	Short message service
SQL	Structured query language
SSD	Solid state drive (hard drive without moving parts)
UTC	Universal time, coordinated
WLAN	Wireless local area network

Table of contents

List of figures

List of tables

1 Executive summary

This deliverable will define the requirements for data handling (quantitative and qualitative) throughout the entire "methodology chain" during an FOT:

- Data acquisition
- Data upload and quality assurance
- Database storage and data management
- Data analysis

Valuable input for this work has come from international sources of FOT experience such as the University of Michigan Transportation Research Institute (UMTRI), Virginia Tech Transportation Institute, the Volpe centre, and others.

Results presented here include data acquisition system (DAS) requirements. Different types of logging needs (for different purposes) will be considered, such as vehicle systems testing, real-time traffic information functions, and vehicle-to-infrastructure solutions.

Experiences from state-of-the-art FOTs on DAS covered: sensors, video logging, vehicle dynamics, vehicle network logging, radar, driver behaviour, system performance, and qualitative data (interviews, questionnaires, and background information). Background information on involved concepts, as well as suggestions on how to arrive at FOT specific requirements, are provided. Aspects of OEM and systems provider co-operation and proprietary issues are handled, as well as issues with personal integrity and privacy.

Further, technical information on data storage, database design, quality management procedures, and suggestions for data analysis tools is provided.

2 Data flow and structuring

This chapter introduces the basic data structure and data flows defined and used in these guidelines and recommendations for an FOT study. In addition to this chapter and the abbreviations (above), the FESTA Glossary is a recommended reference for other common terms and definitions.

2.1 Data management flow and structuring

In Figure 1 the structuring and naming convention used in this document is visualised. This is an example of an FOT structure of data that includes data from an electronic data acquisition system (e.g. on-vehicle, infrastructure or nomadic device, but also services such as geographical, traffic, and weather information), as well as subjective data collection. The Data Acquisition Unit (on the right) comprises sensor systems requiring raw data decoding. The raw data may then be pre-processed, in this case by low-level data processing such as simple filtering or calculation of directly derived results. Both raw data and pre-processed data (derived from raw data) are then stored in the same format and may be used for Performance Indicator calculations done onboard the vehicle. This data is stored locally and can be kept locally for a shorter or longer period of time (until batched wireless uploads or disc pickup). If deemed necessary, smaller amounts of data can be directly uploaded to the main storage location. At one time or another, the data will be moved from the Data Acquisition System (DAS) to the main storage.

Before, in parallel, or after the DAS collection and upload of data, acquisition of subjective data may be performed. Subjective data is also considered acquired from a "sensor" (see picture on left). This data is then similarly subject to manual or automatic decoding, stored directly in the database (pre-processed or not), or used in Performance Indicator calculations.

Note that the "measures" (virtual interfaces) boxes indicate that all data *below* these boxes is to be perceived in a common and generalised way. Functions *above* these virtual interfaces should not see a difference.

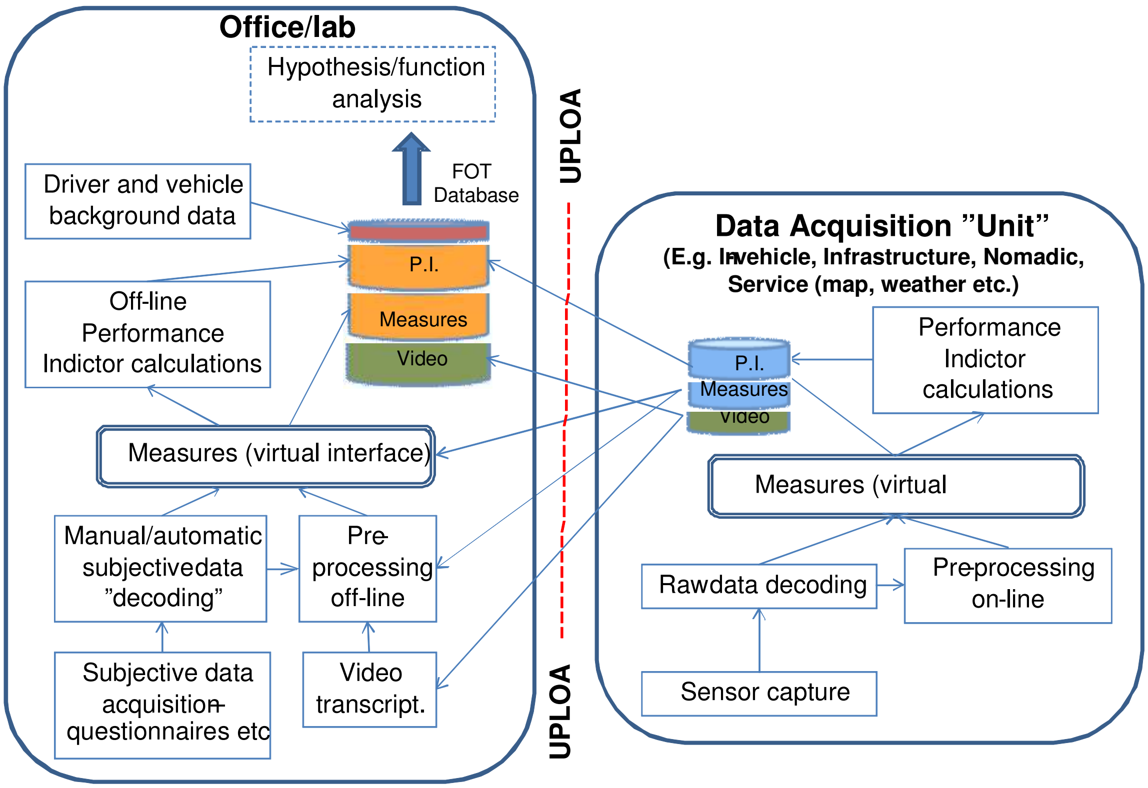


Figure 1: Data structuring

2.2 The measures and sensors tables

FESTA introduces the FESTA matrix, a spread sheet in Excel format containing three tables: “Performance Indicators”, “Measures”, and “Sensors”. The three tables may in a later stage be utilised to create a relational database. The recommendation is that the sensor table is used in preparation of, during, and after an FOT project, in conjunction with the other two related tables. Tables are provided in the FESTA deliverable D2.1.

The performance indicators will use input from measurements. In turn, the measurements are derived from sensors and logging devices installed in the vehicles, and from off-line measurements (questionnaires, interviews, etc.). The sensors to which the measures point can be found in the “Sensors” table (see the FESTA deliverable D2.1 appendices).

The “Sensors” table can be used to collect information about specific sensors, and thus to summarise the requirement specifications for all needed equipment.

The appended FESTA matrix is intended as an aid and gives examples of commonly used sensors, together with specifications, as well as example use. It is not exhaustive and future users are encouraged to modify, expand, or limit the matrix to suit the particular FOT.

Each table contains a key, which is used to identify the connections between sensors, measures, and performance indicators. A database will need this information to define relations and associations between the tables. Therefore it is important to make sure that the appropriate keys are entered when new performance indicators, measures, or sensors are listed. The key list must be complete, and the logical operators must be correct. Properly handled and thoroughly implemented the tables are valuable tools for data structuring and for data requirements specifications.

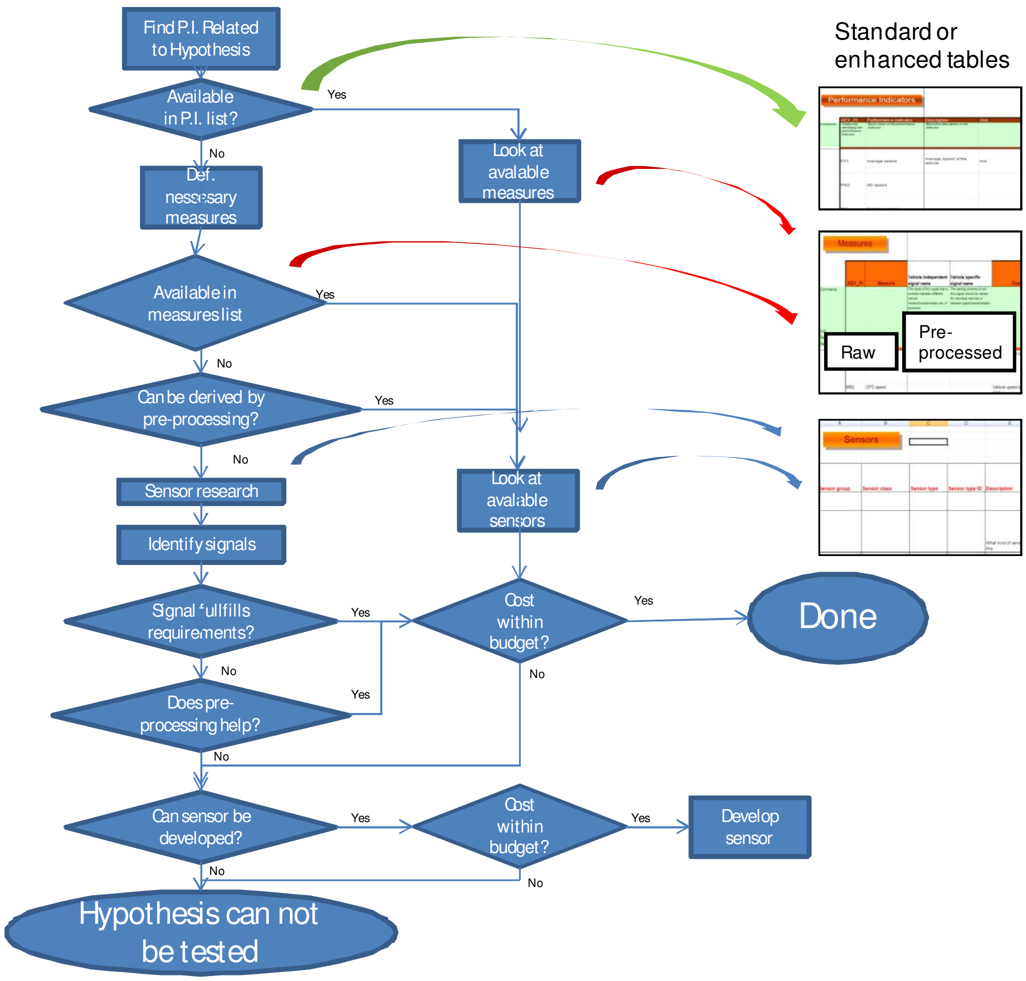


Figure 2: Usage of measures and sensors tables

3 Data acquisition systems and sensors

In FOTs there are many different methods of data acquisition. These include methods to collect different kinds of background data, as well as digitally acquired data from different types of sensors, and subjective data (such as data acquired from questionnaires). As for sensors, in FESTA *all the data sources mentioned above are considered sensors*. In addition to these, there is also data in the form of manually or automatically transcribed data, and reductions of collected data that can be considered acquired data (but with a manual sensor – the analyst). Subsequently all data can be acquired, stored, and processed in a generalised way.

All of these different types of data are used to test the hypothesis defined for the specific FOT. The data to be collected should be defined and based on research questions and hypotheses.

3.1 Data acquisition systems and methods

3.1.1 Background data acquisition

General aspects

The background data about the driver is crucial in conducting any FOT. The data itself needs to be collected and integrated in the driver interaction procedure. Due to privacy issues part of the background data may not be suitable for storage in a database, while other data can be stored and used directly in statistical as well as other forms of analyses. (See the FESTA Handbook.)

Driver background data specifics

Driver background data concern information on demographic, physical and cognitive issues, driver experience issues and attitudes. The following are some examples of driver background data that may be collected during the data collection process:

Demographic issues

- age
- sex
- education
- occupation

Physical and cognitive issues

- visual acuity
- hearing levels
- prior health problems
- personality
- processing speed, vigilance

Driver experience issues

- driving history, driving violations, accident history, type of accidents
- driving experience in years
- type of driving, e.g., private versus professional driving
- distances driven per week or month or year
- type of roads, e.g. mainly city versus mainly motorways

- purpose of the driving, e.g. functional or leisure.

Driving style
- driving style
- driver's level of aggressive driving behaviour

Attitudes
- attitudes towards driving
- attitudes towards new technology in general
- attitudes towards safety devices/safety systems in particular

Data should be collected pre-test, and could be gathered by means of interviews and/or questionnaires, by different tests, e.g. different performance tests, or by specific instruments, e.g., DULA dangerous driving index. For consistency with other data, the driver background information should be considered as acquired from a sensor, so that it can be added into the database and to the FOT's implementation of the FESTA sensor matrix.

3.1.2 On-vehicle data acquisition

An on-vehicle Data Acquisition System (DAS) is needed in FOTs where the focus is to study on-vehicle systems by collecting data from the systems in the vehicle. That is, if a study is not purely subjective (only interviews and/or questionnaires or direct observation), data about the systems to be tested needs to be collected and put into context by collecting other information about the driver interaction with the vehicle and the surroundings. The DAS used can be quite different in different studies and a specific solution cannot be recommended for all types of FOTs. In Table 1 a few different DAS solutions are described:

Table 1: Different DAS solutions

On-vehicle DAS Type	Comments
Fixed on-vehicle DAS with limited wireless transfer	The DAS system is fixed and integrated in the vehicle and can collect data from integrated on-vehicle sources, as well as with connected Nomadic devices and data received from the infrastructure. Parts of the data can then be sent wirelessly to a storage location from where this limited data can be used in analysis of data and system status/quality. This approach requires data retrieval at certain intervals (see section 3.5). The retrieval may include both site based high bandwidth data retrieval as well as data disc pickups.
Fixed on-vehicle DAS with full wireless transfer	Same as above but with the possibility to send the entire dataset from a trip wirelessly to the storage. This requires data integrity checks and validation procedures (see section 3.5).

Nomadic Device based DAS with full wireless transfer	The nomadic device is the data storage device and acquisition device with some kind of connectivity to on-vehicle systems such as CAN. The device can be used as a logger, with GPS providing absolute time and position. As an example, the nomadic device can temporary store the information, and then transfer the data to a storage unit that can be on-board or remote using wireless connections. Also, the nomadic device can be used just as a tool for providing absolute time and position, while other systems are used for logging. Finally, the nomadic device can be used to continuously send summary information (vehicle speed, position, driver assistance system activated) about the tests to a central station where the FOT can be monitored in real-time.

Most of the guidelines and requirements in this document are primarily written based on FOTs and experiences using some kind of on-vehicle data acquisition.

3.1.3 Nomadic devices

General aspects

A nomadic device is defined as “a handheld wireless device such as a PDA or Smartphone”, that is a portable and small system. A nomadic device or an aftermarket device (AD) positioned in the vehicle (in the text both denoted ND) could play two roles in an FOT. The ND could be part of the function/system under test or it could be part of the data acquisition systems, and acquire specific FOT data.

It is important to analyse both roles. When a ND is part of the system being investigated in the FOT a reliable and secure capture (logging) of the functionality of the ND being used must be secured. Very often Bluetooth communication links can be obtained to and from the ND and other data capture devices like “black boxes” (for instance today used to monitor driver behaviour such as speed, position, etc for insurance purposes) will often have access to Bluetooth links as well (and possible other wireless interfaces). In this case the risk of influencing and distorting the performance of the ND function/system can be minimised.

An ND can also be used as a DAS for the testing of other FOTs. For “objective” testing, it is often necessary to have a Data Acquisition System that can be easily installed on any vehicle, without the need to have specific adaptations to the vehicle both on the SW interface side and on the mechanical installation side.

For this reason, it is expected that ND (but rather specialised ADs as mentioned above) may play an important role as DAS for FOTs. Nowadays these systems may include the following features that are requested for DAS:

- connection to GPS signal (used for portable navigation devices)
- connection to GSM/GPRS/3G/…

- cameras (front and back)
- replaceable memory capacity (Compact Flash)
- local connectivity (Bluetooth or Wi-Fi)
- HMI, for subject input (buttons, touch screen)
- accelerometers (in some cases)
- other possible expansions

When there is not a high demand on video resolution or image refresh rate, nomadic devices can also be used as tools to store data as they are easy to install and can be used in different kinds of vehicles. If the vehicle has a dedicated gateway for a ND, this option can be used for further logging of vehicle related data. Capturing vehicle manufacturers' proprietary data can thus be avoided.

Using the local connectivity, Wi-Fi for example, memory could be expanded in ND, with high capacity hard disks. As long distance connectivity can be very important to monitor the FOTs and to guarantee that the function/system under test in the FOT is working properly, short samples of data can be sent using GSM/GPRS/3G.

A possible drawback of a ND, when used as a DAS in itself, is that test subjects must remember to bring the ND to the vehicle every time he/she uses the vehicle. A fixed (and hidden) installation of an aftermarket device (the black box) could then be seen as an interesting option. However, normally the ND to be used will be part of the test subjects "normal life" and the risk of leaving it behind can be seen as minimal.

Sensing with nomadic devices

Nomadic devices (and some aftermarket devices) can be used as sensors inside an FOT by a dedicated data logging SW running on the device itself. This SW should be able to collect data from the sensors connected to the device and to store the data during the time necessary. This operation can be done periodically or can be triggered by certain in advance defined events.

Main sensor information that can be collected by a ND is described in the following list:

- Absolute time, from GPS
- Vehicle position (latitude, longitude, height), from GPS
- Vehicle speed, from GPS
- Vehicle heading (direction of motion), from GPS
- Vehicle position on a digital map, from GPS and map matching
- Accelerometer data (in all directions), from "black box" applications
- Road/traffic image and/or driver image, from internal cameras
- Other sensor data (from other available ports)

With specific software the ND (and AD) could also be used to: detect vehicle position compared with road lane, detect distance from preceding vehicle, and detect traffic signals, level of traffic, etc. or in some cases detect the "jerks" of the vehicle in motion. These could be used as "event trigger" for data acquisition.

In particular, an important effect on driver behaviour is vehicle speed, and in such cases ND devices can be used to measure changes of driver behaviour in speed taking into account the road travelled (derived from absolute position) and the time of the day.

Data so acquired can be stored on board and also sent to:

- a local storage media (on board)
- a ground station for data storage, if compatible with communication costs
- a ground station, to monitor the quality of acquired data and diagnose problems.

An example of a "black box" functionality, needed to capture all the data mentioned above but also (in case the ND will provide the functions to be tested in the FOT) the actual use of the ND to be tested by the subjects, is listed in the following (extracted from an specification document of a "black box" manufacturer):

- GPS: Enables vehicle tracking using coordinates read by GPS (latitude and longitude) and displayed on a mapping system
- GSM/GPRS: Supports the transmission of insurance data from the "black box" to a central data handling centre (for software updates and vehicle tracking)
- Accelerometer: Enables detection of a crashes and reconstruction of the dynamics before and after the event itself.
- Bluetooth: "Vehicle Hands Free" operation: a Bluetooth mobile phone can be paired up to the hands free system fitted on the vehicle. "Data" operation: lets the handheld PC/PDA be used in conjunction with the system's GPS, starting up the mapping navigator on the handheld PC.
- Software: Enables management, processing, recording and transmission of data; it also handles the control of self-diagnosis.
- Installation: Easy, fast and non-invasive on the vehicle's interior and its on-board technology.

3.1.4 Subjective data acquisition

General aspects

As explained before, also subjective data is considered as "sensor" data in the scope of the FOT methodology. All the subjective data will be considered and stored at the logical level as if it were collected from a "real" sensor.

Subjective data include data acquired in a range of ways, e.g., via interviews and questionnaires. Sensors that measure parameters related to the subjects but with physical sensors, such as eye-gaze direction or "objective" driving behaviour (e.g. speed or acceleration), will not be considered here but in the 'On-vehicle data acquisition' and in the 'Specific sensors in FOTs' sections.

Further, subjective data could be also considered the input from the participants about an "event to be considered". A specific HMI component, e.g. an annotation button, can be used by the subject to mark when something of interest has just happened. Further data on the event can be gathered by the participant tape-recording a description of the event, or by filling out a specific questionnaire, or by responding to a telephone interview triggered by the annotation.

In FOTs where only triggered storage is used and where the triggers are partly based on the drivers pressing the annotation button, it is important that data preceding the event (ring buffer) is recorded. The size of the ring buffer will differ between projects.

Subjective data should preferably be stored in an electronic format (e.g. text documents, specific data files related to answers to a questionnaire, data directly uploaded into a database from web-surveys) as well as other means of digital storage.

As for the time of data collection of subjective data, typically the subject performing a test is given some questionnaires to be filled out before, during, or after the test. For subjective data, see the FESTA measures and sensors matrix.

Besides the questionnaire or interview data, related information needs to be filled in by the person(s) administrating the study (e.g., day and time of the test start and end; identification key for the subject; reference to the objective data). In addition to this, data filled in by e.g. an observer of the video data (e.g., weather condition; traffic density) is considered as subjective data.

The following table summarises required and optional information.

Table 2: Information related to subjective data to be stored

Required information	Date and time (hh.mm) of test start
	Date and time (hh:mm) of test end
	Subject ID code
	If present, reference to objective data (file name, location)
Optional information	Time zone (for synchronisation with GPS UTC time)
	Traffic condition
	Weather condition
	Light condition (daylight, dark)
	Special events
	Notes

Subjective data could be collected by a large sample of subjects, since no specific intervention on the subject vehicle is required. Using questionnaires, thousands of subjects could be considered, and this could be a method to:

- prepare "objective" field tests (e.g. selection of subjects with respect to their attitudes towards new technology)
- understand the major issues while using a specific function;
- confirm the results of a small sample of subjects using equipped vehicles;

This suggests that subjective testing could be more beneficial if not used at the same time with the "objective" field test phase, but rather before and after objective tests. When subjective field tests last several months, this could be done by "extending" the duration of "subjective" tests, starting before and finishing after the "objective" FOTs.

Questionnaires and interviews

Questionnaire and interview information is usually gathered from experimenters that are conducting the interview or handing out the questionnaire (on paper or on a computer) to the subject. However, questionnaires can also be distributed by post or realised as a web based survey. In the latter case, the data are stored directly on a data file. One drawback with postal and web based questionnaires is the level of controllability (e.g. is the intended subject the one answering the questions? and does s/he understand the questions?).

When the subject is writing the answers to the questionnaire on paper or he/she is interviewed, the answers (or a summary of main points) have to be transferred by somebody into an electronic format for storage in for example a database.

Subjects may be asked questions on topics such as: *how did they feel when driving with and without the system*, *if they like it*, *how they judge the interface*, *if they were disturbed or pleased by the system*, *if they felt more or less safe*, *if there are situations in which the system failed or behaved in a "considered" anomalous way*. Thus, the subject is in fact used as a sensor to detect the physiological aspects of the system. Subjective data provide unique information on, e.g., the subjects' attitudes towards different systems, important to understand in relation to, e.g., adoption and acceptance of the systems. Subjective data also provide important complementary information to other types of information collected in an FOT. The assessment of the subjects is usually compared with the actual driving behaviour measured in quantitative terms, e.g. by using indicators such as speed or distance to the vehicle in front. In this way it is possible to compare the assessments of the driver regarding the increase or decrease of safety with the effective changes in the driver behaviour introduced by the system.

Using subjective data collection and tests can produce useful information for setting up "objective" tests, or to confirm their results using a larger platform of subjects.

3.1.5 Real time observation

General aspects

In this context, real time observation data is data collected by an observer that directly or indirectly is observing the drivers and systems to be evaluated. Direct observation refers to a situation when a person, the observer, is present in the vehicle while collecting data. Indirect observation refers to a situation when the observer observes and analyses video-recordings of driver behaviour etc. The data acquisition process is usually relatively manual but the results can (and preferably should) be transferred to a digital format and uploaded to the FOT database for further analysis.

Real time observation data help provide a more detailed picture of a driver's behaviour, as well as verifying the information gathered by other instruments. Observations can, e.g., gather information that is not possible to acquire through interviews or questionnaires, in particular retrospective ones. This information includes data on actions, reactions, interactions, and compensating behaviours of which the individual is not aware. Furthermore, direct observations allow the observer to collect not only information on the driver and his/her driving but also on the context in which the driving takes place. Observations cannot, on the other hand, collect information on the driver's interpretation of a certain situation or how the driver experienced the situation. In this case, data must be collected by means of questionnaires or interviews.

There are some important differences between direct and indirect observations that need to be pointed out in relation to an FOT. *Direct* real time observation studies can be combined with the collection of verbal protocols, that is data that is collected by "a think aloud procedure". In addition, direct real time observation can allow the observer to pose questions to the person(s) being observed on, e.g., explanations for a certain behaviour. However, the overall purpose of an FOT is to collect information on as a natural driving as possible. With an observer in the vehicle, there is always the risk of the driver not acting the way he or she would otherwise. In particular, this could be the case if the specific study also includes the collection of verbal protocols and/or questions being posed. Direct real time observations must therefore be carried out with great care and as unobtrusively as possible. Time must also be considered a key issue, as the driver may initially respond to the observer being present but the effect may wear off somewhat after a time period. Nevertheless, if direct observations are to be carried out as part of the FOT they should preferably be carried out towards the end of the test period. Indirect observations, on the other hand, rely on data collected by on-vehicle (video) cameras, which means that the observations per se do not disturb the driver even though the camera(s) may at some point.

Indirect observations are used in most state-of-the-art FOTs while direct observations are more common in more traditional experimental or quasiexperimental studies.

Observation in practice

Observations, whether direct or indirect, could be more or less structured to their character. The observer could observe the driver driving over a certain period of time, documenting behaviours, reactions, situations, conditions, as they occur. The data collected is documented as descriptions but with no "pre-processing" of the data. Such a "grounded approach" may detect the less anticipated. However, structure facilitates data collection and offers some degree of pre-processing of data. It also facilitates comparisons across test participants and/or different systems. Structure is then provided by different types of protocols used in order to document the observations, so called *observation protocols.*

An observation protocol is commonly shaped around activities or behaviours and time, i.e., describe the sequence of "activities"/behaviours as they occur in time and the duration thereof, e.g., the use of the rear mirror or the use of different infotainment systems in the vehicle. In order to facilitate collection as well as analysis, different activities could be pre-coded, e.g., a letter indicating the type of device used (A="rear mirror", B="GPS") and a number describing the specific type of use or interaction (1="manipulation", 2= …..), etc. The protocols are designed before the FOT and should reflect the information that is of interest for the particular FOT. These protocols could be realised in computer for an easier transfer to a database.

min	1	2	3	4	5	6	7	8	9	10
10										
20										
30										
40										
50										
60										

Figure 3: Example observation protocol to be used for registration of activities during one hour.

Observation protocols can also be based on, e.g., rating the drivers level of alertness etc. Analysts may view video data for one minute prior to a particular event, e.g. an accident or near accident, and then give a rating on a scale, e.g. a Likert scale. Reliability can be enhanced by using validated instruments. Examples of instruments that can be used are

- ORD;
- PERCLOS (percentage of time that eyes were closed 80% or more);
- AVECLOS (mean percent eye closure).

Depending on the particular FOT, the type of drivers involved, etc., the duration of a real time observation and data collection varies.

If direct observation is used, the duration of each observation, i.e. data collection period, depends upon the situation. Observation carried out in city traffic and the rush hours may have to be limited to 20 minutes at a time while observations carried out on a non busy motorway may last for 1 hour or more. Other approaches are, e.g., observations and notations for 20 minutes per hour over a total period of a set number of hours. A data collection sequence can also be triggered by an event or a situation, e.g. when the driver is approaching a crossing or a roundabout.

3.1.6 Acquisition of infrastructure data and other service data

General aspects

Information from and about the infrastructure can play important parts in an FOT. Sensors in the infrastructure (e.g. radar, laser, cameras, and loops in road surface) can be used to collect important data such as information about weather condition, traffic flow, speed and throughput in every lane etc. Also visibility sensors are used on the road side to detect the visibility range.

In this case, the information detected by the infrastructure should be collected in raw format or in an aggregated form. If data are collected both on the vehicles and on the infrastructure, it is necessary to have synchronisation between the two sets of data.

There can also be an exchange of information between the infrastructure and the vehicle (e.g. for the driver support function or as part of the FOT methodology). Data can then be stored in the vehicle and/or directly from the infrastructure.

Infrastructure sensing
From the infrastructure, mainly the following kind of data can be available:

- Environmental data (e.g. temperature, weather)
- Traffic status (from radar or loops)
- Visibility range

In fact, when this data is used for traffic control, it can also be made available for FOTs.

In many countries it is required to contact local road authorities before installation of equipment close to a road. Working close to or on roads may (depending on country) require special training or licensing. In some countries it is even required to use a special company or local road authorities for any installation work close to or on roads.

Services
In this context services include all kinds of information gathered by an external source (non subject vehicle) and transmitted to the subject vehicles for some reason. Examples are real time traffic flow information and information about dynamic speed limits. When using such sources it is recommended for traceability (during and after the project ends) to record information about for example version of service, update rates and resolution/precision of the information they have during the duration of the study. It is also recommended to invite the service providers for discussions and possible partnership in the FOT.

3.2 Specific sensors in FOTs

3.2.1 Internal vehicle bus data (CAN, LIN, MOST, etc.)

Most modern vehicle manufacturers features one or several internal vehicle networks such as CAN (Controller Area Network), LIN (Local Interconnect Network), MOST (Media Oriented Systems Transport) or the upcoming FlexRay. An internal network may carry large amounts of useful information for the FOT. For instance, typical parameters sent over the CAN are steering wheel motion, pedal usage, vehicle speed, turn indicator as well as system-specific data from adaptive cruise control and frontal radar (when such systems are applicable). However, there are several concerns with accessing and ascertaining quality on data from an internal vehicle bus.

Accessing the vehicle-bus through standards
Accessing information from an internal vehicle-bus can be highly complex and even void warranty if it is done without the OEM’s permission and supervision. There are a number of standards available which open up for data access. Since 2002, a number of of European vehicle manufacturers, mainly of heavy vehicles, give third parties access to their CAN-bus through the Fleet Management Standard (FMS) [Ref: SAE

1939, ISO 11898, www.fms-standard.com]. The FMS enables access to a selection of CAN signals such as fuel consumption, engine speed or vehicle weight. A similar standard is applicable for busses and coaches [Ref: SAE J 1939, ISO 11898, www.bus-fms-standard.com]. There are also other open standards involving passenger cars such as the OBD standards (On-Board Diagnostics) enabling access to a limited set of vehicle parameters [Ref: Gilberto Geraldo, “Differences Between On-Board Diagnostic Systems (EOBD, OBD-II, OBD-BR1 and OBD-BR2)”, SAE publication no 2006-01-2671, November 2006). The benefits of these standards are easy access, standardised protocols for all models and avoided interference with the network. Anyhow, those standards are not always supplying the most relevant and interesting data for the FOT that is to be carried out.

Accessing the vehicle-bus through OEM cooperation

A description of the entire vehicle network will in most cases contain proprietary information from the OEM since it may reveal detailed information on specific functions and the vehicle system architecture. Thus, for an OEM to provide a network description a non-disclosure agreement is typically required and it can be hard to attain even with the OEM involved in the study.

An option to using the entire vehicle-bus description is that the OEM only provides the description of a selection of signals which enable access to the most important data. An example for the CAN bus would be to provide details on the protocol, IDs for frames carrying the data, bit-positions within the frames for the actual values and the decoding scheme – describing how to translate the data from bits to the actual sensor data with correct unit etc.

However limited the access to the vehicle-bus is, this might still be a proprietary issue. By using a Bus-gateway, the OEM will be able to extract data from the bus without providing any information about how the bus actually works. The function is simple: the gateway is programmed by the OEM to read certain information on the vehicle-bus, decode the information and then pass it on to the FOT logger equipment (either on a bus, e.g. CAN, through USB, TCP/IP or other connection). An illustration of the process is shown in Figure 4.

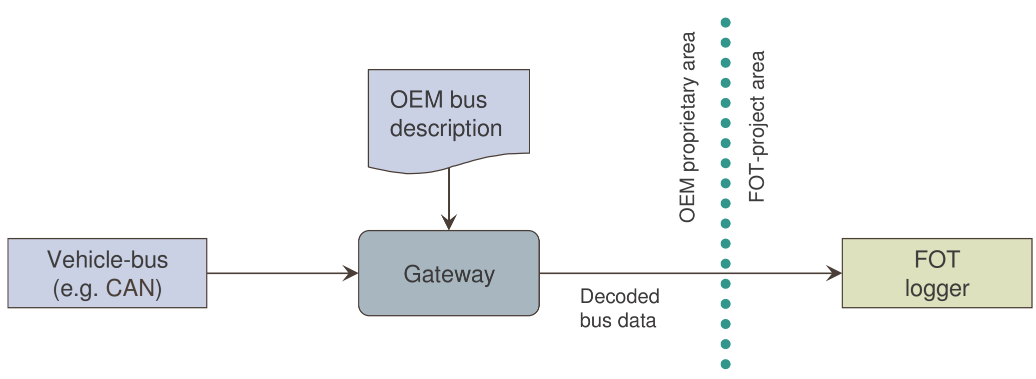


Figure 4: Illustration of process for securing proprietary vehicle-bus data using a gateway.

When physically connecting to the vehicle bus, it is important to follow applicable standards of the bus in order to prevent interference that may reduce network functionality (and thus warranty). A detailed description of the common CAN-bus is provided in reference [“CAN Specification Version 2.0”, Robert Bosch GmbH, D-70442 Stuttgart, 1991]. To interpret the commonly used CAN-bus description file format, CANdb (.dbc), refer to Vector Informatic GmbH [www.vector-worldwide.com]. It should also be stressed that every bus implemented by an OEM might not necessarily fulfil the standards in every detail.

Sensor specifications and details

When acquiring sensor data from a vehicle bus, the information is passed through several stages – as explained further in the upcoming example – before it can be read from the bus. These stages are likely to affect the signal value both in terms of amplitude and frequency and thus need to be closely observed.

In a typical scenario, the actual sensor, located and mounted somewhere in the vehicle, fetches a physical analogue signal with a certain accuracy and reliability. The analogue signal may then pass through an analogue filter to prevent aliasing when the values are sampled for conversion to digital values. Next to the effects of the filter, the analogue-to-digital conversion will determine the granularity of the digital signal. After sampling the sensor values at a certain frequency, either in the sensor itself or in the electrical control unit (ECU) to which it is connected, the ECU will encode the sensor data and pass it to the bus. This encoding might again affect the precision and depending on the bus speed the signal can be sent at a frequency different from the original sampling frequency at the sensor. A difference can cause both loss of data (too low frequency on the bus) and repetition of data (too high discrete message rates). Yet another issue is raised by the non-scheduled buses, e.g. the CAN-bus, where an ECU will need to wait for an open slot before sending its packets. This will cause an unpredictable delay in the signal and will confuse phase-domain signal estimates. Finally there can be additional delay and conversion effects from a gateway delivering the decoded signal.

The scenario outlined above clearly illustrates the multifaceted aspects of capturing sensor data through a vehicle bus. To ascertain that measures of good quality are attained, plenty of contribution is required from the OEM. A thorough description of a vehicle bus and its ECUs will in many cases require involvement by subcontractors as well. Thus, strong cooperation with OEMs is desirable when considering internal vehicle-bus recording. A list of required details for successful data acquisition from the frequently used CAN bus is provided in the list below.

Details required for accessing the vehicle CAN-bus:

- CAN protocol standard used (e.g. ISO 11898-2 or ISO 11898-3 or ISO 11992-1, ...)
- Baud rate (provided through protocol standard)
- Frame type (mainly length, provided through protocol standard)
- Name of the signal
- Source sensor location and mounting
- Source sensor actual frequency (prior to AD conversion)
- Source sensor granularity and reliability

- Frame ID for the frame carrying signal data
- Bit positions of the data within the frame and bit-order (Motorola/Intel, possibly provided through standard?).
- Conversion scheme from bits to signal value (typically factors for $k*x + m$ translation)
- Description of values for enumerate signals (e.g. gear lever position)
- Detailed description of the signal function when applicable
- Frequency typically used for sending signal data over the bus (estimation)

3.2.2 System function/status

A system under evaluation, such as an LDW, ACC or FCW, needs to be continuously monitored to ensure that it is operating properly. The system status signal will thus form a measure to be recorded in the data acquisition system. The status signal will depend on the specific system under evaluation and needs to be provided by the system manufacturer or vehicle manufacturer for integrated systems. Next to status reports, a system may be able to deliver data from its internal sensors. Details on available measures from a system are provided from the system specifications or from the OEM. The Data transfer interface will depend on the specific system, its mounting and possible vehicle integration. Some commonly used interfaces are CAN, LIN and Ethernet. Integrated vehicle systems are often accessed over an internal vehicle bus, requiring strong collaboration with the actual OEM for access.

3.2.3 On-vehicle video

Video is by most state-of-the-art FOT experts considered a very important data source for the identification of driver behaviour and reactions, as well as for the process of analysis to understand the underlying context with regards to the surrounding. When a certain situation or event has been identified for evaluation of a particular system, the video provides the analyst with information about the context of both driver behavioural aspects and the interaction with the environment (if external video is used).

Video can be collected for different reasons, but the two main ones in FOTs are:

1. Driver monitoring – Helps understanding the driver state and actions in the vehicle. The two main views here are firstly the driver eye/head movements in relation to the vehicle/environment/context and secondly the driver interaction with the vehicle and other driver actions (pedal, gear shift, steering wheel handling as well as use of mobile phone, eating, etc.)
2. Environment/contextual monitoring – Helps understanding the driving contexts by collecting information about the surrounding traffic. For a specific FOT the number of cameras needed, the needed resolution, the views captured by the cameras, etc., should be defined by what is needed to be able to address a hypothesis.

Since different FOTs will have vastly different objectives and requirements this document only outlines the methods for defining the requirements for video.

How to define requirements

When several sources of video (cameras) are used, it is recommended to assess the quality requirements for each particular view.

The requirements for a specific video view can be found by the following process:

1. Identify the situations (refer to the FESTA Handbook) that are relevant for a specific hypothesis.
2. Set up a test vehicle with a flexible digital recording system including a camera (or cameras), preferably with the possibility to change the camera lens and to change capturing frequency. The camera/recording device could perform compression, but there should be a “very good quality” mode which should mainly be used in this process. Thus, captured data will be stored without any visual degradation of the image.
3. Use the test setup to capture several minutes (~10 min) of real data. As the efficiency of most modern video compression techniques is affected by motion in the captured video, it is important to record video “live”, i.e. drive the test vehicle. If the hypothesis includes low light situations, such situations should also be included in the recording.
4. Investigate the available modes in the video capturing device. Available options may include settings for frame rate, bit rate, resolution/image size, specific compression parameters, de-interlacing, etc (see below). Make preliminary choices for a small number of configurations/capturing modes to be evaluated, using experiences from prior known FOTs and other known restrictions. (Example: Maximum frame rate at 10 fps, maximum bit rate, etc.)
5. Assess the maximum available storage capacity in the DAS, generally limited by disk space and the size of other logged data. (Example: 100 GB.)
6. Assess how many hours of video capture will be needed at maximum. (At what intervals will the DAS storage be replaced? What is the expected usage pattern of the logged vehicle?)
7. Use hardware (or software) to compress the captured (good quality) video into new video streams, using the configurations chosen above (see item 4). At least assess the “best” setting: highest possible video quality (minimum compression) while limiting maximum bit rate, chosen to satisfy the available disk space and number of recording hours. (Example: 60 GB available disk space and 90 hours of recording boils down to a total maximum bit rate of around 1500 kbit/s.)
8. Evaluate the results and set requirements accordingly (using items 5-6 above). The evaluation comprises both calculations of predicted storage needs for the different configurations, as well as evaluation of video image quality (using item 1 above). The intended goal is to be able to capture video data that will be of sufficiently good quality to be able to address the hypotheses, while still respecting the storage restrictions of the DAS. In this evaluation people from both study design/analysis planning as well as the analysis team should be included.
9. Evaluate the robustness and standards compliance of the resulting video, to ensure proper decoding and viewing of the results.
10. Thoroughly document all settings, methods and standards used.

Note again that it is highly recommended that the requirements on video are founded on the research questions and hypotheses.

Using an interlaced image camera/capturing technology (used in the PAL and NTSC standards) will produce vertical image jittering effects when capturing a moving scene. Without proper and effective de-interlacing, at compression this effect can produce a blurred image (somewhat out of focus). The use of progressive (non-interlaced) scan cameras (digital or analog) and hardware during the complete video capture chain is thus preferable.

Figure 5: Examples of an interlaced and a non-interlaced video frame.

Video compression and quality

Because of storage limitations, compression or limitation of captured video data is necessary, and is thus used in most state-of-the-art FOTs. A number of different video compression techniques are available, more or less suitable for each FOT's particular needs.

Either hardware based or software based real-time compression tools can be considered:

- *Hardware compression:* Takes load off CPU, but higher hardware cost.
- *Software compression:* More CPU load, higher software implementation costs, and usually more configurable. Compression optimised according to context more easily implemented. Possible synchronisation issues incurred by the higher CPU loads.

In general, more complex compression solution required higher loads on the hardware (either the DAS CPU or the frame grabber processor), thus generating more heat.

A (non-exhaustive) list of common and likely methods and standards is presented here (in order of increasing complexity):

- Uncompressed video
- Motion JPEG (MJPEG)
- MPEG-1 Part 2 Video. Used by Video CD, VCD. Highly compatible.
- MPEG-2. Generally digital television and DVDs.
- MPEG-4 Part 2. Used by popular codecs such as DivX, Xvid and QuickTime 6, and numerous hardware-based implementations.
- H.263. Low bitrate format intended primarily for video conferencing.
- MPEG-4 Part 10 (H.264 or MPEG-4 AVC). Used by QuickTime 7 and by HD DVD and Blu-ray.

Several other video compression techniques are also available. The above methods and standards each comprise different features that may or may not be employed by the tools used for compression. Thus, if the tools allow it, each standard can be implemented, configured and used in a vast variety of ways. Not all tools are created equally efficient or robust. For quality assurance feature set, standards compliance and robustness must be investigated and assessed before a final choice is made. Generally, the more complex standards are better at achieving good video quality at a low bit rate. However, in these cases higher costs related to data processing needs (i.e. CPU load *or* hardware cost) must be taken into account when designing the DAS.

Roughly described, the later MPEG standards make use of a compression technique that only sends full image picture frames occasionally, while intermediary frames only contain information changes from the surrounding frames. This technique greatly decreases the video stream size. At the same time it is important to consider that the stream size (i.e. the bit rate) increases with motion or change in the picture. To assess the quality and size of the resulting compressed video stream, it is necessary to test the setup of cameras and the tool configuration in “live” circumstances.

Generally the following configurable parameters affect the video compression and quality:

- Picture resolution (lines or pixels)
- Frame rate (fps)
- Colour settings (black/white or full colour)
- Inclusion only of “regions of interest” (sometimes configurable in frame grabber)
- Bit rate strategy settings (variable, fixed, hybrid)
- Bit rate limitations (kbit/s)
- “Quality” strategy settings (usually difficult to assess, but affects size)
- Other compression method related settings

Recommendations

Make a thorough analysis of video acquisition requirements. Set requirements necessary for each individual view, to possibly achieve a first limitation of video data. Choose a well-tested hardware video frame grabber/compression solution, and choose compression suitable for the FOT. The MPEG-4 part 2 and part 10 (H.264) compression feature sets have been used successfully in other FOTs. Furthermore, an important aspect to consider is that in most cases, the project will not have the budget nor the possibilities to collect everything wanted. Sooner or later compromises have to be made on what to collect (in terms of views, field of view, resolution, frequency, quality, colour/greyscale etc.).

3.2.4 Automatic on-vehicle driver monitoring

In addition to capturing video to try to identify the driver actions, behaviour and state, there are several other techniques to collect this data – often as complements to the video data.

Head/eye tracker

Two of the main issues for many systems are visual distraction and the effects the system has on the driver attending to traffic versus system. In most FOTs to date this has had to be manually coded by watching video – a very tedious and labour expensive procedure. By using eye or head-trackers in the vehicles more or less continuous data of some driver state/attention measures can be obtained (even though it will most likely not be perfect or fully continuous). The problems with head and eye trackers are mainly that there might/will be significant data dropouts due to limitations in driver head and gaze tracking.

State-of-the-art head/eye-tracker technology is relatively expensive, but weighing the benefit of getting more or less continuous information about both some driver state variables and indicators for on-road/off-road driver attention should not be underestimated. If possible it is recommended to use head or eye-tracking systems in FOTs where driver state is an issue (that is, in most FOTs).

Using an unobtrusive system is a requirement for head/eye-tracking systems for FOTs. The driver should not have to initiate or wear any device; the system should be as inconspicuous as possible to the driver. Also, the system should not require any manual calibration in the field. The reason being minimised driver awareness of the data collection to assure as natural driver behaviour as possible.

It should be noted that the Head-tracker technology potentially also offers Driver Identification to simplify just driver identification for each trip.

Other

Adding sensors that the driver has to put on and wear should be avoided (and can be considered a strong guideline) to assure as natural a driver behaviour as possible. There are some sensing technologies (in addition to head/eye-trackers) that can measure different driver variables. Some examples are passive alcohol sensors, seat or steering wheel integrated heart-rate sensors and RF sensors (indicating cell phone usage).

3.2.5 Audio

The continuous recording of audio in FOTs is considered a significant privacy issue. The primary purpose of using audio in state-of-the-art FOTs is to have a driver push button that the driver can push, after which audio capture is started for a predefined period of time (e.g. 1 minute) – allowing the driver to give audio feedback to a specific event. Other triggers can also initiate the audio capture (system warnings etc) but then the drivers are not initiating the capture themselves and this can be viewed as a privacy issue.

A sensor can be used to record sound pressure level or noise in the vehicle. For this kind of measurement there are generally no privacy issues to consider, as the sensor would measure only the sound pressure level with a rate of 10 Hz (for example).

Audio can also be part of the debriefing and interview process.

3.2.6 Extra analogue/digital data sources

In on-vehicle system FOTs there is often a need to acquire information that is not available on CAN or other vehicle bus systems: both digital and analogue information. An analogue/digital I/O device for data acquisition is thus required for these FOTs. Also, anti-aliasing issues need to be addressed.

Examples of extra analogue/digital data sources are: temperature sensors, vehicle battery voltage, and driver push buttons (annotations/audio recording initiation). RF sensors (indicating if there is a cell phone active) have been used in several studies but have been proven difficult to use to a massive amount of false positives.

The requirements for the different kinds of extra analogue and digital sensing need to be defined in the study design and will not be covered here.

3.2.7 *Radar and other non-video environment sensing*

In addition to video there are other sensors that can be used to capture the environment and traffic context that the driver is encountering. Depending on the FOT setup the sensors can either be in the infrastructure or in the vehicles. It can be part of the evaluated systems (or other already integrated sensors), or it can be sensors that need to be added/integrated to the vehicles/infrastructure and integrated into the data collection systems. Examples of this type of sensing are RADAR, LIDAR, Lane Tracking, Sign Recognition (any sensor that acquires information about the vehicle surrounding, this could also include sensors mounted in the infrastructure)

Already integrated sensors (by OEM/Road Administration)

If an environment sensor (required/requested from a hypothesis point of view) is integrated into the vehicle or the infrastructure, great effort should be spent on trying to add the sensor data to the used data acquisition system. There are several issues that have to be taken into account in this integration (possibly not applicable if an OEM is running the study internally or possibly together with non-competing companies):

1. OEM allowing/disallowing access to data: In many cases the low level sensor data from a system integrated for retail use is considered proprietary for the OEM and the system/sensor supplier. Thus, the OEM is likely reluctant or unwilling to share data on a low level. In collaboration with the OEM it is necessary to define the minimum level of data needed to answer the defined hypotheses. The OEM can then possibly pre-process the low level sensor data and provide aggregated higher level data that complies with the minimum requirements, without giving away low level sensor data. This process requires restricted access to the logger raw-data from others than the OEMs and potentially Non-Disclosure Agreements with Universities or Institutions providing support.
2. System interfaces: Much of the needed information from the integrated sensors is often available on CAN. Some data such as low level sensing information (e.g. object tracks from RADAR) may require special interfaces to be implemented both in hardware and software (in the DAS). Accessing the sensors at this low level usually requires supplier consent and special contracts. This may also be the case for data available on CAN.
3. System comparability: If the studied vehicles have different OEM integrated systems, they will provide different quality of data as well as different resolution, range, field of view etc. This will make it difficult to do a direct comparison of Performance Indicators (to answer a hypothesis) between these different vehicles, based on these sensors – even within a project. It should be noted that this is also true for any other sensors that are different between vehicles (for example most data derived from CAN).

Add-on environment sensing

If the hypothesis and performance indicators defined for a study requires information about the vehicle surrounding that cannot be retrieved from the integrated sensor, sensing needs to be added. The process of adding sensing is recommended to include at least the following:

1. For identification of what sensing to use, the FESTA sensor matrix can be used as a base, but a survey of new and improved sensing technologies should be made.
2. Hypothesis and performance indicator definitions must include the requirements on the external sensing – the FESTA sensor table can be used as a base but the requirements must be defined in detail for each specific study and sensor measure.
3. Investigate the issues of integrating found sensing
    a. Does the extra sensing require additional interfaces that are difficult to interface with to a DAS?
    b. Does it require information from vehicle buses (for example speed and yaw-rate from the CAN bus) – if it does, for example a CAN-gateway might be necessary.
    c. Does it draw significant level of power from the battery?
    d. Will it boot and shut down properly with the available power logics?
    e. How can the sensor be integrated without significant effort (for example RADAR must have a plastic cover (non-metal), making integration of for example side looking RADAR more difficult than front or rear looking RADAR)
    f. Does it require repeated calibration?
    g. Does it comply with laws and regulations (for example emissions)? Also consider international driving (since in many FOTs may driver to other countries).

How to define requirements

Since these sensors can vary widely in their specifications it is difficult to give general guidelines to how to define the, but for example RADAR and other object tracking sensing systems, required Field Of View, radial and angular resolution/precision (azimuth) is important to define based on the hypothesis.

3.2.8 GPS

The Global Positioning System (GPS) is a satellite-based navigation system. GPS receivers are using the information gathered from a number of satellites to measuring the absolute position, velocity and heading and a number of measures for quality estimation of the collected data. Also, the GPS includes a *GPS Time* reference time stamp and a difference between the GPS time and UTC. This is useful for synchronisation within a data acquisition unit as well as between systems.

Differential GPS

Differential GPS (DGPS) is a technique for reducing the error in GPS-derived positions by using additional data from a reference GPS receiver at a known position. The most common form of DGPS involves determining the combined effects of navigation message ephemeris and satellite clock errors (including propagation delays) at a reference station and transmitting corrections in real time to a user's receiver, which applies the corrections in the process of determining its position.

GPS error sources

The Multipath Problem: Multipath propagation of the GPS signal is a dominant source of error in GPS positioning (specifically for DGPS solutions). Objects in the vicinity of the receiver antenna (notably the ground and buildings) can easily reflect GPS signals, resulting in one or more secondary propagation paths. These secondary path signals, which are superimposed on the desired direct-path signal, always have a longer propagation time and can significantly distort the amplitude and phase of the direct-path signal. Errors due to multipath problems cannot be reduced by the use of DGPS, since they depend on local reflection geometry near each receiver antenna. Multipath errors can increase in urban areas due to more severe reflection geometry.

Other error sources

There are several other sources of errors in GPS measurement that in most cases can be corrected to a large extent in DGPS solutions. The errors depend on time of day, and satellite positioning (in zenith or low orbits) as well as other atmospheric disturbances. These errors are *Ephemeris Data Errors, Ionospheric Propagation Errors, Tropospheric Propagation Errors, Satellite Clock Errors, Receiver Clock Errors*, and will not be further explained in this document.

For more information see the FESTA sensor and measurements Excel data table.

3.2.9 Geographical Information System (GIS)

One of the lessons learned in state-of-the-art FOTs in the US is that geographical information, such as road curvature, roadside embankments and other on-and off-road information has been underestimated as a valuable source of data in the analysis. That is, this information is used as contextual indicators of events and situations identified in the analysis and provided added insight into both behavioural aspects and how the infrastructure influences system performance. Also, it is used to guide the road authorities to build safer and better roads and traffic environment.

The GIS data can come from several sources. Navigation grade commercial map data is available from several suppliers and some are also working on ADAS grade map information. (See the final report of the MAPS&ADAS sub-project of the PReVENT project: http://www.prevent-ip.org/en/public_documents/deliverables/d121_mapsadas_final_report.htm) In addition to the commercial maps, many countries have national level map information project that are often limited in coverage but with high level of information were available. Lastly, an expensive alternative to getting the GIS information from suppliers is to collect the information within the project itself, using data collection vehicles. This data collection may make use of LIDAR (providing three-dimensional mapping of the entire road environment) and video, as well as very accurate positioning. Using this technology may be feasible if the study is run within a limited geographical region but still requires a significant amount of resources.

The GIS information can be used directly in the vehicles by extracting the information about the current driven roadway and store that directly as separate Measures. This could be road type, speed limit, rural/urban, banking, curve radii etc. By doing this on-line (in the vehicle), one do reduce the necessary post-processing, but adding it in the vehicle also requires additional software and possibly hardware. How this should be handled is up to each FOT. One advantage of performing this on-line can be that one can consider not including the absolute position (e.g. GPS) in the data storage, thus reducing the privacy concerns with absolute positioning. This technique has been used in some state-of-the-art FOTs. It should be noted that if GPS position is not recorded (and on-line map matching is made) it will not be possible to perform quality checks or updates with new maps/information after the data has been collected.

3.2.10 Driver annotation

For drivers who wish to make comments on specific events during driving, a annotation/comment button and a microphone can be provided. In state-of-the-art FOTs drivers have had access to a comment button on the dashboard, which, when depressed, would start recording any verbal comments during a pre-set number of seconds (usually around 20-60 seconds). Use of the button could be encouraged if the driver feels that it is warranted or at agreed (critical on non-critical) events. For consistency, inform the driver about the annotation possibility and provide some simple guidelines on how to offer the comment.

3.2.11 Vehicle metadata

Information about the vehicle is an important part of the analysis and study design.

The type of information needed from the vehicle will differ between studies and details will not be covered here. Recommendation are though that for each type of study, systems, functions and specifications that may act as confounding parameters in a specific analysis should be stored. The reason being that in analysis, even though the systems are not explicitly studied, they may have confounding effect if an inhomogeneous test fleet is used. For example:

- FOTs with focus on safety: Information should be stored about integrated systems that are not explicitly studied, but that directly may contribute to distracting the driver (e.g. active safety systems and real time information systems),
- FOTs with focus on environmental issues: A more powerful engine, automatic gear shift or four-wheel-drive, are most likely to be confounding parameters in an environmental analysis. Also, for trucks it is important to know what kind of cargo is usually transported (influences weight).

3.2.12 Coding/classification/transcription

As part of the data reduction and analysis process, described elsewhere in this handbook, sections of time will often be labelled with classifications according to a coding scheme or syntax. Depending on the study, sections of time can be given categories such as “crash”, “near-crash”, “incident”, “Curve speed warning”, “lane change”, “crash avoidance by steering”, etc. *When classifications are made they are often saved and thus become a new data source which is added to the database*. For example, an index indicating all instances of lane changes in the dataset can be created and saved. *Regardless if the classification of data is performed by a human* *analyst transcribing video or by an algorithm applying kinematic trigger values to the data, this process of classification should be seen as a type of sensor providing a new data source*. Thus, it is comparable to other types of off-line performance indicator calculations (see Figure 1). It is recommended to plan that these new data sources or measures will be created during the data reduction phase after the data has been uploaded to the FOT database.

Classification of events (time segments) into relevant descriptors of situations is an often overlooked problem and should be given sufficient attention. It is important to note that data classifications are important not only to identify events of interest, such as frequency of a warning. Risk estimation requires a “denominator” or exposure measures as well, in order to determine how often a certain event (e.g. lane departure) occurs per something (e.g. near-crash per left turn, lane departure warnings per drowsy event, or crash per rear-end crash type). These denominators may also involve a considerable amount of effort to correctly classify. Even more straightforward performance indicators may need some saved event classifications, for example mean speed per road type, ESC activation during different weather conditions, and so on. In contrast to simulator experiments where each situation is known and planned beforehand as independent variables, FOTs do not offer controlled contextual circumstances beforehand. The fact that test exposure is largely uncontrolled means that analysis is largely conducted by first identifying the important contextual influences, and then performing the analyses to create a “controlled” subset of data to compare with.

One particularly important method of classification that is uniquely used in FOT analyses is the process of using kinematic triggers (e.g. 0.8 g lateral acceleration) to detect critical events (e.g. crash, near-crash, incident). This process involves several steps whereby an event trigger is applied to the data to identify potentially interesting data, the video is thereafter reviewed by human analysts to verify if it was a critical event or not. These events are then often used as important safety indicators in subsequent analyses. The importance of video for validation that an event truly occurred cannot be overstated.

If a critical event has been identified an event coding scheme should be applied to describe the attributes of the found event. It is recommended and desirable to design the event coding scheme to correspond to the European epidemiological crash database coding schemes that are relevant for the region. Examples of variables which can be coded once an event is found are: Event descriptors (event type, crash type, severity, trigger type, fault, driver actions and behaviour), Environmental factors (weather, light, road characteristics and condition), and Driver state factors (drowsiness, distraction state, distraction type, workload level). Other types of coding and transcription procedures using humans to review data are quite common. A number of standardized manual coding procedures exist, for example the reduction of driver eye glance behaviour by video viewing is very common. Many coding classification schemes will be specific to each particular FOT.

3.3 Mechanical requirements

The following description of mechanical guidelines and requirements are primarily applicable on FOTs with on-vehicle Data Acquisition Systems, but significant portions can also be applied to nomadic devices and infrastructure DAS.

3.3.1 Size and weight

The size and weight requirements have to be set by each individual FOT. Since the vehicles in FOTs are driven on a daily basis and over a long period of time by normal drivers, the system should preferably have negligible effect on the driver's use of the vehicle – including limitations in trunk space.

3.3.2 Connectors and interfaces

State-of-the-art FOTs state that from experience as much as 80% of the DAS hardware problems can be deduced to physical connector issues. It is recommended that connectors with some locking between connector genders are used (that is, e.g. simple consumer electronics power connectors and USB-connectors should if possible be avoided). Cable pull-relief should be used when possible. Also, note that connector problems include the cabling to cameras and other satellite sensing and not only to the DAS itself.

3.3.3 DAS mechanical cover and ease of access

There are three main aspects of mechanical cover and ease of access. Firstly it is recommended that a layman without tools is able to find and have visual access to any indicator LEDs (not meaning they have to be in plain sight). Secondly, having the possibility to connect interface devices without having to remove covers is preferable. The interfaces may include an LCD screen for visualisation, mouse and keyboard, and network access (if an on-vehicle DAS is used). For Nomadic Device acquisition this is not applicable for the DAS itself, but for potential interfaces with vehicles (buses) this should be considered. The third aspect is EMC and emission issues.

3.3.4 Crashworthiness and vibration resistance

Depending on the focus on the FOT in question, the needed ruggedness of the DAS and sensors in terms of crashworthiness will differ. For all FOTs the minimum requirements for the ruggedness is that the entire system should operate under the normal driving conditions for the specific FOT, including the harsher situations of normal driving – for example driving on bad gravel roads and into pot-holes, or hitting a speed bump, curb or other smaller obstacle at too high speeds.

If the focus of the FOT is to study systems and situations that include times close to impact, it might be necessary to increase the ruggedness specification. If it is required that the DAS and the stored data should survive a full crash the costs for the system will most likely increase drastically. These kinds of systems are used in crash testing and do in most cases not include the large storage volumes need for example for video storage. If only an Event Data Recorder (EDR) without data sources requiring large storage volumes will be used, ruggedizing for crash events is easier and cheaper.

3.3.5 DAS environmental requirements

Environmental requirements for the DAS mainly concerns temperature, as for sensing the requirements vary widely depending on installation and sensing constraints (see each respective sensor type). If the DAS is placed in a reasonably shielded location inside the vehicle, the need for water resistance may be negligible, although the DAS internal parts should be able to withstand reasonable levels of condensation. If the DAS has active cooling, it is recommended that a simple dust/particle filter is placed by the main air intake of the DAS for dust protection.

The temperature requirements for the DAS are both one of sufficient cooling of the system due to internal heat generation and for ambient temperature.

Internal heat generation

The cooling needed for the DAS is mainly depending on the processor used and the workload on the DAS. If the FOT in focus is using video and the video compression is done in software, the need for processing power is usually drastically higher than if only CAN and other “scalar” signals are recorded. This also then generates significant levels of heat, and the DAS units will need to have proper ventilation and air flow not to overheat. The need for forced ventilation is depending on the DAS mounting position. It is recommended that when considering both hardware and software compression to choose a solution based on the placement of the DAS (due to increased heat generation) and budget constraints.

Ambient temperature

Too high and too low temperatures (both static and transient) do affect the DAS. From the DAS perspective the temperature aspects are mainly as follows:

- Most electrical components are effected by too high and too low temperatures. The actual effect is different for different components and devices.
- Some of the most sensitive devices in the logging system are the hard drives (with mechanical parts) in the cases these are needed (usually due to large storage volumes needed for video or other high-data rate sensing). Automotive grade hard drives are available, although somewhat more expensive than normal consumer hard drives. However, the main issue with these drives is the limited storage capacity, which is significantly lower than consumer hard drives. Flash memory cards or solid state drives (SSD) are available in increasing capacities and is a clear alternative to operating system hard drives and for storage in lower data volume FOTs.

3.4 Electrical requirements

The following description of electrical guidelines and requirements are primarily applicable on FOTs with on-vehicle Data Acquisition Systems, but significant portions can also be applied to nomadic devices and infrastructure DAS.

3.4.1 Power management

The main requirement for the DAS and sensing power management is that the installation should never affect normal vehicle function (cause the vehicle not start or making it difficult to start). This requirement should be enforced in all environmental conditions and is one of the most critical issues for drivers to accept equipment to be installed in the vehicles. If FOT installations are rumoured to impede on the trustfulness on vehicle operation, few people will volunteer in any subsequent study, and the study at hand may be compromised.

In technical terms, the DAS and the sensors (the entire FOT installation) may not draw power so that the vehicle battery charge falls below the level of being able to start the vehicle. The definition of being able to start the vehicle should include multiple start/stops in extremely cold environments. Also, care should be taken so that the system does not draw power (or only minimum power) if the ignition has been turned on but the engine did not start, or if the engine stops without the ignition key being removed.

3.4.2 Interference with on-vehicle equipment

In all on-vehicle installations in FOTs the aim should be to minimise the use of or preferably entirely manage without AC powered devices. The reason for this is mainly that the DC/AC and DC/DC converters are electromagnetic noise (EMC, http://ewh.ieee.org/soc/emcs/) sources for both standard on-vehicle equipment (such as the FM radio) and for FOT installation equipment (especially any analogue I/O device). The level of noise generated from the DC/AC or DC/DC converters differs between converters, but they all do add to the ambient noise. Other parts of the DAS or added sensors can also be generators of EMC. It is recommended (and in some cases required) to use CE certified equipment. If CE certification is not necessary, it is recommended that specific EMC testing is performed on the units to be installed. In the case wireless network connections are employed in the DAS, thorough investigation and consideration of possible electromagnetic interferences must be made to ensure proper function of data transfers.

Attaching any equipment to the on-vehicle bus systems (such as CAN or LIN) has to be done very carefully. More clearly, the manufacturer of the vehicle in question must provide direct and clear consent *and* guidance to connecting any type of device sending any type of electrical signal on the bus system. In most cases this should not be needed or done at all in an FOT implementation. Failure to adhere to this might be dangerous and result in vehicle operational malfunction that may result in significant cost, injury or death, or produce other very unwanted results. Also, even adding only listening/eavesdropping devices should be sanctioned by the vehicle manufacturer before being implemented.

3.4.3 Laws and regulations

Depending on the FOT study at hand different regulations will apply. For example, in some cases CE certification of the FOT equipment may be necessary. In most cases CE certification will not be a necessity, but each project much verify what is applicable to the specific study. If the vehicles are to be driven in non-EU countries the specific regulations for each region should be verified.

For wireless communication and for some sensing systems there are regulatory restrictions on transmitting electromagnetic radiation. A few example of sensing systems that have direct regulatory restrictions are LIDAR, RADAR and any other device that emits electromagnetic radiation. The restrictions may be based on electronic interference or harm. This has to be taken into account for each individual FOT sensor setup.

As a LIDAR instrument emits laser light, it may be subject to restrictions. The international standard IEC60825-1 (European EN60825-1) classifies lasers into hazard classes: from class I to IV (with several sub-classes). Class I lasers are generally considered non-hazardous, even at prolonged exposure (for example to the eye). Moreover, certain lasers are referred to as “eye-safe”. However, for each instrument and jurisdiction, care should be taken to investigate the applicable regulations.

Apart from general regulations and restrictions on the use of equipment that emit electromagnetic radiation, several countries have regulations for equipment employing RADAR or LASER technology on public roads, as these can affect effectiveness of for example authority speed surveillance instruments.

For further detailed coverage of legal aspects, see the FESTA deliverable D6.3.

3.5 On-vehicle storage and data retrieval/uploading

3.5.1 Storage capacity estimation

Factors to be considered for storage capacity estimation

The main aim of the *storage capacity estimation* is to guarantee the availability of free space for recording the vehicle data.

Storage capacity depends on the following factors:

1. Number of recorded signals
2. Sample rates of the recorded signals
3. Sample size of the recorded signals
4. Data collection method
5. Driving hours
6. Data size reduction by
    a. Filtering algorithms
    b. Compression algorithms
7. Data deletion procedure

1 - Number of recorded signals

The number of signals recorded depends on the different FOT designs. It is worth noticing that recording the full information on the CAN bus may not be necessary. Filtering the CAN bus information and recording only the signals relevant for the FOT research questions may save a considerable amount of space and result in a lower investment for the storage supports. Further, the number of cameras (if any) to record videos from may play a major role in determining the storage capacity needed.

2 - Sample rates of the recorded signals

Ideally, the sample rate for each signal should be the lowest possible able to guarantee no information, relevant for answering the FOT hypotheses, is lost in the sampling process. These sample values are reported in the FESTA D2.1 for CAN and sensor data. For videos, the best sample rate depends on which are the events of interest to be recorded. For instance, a 6 Hz frame rate may be enough to recognise the traffic surrounding situations but may not be sufficient to determine a high-speed crash dynamic. Using a dynamic sample rate and a high sample-rate buffer would make possible adapting the resolution of the recorded data depending on the event. For example, the data can be always collected at high sample rate and put into a buffer, and then the data is down sampled before recording if no specific event is detected. If a specific event, such as a crash, is detected, then data can be recorded with the high sample rate without down sampling. The drawbacks of using dynamic sample rate are 1) the recording system complexity increases and, as a consequence, the probability of faults increases as well and 2) more post-processing on the data will be necessary to handle the different sample rates. 3) Database design and searching becomes more complicated because more objects will be needed to keep track of the different sample rates (Chapter 5).

3 – Samples size of the recorded signals

The sample size depends on the nature and resolution of the signal to be recorded. Video data are normally much more “cumbersome” than CAN data at a sample level. Even if video data are collected at a lower sample rate than CAN data, it normally takes more space than the CAN data.

4 - Data collection method

Methods of data collection can be continuous, when all data of interest is logged during the whole drive, or limited to specific events of interest, when a real-time algorithm decides which signals need to be recorded and when they need to be recorded. Examples of possible event of interest are: time intervals in which the lateral acceleration is above a certain threshold or in which the vehicles enters a curve with speed above a certain threshold. Events of interest may vary across FOTs depending on the specific FOT hypotheses and scenarios.

The data collection method determines which percentage of the driving hours is actually logged (see Chapter 3.5.2). For instance, if the data collection method is continuous, this percentage is, in theory, 100% (in practice, it will be lower due to several issue such as logging-system boot-up delay). If the data is collected on a per-event basis, this percentage depends on the probability of the event of interest, i.e. how often the event of interest is likely to happen. In Table 1, a few indicative values for rough estimation of a few of the possible events-of-interest are reported.

Table 3: Probability of crashes, near crashes, and incidents in percentage of driving time.

Event	Probability of the event per driving hour *
Accidents	0.00046359
Near crashes	0.00004253

Crashes	0.00000458

* The values presented in the Table 1 are derived from data from the 100-Car Naturalistic Driving Study Report – VirginiaTech, Transportation Institute.

Note that other events, such as occurrences of safety systems warnings, can be considered. The probability of such events should be part of the knowledge of the FOT designer and should be used for the estimation of the storage capacity.

5 - Driving hours

Driving hours depend on the nature of the driver and the vehicle. In Table 4, a few indicative values for rough estimation are reported.

Table 4: Approximate hours of driver per month depending on the vehicle type.

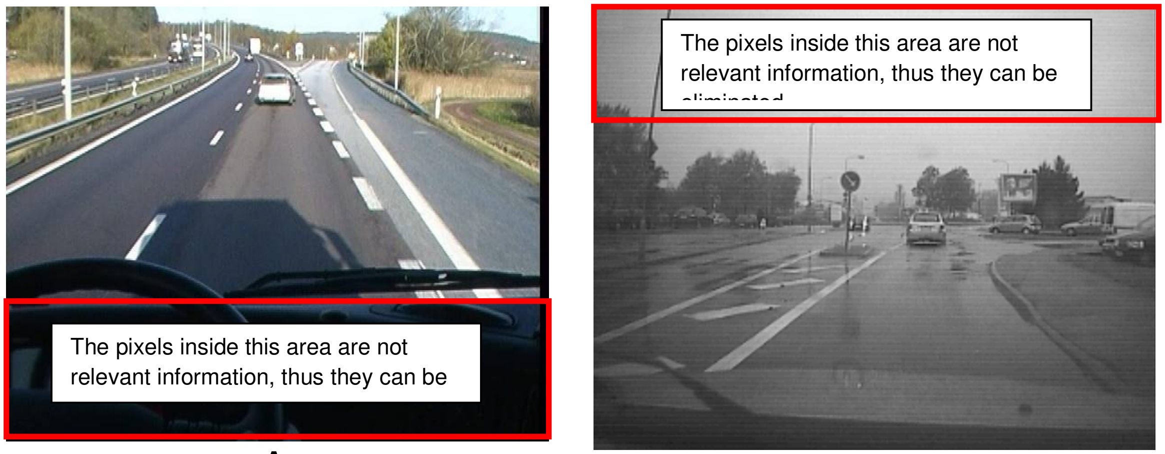
Type of vehicle	Approximate hours of drive per month	Reference
Fleet Truck	180 (1 driver) - 360 (2 drivers)	EG 561/2006
Urban Taxi	180*	Gutachten über die wirtschaftliche Lage des Hamburger Taxigewerbes
Public Transportation Bus	600	Internal Volvo AB studies
Private-use car	30	Internal Volvo Car studies

* Note that this estimation is based on the following two assumptions: 1) each taxi only has one driver, and 2) when the taxi is not moving (the taxi driver is waiting for a customer with engine on) this time does not count as driving.

6a - Data size reduction: filtering algorithms

Not all the data coming from a CAN bus, a camera, or a specific sensor may need to be recorded. Filtering/selection algorithms may help to select only the signals from the CAN or the pixels from the video which are relevant and deserve to be recorded (

Figure 6). The benefit of using such a filtering is to reduce the data size, the drawback is to increase the complexity of the system and introduce new possible source of errors and malfunctioning.



A B

Figure 6: Example of filtering algorithm on cameras data.

Panel A: Since the dashboard pixels are not relevant information, a filtering algorithm eliminates the dashboard pixels from each frame before recording. Panel B: Since the sky pixels are not relevant inf information, a filtering algorithm eliminates the dashboard pixels from each frame before recording. Panel B: Since the sky pixels are not relevant information, a filtering algorithm eliminates the sky pixels from each frame before recording. By applying these filtering algorithms video size is reduced up to 30%.

6b - Data size reduction: compression algorithms

Compression algorithms can help to reduce the data size. Compression algorithms can be lossy or lossless. A lossless compression (such as zip) can be used for CAN and sensor data whereas a lossy compression (such as MPEG4 and MP3) is normally acceptable for voice and video, respectively. Zip compression reduces the CAN data about 90% but has so far been rarely used in state-of-the-art FOTs since it adds significant complexity to the logging and data handling systems. MPEG4 compression can reduce video to about 2% of the original size and MP3 compression can reduce voice data also to about 2% of the original size. The benefit of using such compression algorithms is to reduce the data size, the drawback is to increase the complexity of the system and introduce new possible source of errors and malfunctioning.

7 - Data deletion procedure

A safe data deletion procedure implies that no data is deleted in the vehicle until a copy of the data has been backed-up, verified, and stored in a safe place. The procedure for verifying and storing the data (see Chapter 3.5.4) may take some time depending on the different FOTs. As a consequence, if the safe deletion procedure described above is used, storage capacity estimation should take the time for data transfer, back-up, and verification into account. If data is transferred by changing the storage support (e.g. substituting a full external hard drive with an empty one; Chapter 3.5.4), then the data deletion procedure does not affect the estimation of data storage.

Example

In the 100 Car Study (100 Car Naturalistic Driving Study – Phase I, experimental design, DOT HS 808536): vehicles were equipped with 5 cameras. Data from the camera was collected with a 30 Hz sample frequency and compressed with MPEG-1. Data on driving performance was collected with a 10 Hz sample frequency. It was estimated for vehicle to drive 6 hours day and on-board HD for data collection were calculated to last 2 weeks. However, data download was scheduled every week. Data was deleted with a remote wireless command. This command was issued only after the data had been downloaded, backed-up and verified.

Equations for the estimation of storage capacity

All factors above described influencing data size can be grouped in three clusters depending on their relation to 1) the recorded signals properties, 2) the FOTs experimental protocol, and 3) the data processing implemented on board. Figure 2 presents all the above mentioned factors, to be considered in order to evaluate of the storage capacity needs, divided into these three clusters.

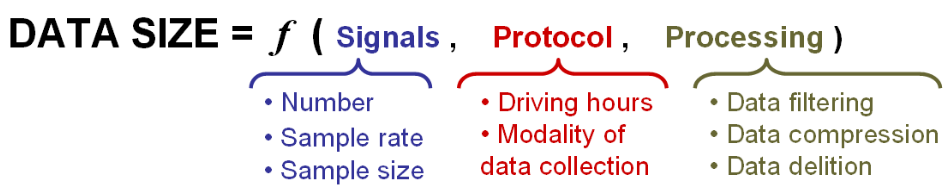


Figure 7: Factors influencing data size.

1) the *signals* recorded, 2) the FOT experimental i, and 3) the on-board data *processing*.

The general equation in Figure 7 can be represented in terms of more simple and defined equations in order to estimate, analytically, the storage capacity once the signals to be recorded, the experimental protocol, and the on-board processing are known. The following Equations 1-5 are derived from the general equation in Figure 7 and guide the experimenter to estimate storage capacity by including all information about signals, experimental protocol, and processing in a numerical format.

Equation 1: (estimate of the signals size per time unit)

$$S_{Size} = S_{Number} \cdot S_{SampleSize} \cdot S_{SampleRate}$$

Where:

S_{Size}: size of recorded signals

S_{Number}: number of recorded signals

$S_{SampleSize}$: size of the samples

$S_{SampleRate}$: sample rate (this variable defines the time unit; normally it is expressed in Hz, thus the time unit is s).

Equation 2: (estimate of effective data size after taking into account experimental protocol and processing)

$$D_{Size} = S_{Size} \cdot P_{Event} \cdot T_{Driving} \cdot F_{\mathrm{Re}duction} \cdot C_{\mathrm{Re}duction}$$

Where:

D_{Size}: data size per time unit
P_{Event}: probability of the event of interest
$T_{Driving}$: driving time
$F_{Reduction}$: filtering reduction rate
$C_{Reduction}$: compression reduction rate

If 1) the data is copied from the vehicle and not "picked up" from the vehicle and 2) a safe data deletion procedure is used, then, the time necessary for data transfer, back-up, and verification should be considered into the equation above as follow.

Equation 3: (estimate of effective data size after taking into account the time for uploading – not necessary if data is retrieved by substituting the storage support)

$$D_{Size} = S_{Size} \cdot P_{Event} \cdot (T_{Driving} + T_{Upload}) \cdot F_{\mathrm{Re}duction} \cdot C_{\mathrm{Re}duction}$$

Where:
T_{Upload}: data upload time

At this point, storage capacity depends only on how often data is retrieved / uploaded from the vehicle. Thus, finally:

Equation 4: (estimate of the storage capacity based on how often data will be uploaded or retrieved)

$$C = \frac{D_{Size}}{T_{Max\mathrm{Re}cording}}$$

Where:
C: capacity
$T_{MaxRecoridng}$: maximum time interval in between data upload/retrieval processes

Note: 1) filling up the storage support may be critical and 2) intervals for retrieval and uploading may present some variability. These 2 factors should be taken into account by guaranteeing enough tolerance on the final storage size. Since no space to record data would result in data loss a 20%-50% on storage size tolerance is recommended.

If storage capacity is predetermined (for instance due to the internal memory of the logging equipment), then Equation 5 can be used to determine how often data needs to be uploaded/retrieved from the vehicle.

Equation 5: (estimate the maximum recording time, which is a higher limit, for how often to upload/retrieve data) once storage capacity has been decided)

$$T_{Max\mathrm{Re}cording} = \frac{D_{Size}}{C}$$

Table 5 shows some example of data size estimation derived for some common types of vehicle assuming continuous data recording, driving hour as expressed in Table 4, no filtering algorithms, in a month time.

Table 5: Examples of data size for 1 month driving.

Type of data	**MB/h**	**Data size for 1 month driving (in MB)**			
		Fleet truck 2 drivers (360h)	**Taxi (180h*)**	**Bus (600h)**	**Private car (30h)**
CAN**	6	2160	1080	3600	180
Video†	140 (compressed) – 780 (uncompressed)	50400-280800	25200-140400	84000-468000	4200-23400
Voice‡	12 (compressed) - 300 (uncompressed)	4320-27000	2160-54000	7200-180000	360-9000

* Take into account that this estimate does not consider the waiting time (when the taxi engine may be on even if the vehicle is not moving) and the logger may be collecting as in a normal driving situation.

** CAN refers to 50 signals recorded from one CAN bus with a 10Hz sample frequency.

† Video refers to a B&W, 128 grey-level, 320x240 pixel camera recorded at 6 Hz compressed with the DJLS algorithm.

‡ Voice uncompressed refers to WAV 44100Hz, mono, 16 bit. Voice compressed refers to MP3 compression with a fixed bit rate of 128 kbps. Please note that the FESTA project does not recommend continuous audio voice registration due to personal privacy issues (section 3.2.5)

Please, note that the numbers reported in Table 5 depend on many factors (such as the type of vehicle, type of camera, compression algorithms used, etc.). Considering the high numbers of factors (and factors combinations) which play a role in this matter, it is not possible to present an exhaustive table. Also, it is worth noticing that video, and voice in particular, may not be recorded continuously as assumed in the table above. However, the values reported in this table are intended to give the reader an idea of the different data sizes in play for the most common combination of types of data and vehicles. Note that this setup may not be typical for following FOTs in terms of number of cameras, resolution, frequency or compression. Unique calculations should be done for each study.

Examples

Below, a few storage capacity estimations from internal Volvo and Volvo Car FOTs are reported. Once more, the numbers presented below depend on many factors which are not reported in detail, however, these numbers can be used to get an idea of the data sizes in play before starting an FOT to help the design of the FOT.

In a truck:
(1 CAN + 7 cameras + GPS) x 8h drive per day x 5 day result in about 20 Gbyte.

In a bus:
(4 CAN + J1587 + GPS) x 20h drive per day x 6 days result in about 20 Gbyte.

In a private car:
(1 CAN + 1 Video compressed + GPS) x 1h drive per day x 30 days result in about 20 Gbyte.

3.5.2 Data collection methods (continuous & triggered)

As mentioned above in Chapter 3.5.2, data collection methods can be continuous, when all data of interest is logged during the whole drive, or limited to specific events of interest, when a real-time algorithm triggers the logging of a predetermined set of signals in a decided time.

For continuous logging, the main issue to take into account while defining data retrieval and upload procedure is storage space. In fact, continuous data collection may create a huge amount of data, especially if video data is recorded. When triggered data acquisition has been chosen as the method of data collection, great care has to be taken to define the trigger definitions. Thus, from a sensing and logging point of view there are a few aspects that are important to point out.

- Even if triggered logging is used for the evaluation of effects, most studies will require baseline data. It should be possible configure and acquire baseline data also for these cases. The baseline data acquisition can be collected in different ways (FESTA D3.5 ANNEX I). If baseline will require continuous recording, storage and retrieval/uploading should be decided accordingly.
- The probability of the triggered event plays a major role in data storage estimation and, as a consequence in data retrieval / upload frequency. Thus, if this probability is not well known a priori, data storage estimation should take this into account by applying a bigger tolerance on the calculations.
- If the probability of the events of interest triggering data recording is not known and, as a consequence, data upload/retrieval cannot be accurately scheduled, then, data upload / retrieval may be controlled by communication / monitoring of the available recording space on the vehicle. This can be implemented wireless using automatically generated e-mail or SMS or installing a display showing this information to the driver.

It should be noted that several state-of-the-art FOT experts have stated that using only triggered acquisition in FOTs may result in incomplete data, making analysis difficult or inconclusive. The issues are e.g. problems comparing treatment period with baseline (or lack of baseline) and missing important events. It also makes additional use of the data (non-project specific) less probable.

As example of triggers, the following triggers are a subset of what was defined in the 100-Car Study (the Naturalistic data collection in 100 cars for one year, performed by Virginia Tech – Report on the results of Phase II, Appendix B). Note though that this was a naturalistic data collection study (aiming to identify incidents, near crashes and crashed as defined in the 100-Car Study) and not an FOT. Triggers in FOTs are more likely to be system activation based (if at all used on-line).

- Lateral acceleration - Lateral motion equal or greater than 0.7 g.
- Longitudinal acceleration - Acceleration or deceleration equal or greater than 0.6 g.
- CI button – Activated by the driver upon pressing a button located on the dashboard when an incident occurred that the driver deemed critical.

- Forward Time To Collision (FTTC) - Acceleration or deceleration equal to or greater than 0.5 g coupled with a forward TTC of 4 seconds or less.
- All longitudinal decelerations between 0.4 g and 0.5 g coupled with a forward TTC value of ≤ 4 seconds and that the corresponding forward range value at the minimum TTC is not greater than 100 ft.
- Rear Time To Collision (RTTC) - Any rear TTC trigger value of 2 seconds or less that also has a corresponding rear range distance of ≤ 50 ft. AND any rear TTC trigger value where the absolute acceleration of the following vehicle is greater than 0.3 g.
- Yaw rate – Any value greater than or equal to a plus AND minus 4 deg change in heading (i.e., vehicle must return to the same general direction of travel) within a 3-second window of time.

3.5.3 Additional memory requirements

There are a few different FOT approaches that require larger main memory capacity (RAM). FOT activities that use any type of triggered data acquisition (see Chapter 3.5.2) and have high data-rate data sources will need significant amounts of main memory to handle the necessary ring-buffer for pre-triggering storage.

3.5.4 Data retrieval/uploading procedure

Data retrieval /uploading is a procedure aimed at making sure that all data collected in the vehicles during the FOT are conveniently backed-up and stored in a safe place without any loss. The Data retrieval /uploading procedure has 4 main goals:

1) To assure that a copy of the data collected in the vehicle is stored in a safe location
2) To prevent data loss by having multiple copies of the data collected in the vehicle stored in different safe places.
3) To verify that no data loss occurred in the copying process
4) To ensure that storage space is newly available in the vehicle once the recorded data has been safely transferred, backed-up, and verified.

The 4 main goals of data retrieval /uploading are achieved in 4 chronological steps as shown in Figure 8.

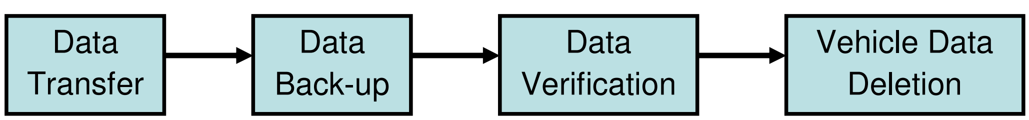


Figure 8: Data retrieval /uploading steps.

Data transfer **is aimed to assure that a copy of the data collected in the vehicle is stored in a safe location.** ***Data back-up*** **is aimed to prevent data loss by having a multiple set of the data stored in different safe pl places.** ***Data verification*** **is aimed to assure that no data was lost during** ***data transfer*** **and** ***data back-up*****.** ***Vehicle data deletion*** **is aimed to ensure that storage space is newly available in the vehicle once the data has been safely transferred and backed-up.**

Data Transfer

Depending on the support used to record the data in the vehicle and the data size, different data transfer modes can be implemented. If the data is recorded on a flash card or on a movable hard drive, the driver (or a technician) can substitute the support, with a new, empty one, when full of data, then make sure that the support full of data is conveniently stored. If the data is recorded in an internal hard drive or memory, then the data transfer method will imply some kind of connection. The connection can be wired (e.g. USB, FireWire, serial) or wireless (e.g. Wi-Fi, GSM, GPRS). Depending on the different 1) kinds of connection technology, 2) transfer protocols, and 3) application protocols different transmission speeds can be expected (Table 6).

Table 6: Range and transfer range for different possible connections for data uploading.

Connection	Range	Usable transfer rate (up to)	Time to transfer 1 GB [s]
USB 2.0	5 m (cable)	40 MB/s	≥ 25
FireWire 400	5 m (cable)	49 MB/s	≥ 20.3 (FireWire800 => 10)
Ethernet	(depending on network topography)	75 MB/s	≥ 13
Wi-Fi	(network coverage)	31 MB/s	≥ 34
Bluetooth 2.0	~10 m ~100 m with external amplifier	0.26 MB/s	≥ 3938
IrDA	~1 m	9.6 kbit/s (SIR: Serial Infra Red) 2 MB/s (VFIR: Very Fast Infra Red)	> 853333 (SIR) ≥ 500 (VFIR)
Turbo 3G	Like GSM/ GPRS, depending on network coverage	0.9 MB/s	> 1137
GPRS	(network coverage)	57.6 kbit/s	≥ 138889
GSM	(network coverage)	14.4 kbit/s	≥ 555556

Data transfer presents two major issues: 1) it may be time-consuming and 2) data can be lost during the transfer. Depending on the data transfer mode, these two issues can play a different role. Table 7 presents an overview of benefits and issues related to the different data transfer modes and rates the different modes in terms of time efficiency, cost efficiency, technical complexity, data security, and driver comfort.

Table 7: Pros and cons for different data retrieval/ uploading modes.

Transfer mode	Time efficiency	Cost efficiency	Technical complexity	Data security	Driver comfort

Data is “picked up”	+ Very fast from a vehicle point of view.	– May require hiring someone to pick up the data. – May require an extra vehicle chasing the FOT vehicles.	+ Very reliable because simple.	– Data may be misused between the “pick-up” and the final storing.	– May require the driver to go somewhere or to move the data.
Data is transmitted with wired transmission	– May be slow for big amount of data.	– May require hiring someone to do the downloading.	– The equipment needs to be safely accessed.	+ Can be very secure if the driver is not able to access the data.	– May require the driver to go somewhere or to move the data.
Data is transmitted with wireless transmission	– May be slow also for relatively small amount of data.	+ Download can be automatic. – May be expensive depending on the network providers.	– Automatic wireless download is very complex. – Needs “hot spots”.	+ Can be very secure.	+ The driver does not need to do anything.

Data Back-up

Data should be backed-up and stored in a safe place as soon as it is available. Ideally, the backed-up data and the main copy of the data should be in two different safe places. The back-up procedure may be time-consuming depending on the amount of data. Data can be backed-up on different supports such as hard drives, DVDs, or network storage. Many service providers offer space accessible via internet. If data will be organised into a database, as it is recommended by FESTA (Chapter 5), then the back-up procedure needs to be harmonised. In fact, also the database will have a back-up. If the data included into the data base will be exactly the same, but with a different structure, as the data recorded, then the database back-up may make a back-up at this stage unnecessary. However, chances are that the database will not have all raw data (because data may get filtered, resampled, normalised, etc.), in this case it is important to make a back-up at this stage to make sure no raw data will get lost.

Data Verification

Data verification should be an automated process and should check that both the main copy of the data and the baked-up copy are a perfect image (no data loss) of the original data. Data verification may also be time-consuming, however data verification is important since the probability of errors during the copying process is high due to the potentially huge amount of data handled.

Vehicle Data Deletion

Data from the vehicle should be deleted only once the data has been backed-up and verified. As a consequence, the time for data transfer, back-up, and verification should be considered into the estimation of the storage space capacity. This last point does not apply in case data is “picked-up” instead of copied from the vehicle (see Chapter 3.5.1).

Example

In the 100 Car Study (100 Car Naturalistic Driving Study – Phase I, experimental design, DOT HS 808536): a wired data transfer was implemented. Two chase vehicles were used to download the data. Continuous monitoring of the vehicle was implemented via wireless. Via wireless 1) the chase vehicle received GPS coordinate of the cars and 2) it was possible to know immediately whether there was any data collection failure. Data was backed-up twice in DVDs. Then, data was reduced and a new back-up was made. Another copy of the data was stored in the server for data analysis. The on-board data was deleted with a wireless remote command only once the data was verified.

Data Loss

Despite its apparent simplicity, data retrieval/upload has been proven to be a critical procedure: experience from previous FOTs evidences that data loss at this stage is often experienced even if it could be totally avoided with a robust and well-tested data retrieval/upload procedure. For this reason, we reported below a few tips which may seem obvious even if FOT history proved they are not.

To prevent data loss during the data upload/retrieval procedure it is important to:

1. Verify data is consistent before deleting it from the vehicle. In case, data is not consistent, the retrieval/ upload procedure should be performed again.
2. In case data is picked up and the data is not consistent, the vehicle data logger should be checked as soon as possible.
3. Monitor the state of the data logger so that any issue leading to data loss can be recognised and fixed as soon as possible.

3.6 System configuration

Depending on the type of FOT the DAS system configuration needs are significantly different. One basic recommendation that applies to all FOTs is that it should be possible to find and review the configuration of a specific DAS after the study has been finished. Other recommendation, such as the need for quick or on-line configuration changes only apply to some FOTs.

3.6.1 DAS inventory management

A system for basic inventory management is recommended for FOTs with more than a few vehicles in use. Either simple tools as spread sheets or more advanced database tools may be used, depending on the particular needs of the FOT. The purpose of an inventory management system is to keep track of all pieces of hardware used in the project, together with the relevant particulars, using well-defined procedures. For such a system to be efficient, sensors, DAS units, vehicles and all other equipment need to be included, as well as relevant supporting procedures developed. The employed system should allow for arbitrary grouping of hardware into "setups", and be able to track changes to these setups. In this way, a complete DAS installation – including sensors, devices, computer, antennae, etc. – for a particular vehicle can be followed through the length of the FOT. For any one point in time it should be possible to deduce the exact hardware configuration of a particular installation.

Several commercial and free software products are available, but it is likely that some amount of customisation need to be made to any tool and procedure used. The FESTA sensor matrix should be used or included in an inventory management system.

3.6.2 Configuration tools and traceability

In addition to an inventory management system, it is appropriate to employ a system (and supporting management procedures) for configuration management. Most DAS implementations include not only hardware, but also software versions and additional information such as (but not limited to) software configuration files, start-up scripts, device calibration files, and other file based data needed for proper operation of the DAS. For each vehicle or installation all this information should be stored together with information about the present DAS hardware configuration.

Similarly to above, it is important that traceability of software configuration changes is maintained. In this way, the exact software configuration of a particular installation at any particular time can be reproduced and reviewed. Possible problems can easily be traced back to particular revisions of the software or configurations. For software, source code, and all other file based data, version control software could be employed. State-of-the-art FOTs have used similar systems for tasks such as configuration management, "bug tracking", and system service documentation.

3.6.3 Switching between configurations

For FOTs where the vehicles will have several (primary) drivers and the study time for each driver is limited, methods for switching between configurations in each vehicle are necessary. This changing of configuration may include only changing driver ID, but it may also include driver specific configuration settings. In addition to having to change DriverID when a new subject is introduced to a vehicle there are a few other situations where it may be important to know who was driving or what was done with the vehicle. If this is not separated from the subjects, these Trips may be erroneously classified as a subject drive. Here are some examples:

- The vehicle driven back for overhaul/maintenance
- Validation testing just prior to vehicle delivery
- Other engineering testing

The actual switching between configurations can be implemented both with hardware and as software. That is, one can design a switch box that physically connects to the DAS, with which a dial or turn knob can be used to set the current configuration to use. For software, some way of interacting with the operating system is probably necessary. It can either be the connection of interface peripherals such as keyboard, mouse and screen, but it is recommended to use remote desktop tools over wireless (e.g. WLAN or 3G) or wired networking (Ethernet).

3.7 Acquisition of data

The automatic data acquisition system (DAS) is often dependent on its host system for power supply. If, for instance, installed in a vehicle it is of great importance to strictly control the power supply to the DAS since it might drain the vehicle battery if the DAS is running over night. When controlling the power supply to the DAS the start-up and shutdown speeds must be optimised to reduce loss of data. Loss of data can occur both during hardware initiation when no software is started and during hardware termination when no software is able to trigger on a vehicle restart. Power control need to be managed by the DAS power management unit and by the DAS software. The following sections will apply for a DAS installed in a vehicle or on another host with limited power resources requiring shutdown when idle.

3.7.1 Start-up

Normally the data acquisition will start as the vehicle ignition is turned on. Once the hardware is running, the software shall be initiated and started as soon as possible. In order to minimise the data lost during the start-up procedure, the hardware and software must load and initiate as quick as possible. One further possibility is to start the DAS upon a pre-start signal from the vehicle. For instance, there could be a signal from the vehicle locking unit generated as soon as the unlock command is received from a remote key. If and how a start-up signal can be achieved depends largely on the vehicle manufacturer and on their collaboration in the FOT. Nevertheless, a state-of-the-art DAS should be prepared for this function since it might save from a few seconds up to minutes of lost data at the start of a trip. The start-up time (or the duration where data is lost) should be well monitored and documented, preferably as a property associated with each recorded trip (since it might differ with temperature etc.). Start-up of the DAS hardware shall not be done if the voltage is too low.

3.7.2 Acquisition of data

A state-of-the-art FOT will continuously record data from ignition on till ignition off. For specific purposes, different event based recordings may be used to record only certain parts of the trip. Independent of the trigger criteria the DAS hardware shall be kept powered on and running during the entire trip. To ensure that the host power system is not overexerted, the power management unit must continuously monitor the power supply and initiate shutdown if a permanent voltage fall is detected. The system must not shut down on temporary variations such as the drop during engine crank. For such circumstances, an energy reserve may be required.

3.7.3 Shutdown

When the DAS recording has stopped, typically after ignition off, the DAS should be kept running for a short time (typically a few minutes) in case the vehicle is started again. If this time limit is passed the hardware should shut down as fast as possible since it is unable to initiate a recording during shutdown. The power management unit should then, after a short interval, cut the power supply to the DAS – independent if it is properly shut down or not.

3.7.4 DAS adaptation for wireless uploads

If the DAS is equipped with wireless upload that has limited access points, the actual upload might need to be performed when the vehicle is not running, e.g. after parking next to a WLAN access point. The data upload is likely to require more time than the few minutes provided after ignition off before the DAS shutdown is initiated. To work around the automated shutdown, a keep-alive signal can be sent from the upload software to the power management unit to prevent forced shutdown. Once the keep-alive signal is dropped the shutdown shall initiate as normal. It is recommended to use a security time limit (typically in the magnitude of hours) in the power management unit when a keep-alive signal is received.

3.8 Synchronisation

There are several types of synchronisation that need to be considered when designing and working with FOTs, depending on the focus of the FOT and the sensors and systems used and evaluated. Note that in the following context FOTs are considered to be used for logging of data and not as part of a hardware-in-the-loop component (for for example prototype system development). This means that data from the different data sources do not necessarily have to be available for storage close to the real-world event, as long as they are individually time stamped for off-line recreation of the time-line. The recommended method to perform synchronisation between the different sources of data is to use GPS UTC time, when possible and applicable.

3.8.1 Time stamping versus real world event

For each sensor, there is a time (hereafter called latency) between when the actual real world event takes place and when the data from each respective sensor is time stamped in the logger. This latency will be available for all data sources but the level, stability and predictability of latency differ greatly between data sources. In most FOTs there is a need to specify the requirements for what is allowed levels of latency is, based on the hypothesis. It is recommended to explicitly evaluate what the latency is for each data source (and for some sources individual measures) and compare this with the defined requirements (based on hypothesis – needs to be done as part of hypothesis generation). Ones the latency has been measures for a data source/measure and it has been established that the latency is not fluctuating/jittering too much, the time stamp time can be corrected by offsetting with the latency. For some sources the fluctuations and jitters in the latency is significant, making this correction of time difficult or impossible.

For data sources that are not controlled by the DAS implementers, such as vehicle CAN, it may be even more difficult to obtain the necessary information for the latency from real world event until time stamping. The reason for this is that for data sources like vehicle CAN, the information is packeted and sent by each node and priority scheduled. This means that the data flow might not be equally spaced and may have a jittering in the latency that is more unpredictable than for many other data sources. For measures on these buses that are related to physical events it is often still possible to evaluate latency if a reference sensor is used. For example, one can set up an experiment to measure the latency from time information about a safety system warning is available on the CAN and the visualisation of a HMI – this may show a significant delay which has to be taken care of in analysis. Also, comparing acceleration sensing between on-board and reference sensor gives a good indication of the latency.

3.8.2 Integrated sensing synchronisation

Depending on the methods of analysis and the implementation in the database, the needed level of synchronisation as well as the importance of measuring latency between different integrated data sources in a vehicle will differ. In Table 8, issues and methods for calculating the latency for some on-vehicle data sources are shown. It is recommended to use an internal continuous clock for synchronisation between integrated data sources. This clock is then preferably synchronised with GPS time. At least the latency should be calculated between GPS time and system time. The easiest way to get GPS time information is to read UTC time from a GPS receiver. Using this approach, when the vehicle is under a tunnel or in a location where the GPS signal is unavailable, the internal time information is still available from the internal clock. It is thus possible to perform a synchronous analysis with other external GPS synchronised data sources.

Table 8: Methods and issues in calculating latency for on-vehicle data sources.

Data source	Methods and issues in calculating latency
Video	May produce significant jitter/fluctuation on itself and other data sources if the data. Rough latency can be found using a synchronised LED light (measured by digital I/O)
GPS	The latency can be calculated very precise since the GPS time is the actual acquisition time
CAN	Difficult to establish latencies for internal systems, but with reference sensors it is possible to get the latency for some measures.
Acceleration/Yaw rate	Latency can easily be measured using a reference sensor, although care has to be taken with synchronisation between the reference sensor and the FOT DAS system.
RADAR	A lab setup with reference sensors of tracked object motion can be used

3.8.3 Synchronisation with nomadic devices

The absolute time accuracy derived from GPS depends on how this time synchronisation signal is acquired and used. All acquired data should be associated with a time stamp that is represented as absolute GPS derived time in order to enable comparisons with data stored by other data sources than the nomadic device.

GPS receivers can be easily connected to nomadic devices, as it is done for portable navigation systems. For use of nomadic devices inside FOT, it is recommended to use nomadic devices that have easy interface with GPS, so that the absolute time information is available.

Another possibility is to run on the ND a program that synchronises the internal clock time to the signal received from GPS. These applications are called NTPD (Network Time Protocol Daemon), and are the same applications used on NTP servers that provide reference time on a network. In that case, the time associated with the data is the time derived from the internal clock, and can have an accuracy that is much better than one second.

3.8.4 Synchronisation of infrastructure systems

In the case of infrastructure, two ways can be used to achieve synchronisation of data. One way is to use also here UTC time derived from GPS, if useful corrected with the time zone difference. The other possibility, when a server is available that is connected on the Internet, is to use synchronisation services that are available on-line, usually called SNTP (Simple Network Time Protocol). Also here the time zone should be clearly defined, and accuracy of one second is easy to achieve.

Using the data from the infrastructure inside an FOT could be critical in term of synchronisation, in time and space.

Time synchronisation
Depending on the type of data the synchronisation could be more or less critical. Weather information, like visibility range, are changing slowly, in minutes, so a good quartz clock synchronised every day or week is sufficient. Different it is the case of traffic data, where a 10 or at least 100 ms synchronisation is needed. In this case a GPS receiver is required, to guarantee precisely synchronised data.

Space synchronisation
All sources of information have to be localised, with good accuracy (in some applications to significantly less than 1 metre) to allow being associated to on-vehicle information.

As for time synchronisation, in case of traffic data it is important to localise vehicles with accuracy. But in this case the problem is more on the vehicle side, where the localisation is, normally, less accurate. This makes difficult to process data to identify the FOT vehicle inside the vehicles recorded by the infrastructure sensor (association). In this case the precise timing available from the GPS receiver could be very useful: specific software could be developed to find correspondence between data sensors acquired on the infrastructure and on the vehicle. But this could be very complex in case of dense traffic.

As a conclusion it could be suggested that in case of information slowly varying in space and time, these information could be stored locally and transferred to a central server that could also fuse them with other sources of the same info (e.g. weather data from terrestrial and satellite station).

In case of traffic data and safety related FOT, this data have to be stored as raw data, accurately synchronised in time, to allow the reconstruction of the scenario in the following data analysis phases.

3.8.5 Synchronisation of cooperative systems

Systems with communication between vehicles can also be realised without a central infrastructure that is a fixed station connected to all vehicles. In fact, some systems exist that rely only on the cooperation between the vehicles involved. One example is given by the applications developed inside the WILLWARN subproject inside PReVENT Integrated Project. Here, communication only between vehicles – without a central station – is the base for driver assistance functions. In this case the functions are realised through an ad-hoc network that is created dynamically, comprising the vehicles that at that given moment are connected. This means that the network continuously changes with nodes entering and exiting at any time (ad-hoc network).

In such an application, a central logging system is not possible, so the logging has to be done on every vehicle for the information that is available on that vehicle. Also here, it is recommended to use the common reference time provided by GPS. This should be easily done because many of these systems are using GPS localisation, so that the information is already available.

3.8.6 Synchronisation with interviews and other subjective sensors

In the case of interviews and questionnaires the time accuracy is not always critical. It is enough in these cases that the interviewers note date and time (hours and minutes) of the interview or questionnaire, so that these values can be included in the FOT database at digitalisation.

If the subject is requested to indicate and/or comment events during the driving (for example by pushing a button), this should be time stamped when logged if possible to be able to synchronise the event/comment with other data. The accuracy needed is in most cases less than 5 seconds.

Post trip collection of comments could produce errors but are sometimes necessary and extra care needs to be taken to manual synchronisation. For post-hoc structured comments or questionnaires on video or events (for example the driver or someone else commenting on driven events), it is important to define a process of linking the events to absolute time. Preferably this is done with digital tools for the questionnaires that automatically can be linked to events in for example objective data in a database.

The previous discussion is referring to a type of synchronisation that could be indicated as “event driven”. The driver has seen something of interest for the FOT and pushes a button to allow the recording of vehicle data and write down (or record) a comment, explaining the reason why this event is important (misbehaviour of the system, false alarm, missing alarm, etc.)

There are applications that require a less precise synchronisation in time and space. This is the case for example of traffic applications. A good description of the route and the timing could be necessary in order to correlate the “perception” of the driver with real infrastructure data. In order to accomplish this it would be preferable to provide some equipment, e.g. a nomadic device, to allow the gathering of this information.

3.9 DAS status and malfunction management

3.9.1 Self diagnostics and layman feedback

To simplify laymen feedback on system status (for example over the phone) to the responsible technicians at times of system problems, LEDs or a similarly externally viewable GUI about the system status can help and are recommended. The LEDs should then preferably give feedback on the DAS being operational and storing data, as well as the operational status (good/bad) of each of the individually attached data sources. The status of kinematic or system triggers could also be displayed.

3.9.2 System status uploads

In order for the people responsible for DAS and sensing in the project to be able to continuously monitor the status of the test vehicles while on the road, a remote (wireless) transfer of the system and sensing status is preferred. See Chapter 4.1.

3.9.3 Malfunction management

Malfunction management includes the process of identifying a problem in the vehicle to contacting the driver and exchanging sub-systems or the entire vehicle. For details see Quality Management (4.1.2).

3.9.4 Spare system management

For quick and efficient problem solving it is recommended to keep spare parts for all components pertaining to the DAS. Computer system hardware (hard drives, etc.), extra sensor equipment, cameras, and other various devices being part of the instrumentation of the studied vehicle should be included. All spare parts are to be managed by the inventory management system (see section 3.6.1) so that different parts used in the FOT can be traced, and so that the project does not run out of spares unexpectedly. Testing and calibration of spare parts should be planned and performed within the process. Supporting management procedures need to be developed as well.

Experiences from state-of-the-art FOTs suggest that, if applicable and depending on the scope and size of the FOT, it is recommended to always keep one extra vehicle for quick replacement. If not applicable, at least a fully equipped DAS system should be kept “on the shelf” – prepared and calibrated for immediate use.

It has been suggested that a 10% extra number of spare parts should be kept within the FOT, and that this number be added specifically when estimating the total project cost.

3.10 System installation

3.10.1 Installation procedures

Before initiating the installation procedures, an installation specification document shall be prepared. The installation specification must in detail describe how each component of the system shall be installed. Specifically, the installation specification shall provide solutions to the following topics:

- Mounting: positioning, means of attachment, accessibility, safety and security.
- Cabling: dimensions, shielding, drawing, mounting, tolerance and labelling.
- Connectors: soldering/pressing, robustness, impedance and labelling to avoid mix up.
- Power supply: consumption, fuse, voltage, source and switching.
- Environmental endurance: effects on electromagnetic disturbances (EMC), temperature, humidity, vibration, shock, electric safety and dirt.

An FOT data acquisition system is often installed next to other systems in a vehicle or external infrastructure. Therefore it is of great importance that the FOT system installation is adapted to the requirements set by all other systems. If this is not done properly the installed system could generate disturbances that might void warranty of the original systems – or even an entire vehicle. To ensure that no interference will appear, all systems that possibly may be in conflict with the installed system need to be identified. An adaptation plan must then be developed for each system to ensure that they will be able to operate properly after installation.

The actual installation work needs to be done by operators that are authorised to work on the actual host system, e.g. road authority authorised personnel for roadside installations and authorised vehicle mechanics for vehicle installations. During the installation work, all changes to the host system (if any) must be documented in detail.

If several units are to be installed, e.g. several vehicles, it is recommended to select one unit as prototype. The prototype unit can then be installed in cooperation between mechanics and engineers to ensure that all considerations are met and that the best solutions are found. The prototype installation will then revise the installation specification continuously during the work. Using a prototype is especially recommended for estimating installation time required for further installations.

3.10.2 Installation verification and calibration

When the system is installed it needs to be verified and calibrated before the data acquisition starts. The verification will refer to the installation specification and verify that all requirements are met. This involves verification of each component and sensor individually as well as the system as a whole. Specifically, if there are other systems that could be affected by the installation, all interferences need to be monitored to verify that no conflicts can be generated.

To ensure data validity and quality a calibration and verification scheme is recommended. Through a pre-defined set of actions, each sensor is measured and compared to an external reference sensor. The deviation is monitored and used for calibration of the sensor measurement. The scheme for calibration and verification will depend largely on the actual system installed since some systems require a carefully prepared calibration while others are ready to use at once. For data quality aspects it is important that all installed systems of the same category are calibrated and verified using the same procedures.

3.10.3 Dismounting the system

When the data acquisition is finished and the system is to be uninstalled, the installation documentation shall be used to ascertain that the host system is restored to its former condition. The list of possible interactions with other systems that was developed during the installation must be consulted to ensure that solutions applied to reduce interference are restored. Finally, all proprietary data need to be removed from the FOT system before it is disposed or reused in future projects.

3.11 Proprietary data in FOTs

The concerns regarding proprietary data are to keep the CAN/LIN/MOST specifications OEM-confidential and to hide the actual system design to prohibit reverse engineering based on data collected within an FOT project. Regarding the first issue – to keep the CAN/LIN/MOST specification OEM-confidential - there are two cases to be distinguished.

- When the OEM is strongly involved in the data acquisition process during the FOT execution, which means the data loggers are adapted and installed in the test vehicles and data quality management is handled by the OEM or a OEM authorised subcontractor: In this case confidentiality of the CAN/LIN/MOST specification is not an issue within the project.
- When the OEM doesn't handle the data collection by himself, the usage of CAN gateways is proposed. The CAN gateway has to be programmed by the OEM to provide the data from the CAN/LIN/MOST bus according to the agreed logger specification (see section 3.2.1).

The second issue – reverse engineering of functions and systems – is much more difficult to handle. Of course it is unlikely that low level system data will be made public in any project. However reverse engineering based on low level system data is also an issue within the FOT project. For example it could be necessary to record low level data to answer a specific research question. On the other side, data analysis is typically done by universities and research institutes to assure independency of the results. Each project will have to handle this and define what is needed. In some cases it may be necessary for the OEMs/suppliers of nomadic and after-market devices to handle detailed low level data and aggregate it on a certain level before it is provided to the project partners responsible for data analysis. A general recommendation to future FOT projects is to define in advance, what level of system data is needed to answer a specific research question and whether the involved OEMs are able to provide this data to the project.

In some FOTs OEMs/suppliers of after-market devices might be interested in the acquisition of additional data, which is not directly related to the project and proprietary to the OEM/supplier. This should be allowed. One possibility to deal with this could be to foresee specially protected storage capacities for this data. Another possibility is that the OEM separates the additional data from the project data before the data is provided to the further project for analysis.

It should be noted that in some FOTs where there are no OEMs involved, systems suppliers can still be partners. The issue of proprietary data is as strong in this case as in the case of OEM involvement.

3.12 Personal integrity and privacy issues in data acquisition and analysis

This chapter deals with the technical implications of personal integrity and privacy issues on the data acquisition and analysis in FOTs. However recommendations for the definition of necessary legal arrangements depending on the specific FOT are not covered here.

Different levels of data security should be implemented in order to cover personal and privacy issues properly. Namely not all project partners have the permission to access all of the project data. The data access right of a project partner should depend on his specific role in the project.

In section 5.2.6, which discusses different types of data used in the database design different levels of data security are proposed. Driver data allowing direct conclusions about the identity of the driver is considered to have highest requirements regarding data security.

Video data of the driver’s face and GPS data are typical examples of the data that belongs to this category. Access to this data needs to be controlled through e.g.:

- Viewing video data is only possible on-site at specific places.
- Access to GPS data/video data requires a certain level of data access right.

Some types of metadata like the vehicle serial number also belong to this category of data. For this kind of data pseudonymisation is required.

Dissemination of driver data outside the project needs to be approved explicitly by the participants and some additional “tricks” (e.g. blur people’s face on video, only road type and not GPS data is provided etc.). Even though the initial consent of the drivers could be enough also for external dissemination in some cases, it is still recommended to inform drivers explicitly about any external dissemination.

Depending on different laws in different countries it might also be necessary to implement a GPS data based control, which deactivates the video recording, when required. It is up to future FOT projects to take care of this issue.

4 Data quality

In all FOTs assuring data quality in the data collection and data management process is very important. The procedures for data quality assurance before, during and after the data collection should be well defined. Specifications and plans should be written for each individual FOT. In this chapter focus will be on objective and not subjective data quality assurance. The quality assurance issues with questionnaires and interviews are not covered here. (See Deliverable D2.4.) Also, the focus is on FOTs with a DAS system that has some kind of wireless connectivity to a central location where the quality management team can assess the quality and from there take appropriate actions to correct any problems.

It is recommended that a quality management team is appointed for each individual FOT with roles such as:

- Daily quality overview
- OEM contact person
- Subject contact person
- DAS and sensor maintenance person
- Vehicle maintenance person

4.1 On-line quality management procedures

When the vehicles in the study are on the road and collecting data it is important that it is possible to assess the status of vehicle, DAS and data without having to physically access the vehicle. In most state-of-the-art FOT wireless transfer of some of the data/status has been used. This section outlines state-of-the-art methods for performing this on-line data

4.1.1 Remote automatic upload

In state-of-the-art FOTs, different transfer techniques such as simple text messages (SMS) or GSM/3G, have been used. The transfer methods can be custom made or already available in the vehicles (but modified – for example fleet management systems for trucks sending additional status information). The use of Wireless LAN (802.11x) is likely to be more widely used in future FOTs.

In FOTs where *all* data is transferred wirelessly after each Trip or at short regular intervals, the on-line quality procedures can be more or less the same as the off-line procedures. That is, since in this case all data is available for the quality management team immediately, they can apply all off-line data quality algorithms and procedures on the full data set directly. The result of this full analysis can then be used directly to apply the appropriate action if a problem is found. Even though this is the most complete method of quality assessment on-line (compared to only summaries being sent), timing issues has to be addressed. A maximum delay from the time of actually collecting the data until it has been analysed for quality and status should be defined. If this is not done and the evaluation drags out in time, the vehicles on the field are potentially not collecting the required data. The maximum delay should be set based on the accepted levels of data loss and the length of the study. Also consider possible problems with electromagnetic interference (see section 3.4.2).

If only a limited amount of data can be uploaded wirelessly the methods of quality assessment is different. In this case (the most common so far), the data to be send has to be processed in the DAS and stored separately for upload scheduled upload.

The following is a list of example variables/measures that may be of interest to store in the vehicle DAS summary files for per trip data upload:

- DriverID, TripID, DAS-version etc (identifiers)
- Start and end times for logger and for each individual data source (compare start-up/shutdown times)
- Total count of data packets per data source and possibly per measure
- If possible, each sub-system (data source) should provide status and health information to be added to the summary
- Driven distance (if possible give vehicle odometer since this can be used to identify missing recordings from one report to the next – that is, if the DAS failed to start and did not record anything there will be no record of any problem if the odometer is not used).
- Averages, standard deviation, min and max for all measures.
- Identification of time gaps (missing data/dropouts) for each data source
- Road type (needs on-board map)
- For CAN and other vehicle bussed, include if available Error counts
- Data distributions for histograms of important measures for each trip. Using histograms gives a good overview of the distribution of data and possibility to have more fine-grained control of warning and alarm thresholds for specific data.
- For RADAR and other object tracking sensors histograms of number of targets and possibly distribution of targets are useful
- Use of different sources to verify each other – examples:
    - GPS speed and compared with CAN speed
    - Odometer compared with integrated speed
    - Time synchronisation comparison between sources using common measures
- Image verification: Video is often an important data source and also needs quality/status verification. There are several ways to do this. Here are two examples:
    - Image histogram data (possibly only over one or a couple of rows and/or columns (per camera). This can be used as an indicator of video image quality but the camera could still have moved or the sensor could have become dirty. Manual inspection is probably the safest way for validation of camera views (although time consuming).
    - Sending one or a couple of images (possibly at lower resolution) for each trip to be visually inspected on a quality management web-page
- More detailed information about evaluated systems: For the systems being evaluated in a specific FOT, it is recommended to send more information such as system status, histograms (for important measures such as Time To Collision or Lane Position) and transitional data about when warnings were issued or systems activated/deactivated. The transitional data are usually relatively small datasets compared to time-history data (for example recorded at 10Hz) and should if possibly be transmitted on-line. By sending this more

complete set of data specifically about the systems being studied, preliminary evaluation of the systems can be started quickly (while the full data has not yet been collected), and it produces a good quality evaluation platform (if such a system stops delivering good data, the entire study is immediately jeopardised).

- Depending on the amounts of data possible to transfer log-volume data may be fully uploaded (which other data such as video must be scheduled for pickup). In this case, much of the histograms and averaging is not necessary in the DAS, but can be performed off-line.

When the data has been uploaded and put into a database trip statistics can be calculated per vehicle or driver. It is recommended to use this to identify abnormal driver/usage behaviour early in the study so that if necessary drivers can be exchanged (if possible). Such situations would be if the driver does not drive at all as much as stated or under the conditions specified. This driver monitoring early in the study is highly recommended so that the study schema of the specific FOT is kept.

If an FOT is to be executed across country borders and include roaming for the wireless services, investigation of the cost/benefit of using the quality data upload systems outside of the “home country”, should be made. That is, due to high costs for wireless roaming, data upload may have to be restricted either on a service level or on a geographical level (the system does not allow wireless transfer outside of specified GPS coordinates).

It should also be noted that for DAS and infrastructure based systems, more or less the same procedure can be applied as when dealing with a vehicle. The main difference being is that one does *not* have to interfere with the subject/driver for maintenance.

For nomadic device studies the same procedures apply as for an on-vehicle integrated DAS study.

4.1.2 Automatic and manual quality checks

When the quality assessment data has been wirelessly transferred from the vehicle to the database/server, there are several ways of assessing quality. It may be tempting to do the quality assessment fully automatic, but state-of-the-art FOTs has indicated that by doing this, the risk is that the driver is contacted in cases where the error is not significant for the study or when just because the driver did something special on a Trip or two. Due to this, it is recommended that the quality assessment should be set up in different steps and that before (if) employing a fully automatic system the algorithms for the assessment should be thoroughly validated. As a first step an automatic thresholds based system can be applied for some hard and very important measures. For most data a tool (possibly browser-based) can be build which indicated problems for further analysis before driver contact. This tool should preferably be checked by a one responsible person each day.

It should also be noted that in some FOTs where the automatic quality assurance methods have been validated, the DAS itself could directly inform the responsible person about problems.

In the following an example of a process of on-line data quality check for an FOT with limited wireless transmission (not full wireless data uploads):

1. At each Trip with a specific vehicle, the DAS stores a summary information file (including items described earlier). This is done in addition to the normal logged time history and transitional data.
2. Before DAS shutdown (after trip is finished) the summary and status data is uploaded to an FTP server (or another transfer solution, such as VPN or directly into a database). The system is trying to send data until the power management system starts shutting down (e.g. 5min).
3. If data is not directly uploaded into a database, an automatic database upload application is reading the transferred files and uploading them into a database at regular intervals (or initiated by the transfer).
4. All information in the database is published on a web page with colour coded warning/error (yellow/red) on each uploaded signal (Histogram/min/max/average), if it is above/below threshold levels set for each signal.
5. This web-page should then daily be visited by someone in the quality management team.
6. In parallel, for some of the more important signals, if the error threshold levels are exceeded, an email message can be sent to responsible persons.
7. If any problems and depending on the type of problem, different actions need to be taken. (See 4.1.3). Note that is recommended that before any subject is contacted due to problem with in the study, the effects of the contact itself should be evaluated against what the problem is (and how it effects the study results) or if it is likely to be repeated. It should also be made clear who makes the decision to contact a driver in case of a problem – this is especially important when/if non-core FOT members have access to data quality information (such as suppliers having their own system quality oversight – they should probably not contact the driver directly if there is a system problem).

4.1.3 DAS and sensor maintenance

It is recommended that a specific subject/driver only has one contact person throughout the study. The process for contacting the driver should be clear and preferably there should be a list available within the project with contact information for the contact-responsible for each vehicle. All communication with the driver should then be at least initiated via this person.

When a problem has been identified and validated as far as possible without access to the vehicle, maintenance procedures will differ depending on the type of problem, study design (including vehicle ownership, etc.). The process of fixing a problem with DAS, sensing or systems in the vehicle may include the following:

1. Identification: there is a problem (see 4.1.2 above)
2. If the vehicle is brought in for replacement, in some cases a non-system (DAS etc.) equipped temporary vehicle could be made available to the driver. In several state-of-the-art FOTs an identically equipped vehicle has been given to the driver in exchange. When this is possible, it is recommended,

3. Note that the vehicles should be run through a last “system validation” test before handed over to the driver, even if only minor adjustments has been made.

If the study is run over a longer period of time (e.g. more than six months), a scheduled vehicle or DAS/sensor replacement may be an option to avoid problems (for example every three months).

4.2 Off-line quality management procedures

4.2.1 QA: Objective data

Quality assurance before data is uploaded to database

Before uploading objective data from the vehicle, a well-defined algorithm should be applied to all these data in order to verify their:

- size, to be sure that something has been recorded in the vehicle database
- state, for example for each variable, it is important to verify if the raw is completed or lost
- quality, for example, a froze value of a continuous variable

A short report must be edited or just coded (warning and/or red LED on a DAS display) if any problem is observed on one or more sensors signals. There are many ways to test each signal: for example, the algorithm can compare theoretical characteristics to real measures: maximum and minimal possible values, maximal dynamics response, consistence with another linked measure, frozen value of a dynamic parameter during a given duration.

Quality assurance of video data

To catch problems with camera failure or other video related problems, a video checking strategy should be implemented. A tool for viewing one or several images per trip can be useful. Moreover, a function to verify at least the size of video files is necessary - the size is somewhat proportional to recording duration.

Driver id verification/input

To be able to study within and between driver differences the driver ID for each trip has to be identified. It may be necessary to have a process that allows analysts to view for example one image per trip and match this with the ids of the drivers allowed to drive a specific vehicle. A vehicle may have several drivers, but maybe only one or few of them are to be included in the study. If a driver is unknown the data for the particular may have to be removed. A software tool for doing this manual identification of drivers is preferable. Be advised that some eye trackers (if available) provide DriverID functionality.

4.2.2 QA: Subjective data

A number of factors contribute to the quality of the subjective data collected by means of interviews and/or questionnaires.

In an interview situation the factors include the interviewee (in terms of, e.g., social skills, ability and willingness to respond as well as experience and knowledge); the interview situation and the interviewer (in terms of social skills; training; motivation, etc.), the content of the interview (in terms of, e.g., complexity; the sensitivity of the topics addressed, etc.) as well as the structure. If a questionnaire is distributed, the factors include the respondent (ability and willingness to respond as well as experience and knowledge) and the content and design of the questionnaire (e.g., complexity; number of questions, formulation of questions, etc.). The interview situation is affected by more factors but at the same time allow for control which contributes to higher quality data.

In order to address the validity of the data, the formulation of the questions (and possible answers) is a key issue, even more so when designing a questionnaire to be distributed to respondents. Questions must evidently be formulated in a way so that they measure what is intended to be measured. However, questions must also, e.g., be specific; not too complicated; and be formulated in terms that can be understood by the interviewee. Hypothetical questions are the most difficult questions and should be avoided. Data quality can also be improved by designing the interview/questionnaire so that it checks for consistency, i.e., the same question posed in different ways. Pilot tests must *always* be carried out in order to ensure the clarity and completeness of the questions.

Questions can be open-ended or close-ended. Open-ended questions do not supply any answer categories while close-ended questions supplies do. If close-ended, the answer categories should be as few as possible in relation to the questions; be relevant in relation to the type of question; be mutually exclusive; be reasonable and make sense; and allow the respondent/interviewee to be able to answer the question. The answers to open-ended questions will take longer to analyse than close-ended. Missing data is more common for open questions than closed. Furthermore, most often these answers must be coded which in itself may result in errors. This can be avoided by the support of a clear and consistent code key. Furthermore, in an interview situation, the interviewer can summarise the answer or group of answers, and allow the interviewee to agree or disagree and/or to comment on the interpretation. Consistency in coding can be checked by comparing several independent analysts' coding of the whole or a subset of the collected data.

Questions can also be direct or indirect. An indirect question directs the interviewee's attention to another person (or to other persons) than the interviewee and can be a way to address more sensitive questions or areas where a "true" answer may not be anticipated.

The interviewer plays an important role in collecting data in an interview situation. In order to ensure the quality of the communication, the interviewer must show the interviewee respect; must be able to listen as well as be able to communicate "active listening"; should not be afraid to wait for the answer; should have good knowledge of the issues addressed and the specific conditions; and must never show dislike, irritation, or stress. Interviewer bias, that is the influence of the interviewer on the participant's response, can be avoided by administering a questionnaire. However the interviewer may also increase the quality of the data collected by, e.g., being able to explain questions and using probing questions.

Data missing due to different reasons is a threat to the quality of the data, whether an entire interview or questionnaire is missing or the answers to individual questions are missing or answers are not readable. In addition, data can be missing due to the respondent providing an answer, or rating, outside allowed categories. In the case of a missing questionnaire or interview, efforts must be made to ensure that data collection is as complete as possible and reminders must be administered. Furthermore, overall the number of questions should be thought through, to the goal being to limit the number of questions. In addition, the number of open questions should be as few as possible in order to reduce the effort of the respondents. Some problems associated with data missing can be avoided by choosing interviews instead of questionnaires. The interviewer can, e.g., explain the question if poorly formulated and therefore not understood correctly. The interviewer can also probe for answers when open-ended questions, ensuring not only data being collected but also more in-depth information. On the other hand, the questionnaire most often allows the respondents to answer the questions at a time of their own choice, which may ensure a higher response rate. If data is missing it is important to determine if there is a bias. For instance, one should check whether missing data results from a specific group or category of participants and how this may bias the analysis of the data. As a rule of thumb, if missing data is less than 10% and is randomly distributed, the analysis may not be significantly affected.

5 Data management and storage

As state-of-the-art FOTs has proven that various types of studies demand different data models and hardware specifications, this section will not describe a generic solution for all types of FOTs.

The following sections will focus on large FOTs where hundreds or thousands of hours of raw data are collected. The considerable amount of database data and, if applied, video data will test hardware to the limit. Effective data storing and index within the database as well as cost effective hardware is vital.

A smaller FOT might use a less complex solution, using a Microsoft Access database (still 5.1.1 might apply) on a shared disk or even a spread sheet as Microsoft Excel.

This chapter will also include guidelines to handle confidential and proprietary data in the database as well as aspects of physical and logical access.

5.1 FOT database design

5.1.1 Database design

The FOT specific demands make it difficult to present a generic data model, but generally different FOTs will face some of the same challenges. The two suggestions for database models presented here address some of these issues. They are however not complete, and should only be used as a suggested base for further development of an FOT specific complete data model.

FESTA DB model 1
[See *Annex B.*]

State-of-the-art FOTs have had a de-normalised layout depending on the data collected. Strategy: a measure equals a table column in the database (see the measures and sensors matrix). To avoid keeping all columns within the same table, splitting the data into several main table(s) and sub system tables (sub systems as LDW or eye tracker) is necessary.

Pros: since measures are split into different tables there will not be too much data in each table which makes it easier from a performance point of view.

Cons: the design may end up with lots of database tables and it may be difficult to keep track on where measures are stored. Additionally, it will not be possible to activate/de-activate measures during the study.

FESTA DB model 2
[See *Annex B.*]

Strategy: instead of storing data in different tables, a “value table” can be created to store any value. Instead of using a column as the reference, measures are stored as rows in a measure table. Note that this approach has not been used in any published large scale FOT. The usefulness and performance of this approach is still to be proven for FOTs.

Pros: The measures as can be used as references for other tools (such as a quality assurance system), and that measures can be added/removed or activated/de-activated easily.

Cons: The value table can grow huge and care must be taken to ensure response time. To store 1000 hours of data, 200 measures at 10 Hz will result in 7,200,000,000 database rows. Different databases deal with this issue in different ways. So-called partitioning and/or block compression are two known methods.

Trip id and time are cornerstone indices in the database designs. This means that sensor data must be time stamped when inserted in database. State-of-the-art FOT relational databases use a common sample rate to ensure the validity of trip id and time. If different sample rates are needed within the same FOT database, the different datasets should be organised into different tables. Furthermore, data with frequency differing from the default one (e.g. 10 Hz) should be clearly marked as potentially incompatible with the main data. If the need to join this data with the main data arises, it is suggested that data from the deviant table is extracted, re-sampled and inserted into a table with the common sample rate. Potentially, this is only done for performance indicators derived from data sampled at a higher frequency.

Transitional data can be stored separately in tables with only the data when transitions occur. For instance store only data about gear change when it actual happen with the relevant data type. Although the trend is, despite huge storage overhead, to handle transition data the same way as measure data to simplify analysis.

Events data can be described as shortcuts or pointers to specific events within the database. *Example*: when during data analysis *break events* are found, the found events can be inserted in the “events table” (with data type ‘break event’). It is up to the FOT to define what an event is and the level or algorithm that defines it.

Using annotations is another way to create shortcuts to events in the database. Annotations can be defined manually when watching the video. As the driver does something that might be interesting (using a cell phone or making a left turn), the analyst can record start and end of the event and maybe comment the situation.

There are two types of background data in the model; driver background and vehicle background. Relevant data should be defined in the FESTA measures and sensors matrix. Driver information is stored in the “*driver table”* but any data to identify the driver should be kept securely and separately. Data to classify the vehicle is stored in the “*vehicle*” and “*vehicle properties*” tables. Subjective data as driver interviews are stored in appropriate tables also.

5.1.2 Adding different data characteristics to the database

Subjective data into the database

To reduce errors automatic transcription of subjective data is always preferable, when possible. However the solution has to guarantee in particular the data synchronisation, since a mistake in the driver information interpretation could be recovered, if the raw data is still available, but a missed synchronisation cannot, in general.

Then for example it is important that the comments during the driving will be stored using a voice recorder, not on paper (this solution will also disturb the driver, since it requires to stop the vehicle), to allows also the automatic time synchronisation of the message.

There are several ways of storing and using audio protocols in the database. First, the audio file can be stored (MP3 compressed or in some other way), and if recorded in the vehicle during trips it can be stored synchronised to the actual event. Secondly, transcription of the voice message can be used, but will require huge amounts of work. Voice data require relatively large amounts of disc space, but, after the transcription, the audio file could be stored on backup discs and removed from the main storage.

Time history data into the relational database

Most of the data in a typical FOT is stored in time history tables. For *FESTA DB Model 1* it is important not to create tables with too many columns. For both models, but *FESTA DB Model 2* in particular, specific database tools and functions could be considered:

Table partitioning: The aim is to help the database engine find requested rows as fast as possible. Instead of storing all data for a table in a single data file, partitioning of the table can be used. The method allows spanning the table over multiple files, each with its own defined logic in it. *Example*: one partition per trip could be created automatically. However, as too many files are not desirable either, partitioning on vehicle type or vehicle id should be considered instead.

With or without table partitioning multiple files for a table can still be specified and spread to different disks to increase performance.

Block compression: Block compression could be applied either using software or hardware. The idea is to use pointers to data instead of storing vast amount of identical values. This approach is extremely powerful when dealing with repetitive data and makes it possible to decrease database storage significantly.

Pre-defined indices: Some important performance indicators may already be defined early in the project, and those can be used to design specific indices. Min/max values or an index for a value that will be used frequently are typical examples. When applying indices to values, be aware of the size and characteristic of the data; some measures will not be suitable for an index.

Time history data can be of different database types (char, number, date, enum, boolean). Once again different vendors may implement this in different ways. The database type for each measure should though be described in the FESTA measures and sensors matrix.

Background data into the database

Background data can be essential to understand driver behaviour and the information should be easily accessible. The driver background data may include age, km per year, income, social status, etc. The database can also store data about the vehicle, such as year model, installed safety systems, and mileage.

Adding tables to the database

Tables are initially created manually based on information in the FESTA measures and sensors matrix. Any change to the database design must be documented thoroughly.

5.1.3 Events/classifications in the database

An FOT database can consist entirely of events if a triggered data collection approach is adopted (as opposed to continuous data collection). Thus, the database could simply be an event database in some FOTs. In other FOTs, where data collection is continuous for longer periods of time, the ability to find and classify events of interest is of central importance. As described in section 3.2.12 above, the classification and use of "events" (classified time periods) as a means of partitioning data according to observed events or changes in driving circumstances is an important aspect of FOT analyses.

Some events are straightforward and simple to identify, for example hard braking defined as peak deceleration > 0.7g, and may not need to be saved as a discrete or transition variable. However, many events involve a considerable amount of effort to find and validate, such as sorting through true- and false-positive curve speed warnings or finding critical rear-end events. As the data processing effort to achieve some event classifications is considerable, they are worth saving into a discrete variable database or index to facilitate data query and analysis. Event pointers should be saved to speed up analyses and can be used in combination to describe more complex situations with multiple events.

The following is a suggestion for naming convention of data segments. For an FOT "X", when the "ith" Subject (Driver) turns on the vehicle key at a time Ti and turns it off at a time T2, during a day "YY/DD/MM", then the file name could be: "X-i-YYDDMM-T1T2". For example: "SL-15-092303-1156-1234.mat" is the 15th driver of Speed limiter FOT who drives on March 03 2009 between 11:56' and 12:34'. We can add to these names a suffix "O" for objective data, "S" for subjective data and "Vi" for video recordings (i: 1 to 4 for camera identification).

5.1.4 User data spaces and data sharing

It is vital to keep track on where data are created and manipulated. The collected data should be kept as original data in a read only space. It might only be the uploader that can insert new data and FOT database owner to update or delete.

As the analysis work begins there will be need for the analyst to store relevant data in a private user space. If the material is relevant for other users of the database, there must be a way to share this data in a project internal space. If the study is collaborative study with many stakeholders it should also be possible to share the data within the organisation.

The approval process for data sharing should be described in the suggested system overview document (see *Annex A*) and basic meta description of the data is needed (as data origin and function/method/algorithm applied to the data).

Some of the data could be of public interest and therefore exported or accessible via web interface. Although sharing this data must be approved from all stakeholders in the FOT and/or on an aggregated level.

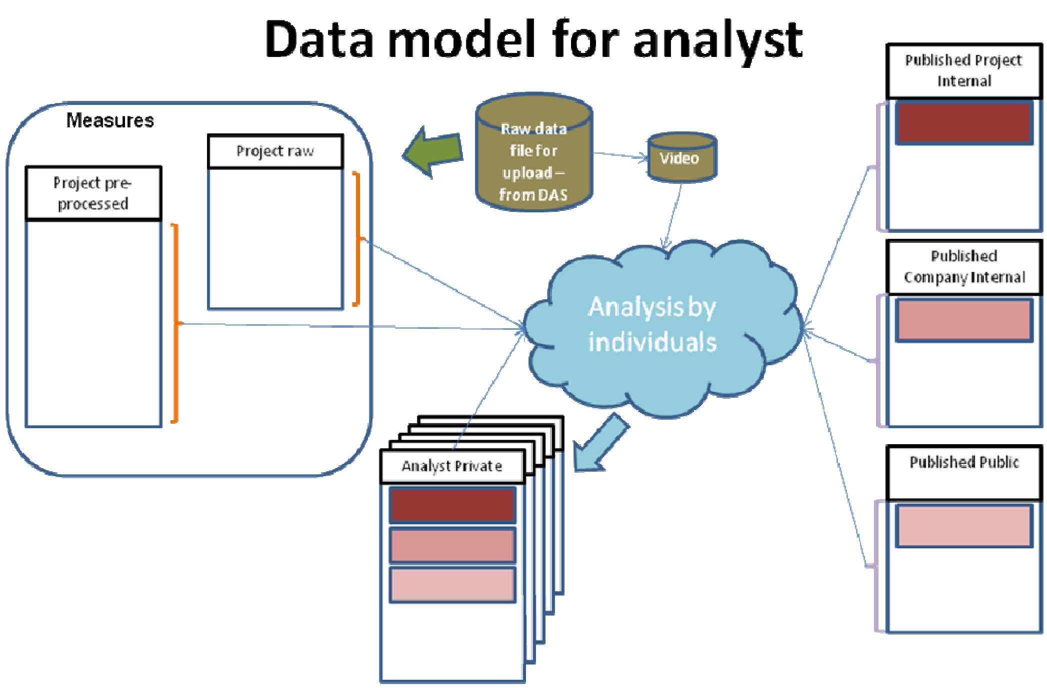


Figure 9: Data structure for private, organisational and public data

5.1.5 Hardware

Designing hardware architecture for this type of project is a bit ungrateful. The database server will rest most of the time and then run for 100% of its capacity during analysis. This makes it difficult to estimate cost for hardware and analysis must be done with user expectations of time to search and extract the data.

UMTRI has proven that mid range servers, focusing on CPU and memory, can be used and answer the demand of the analysts. If the supporting organisation can provide flexible solutions, such as server virtualisation and/or clustering, the FOT study when running analysis on the data can be prioritised. When the project ends, machine usage can be set to a minimum until a subsequent study needs to use the data. The downside is of course that this infrastructure is a major investment and can be complex to administer.

5.1.6 Storage

Disk cost has decreased significantly last years and very fast and reliable disks can be used even with a limited budget. But as storage cost drops, FOT studies tend to store more raw data from sensors and video with higher quality and better frame rate. The database and video data can be stored using large internal disks or external USB disks. But in most cases storage at some kind of disk cabinet, NAS (Network Attached Storage) or SAN (Storage Area Network) is most appropriate.

A storage setup with some kind RAID configuration should be considered, in order to be better prepared if a disk crashes or some data blocks are corrupted. RAID configurations usually have a higher disk cost (as some disk space is used for redundant data storage), but the insurance is worth it.

The database should use faster disks than the file server and disk cache is a good way to increase the performance of the system.

5.1.7 Risk management

An FOT study can generate huge amounts of data; especially when video is used, and the management must decide on the need for backup and acceptable downtime for recovery of the FOT database. It is up to the steering committee of the study to have a documented backup policy and crash recovery strategy. Some studies can handle downtime up to a week as some must be up and running as soon as possible. The backup strategy might vary during the lifecycle of the study (collecting phase, analysis phase). If so each phase and strategy should be documented.

Disaster recovery (when local database and backup hardware are destroyed) must be taken care of and there should be an offsite backup of the data.

5.2 Implementation of database and other data storage

5.2.1 Database

Storage of all data but video should be stored in a relational database, supporting ANSI SQL.

Reduce space
To reduce the number of columns only core data could be stored. Then values can be calculated when running queries. This strategy may reduce disk space but the database server will have to process more data than necessary. An opposite strategy is to simplify the work of the analyst by pre-calculating values and store them in the database before the analysis.

Null
The implementation must also consider what to do with data loss from a sensor. It is up to the DAS to manage this but it is important to be aware of the decision when analysing the data. If a sensor gives no data, a null value can be inserted. State-of-

the-art FOTs suggest using the last known value makes analysis easier. The problem with data that is actually not valid has to be dealt with.
If a sensor *never* has had a value (when starting the vehicle) NULL might be used to describe this scenario.

5.2.2 *Video data storage*

In state-of-the-art FOTs two ways of storing video data have been identified. There are mainly two different strategies available:

- *File server*: Store video file on a file server.
- *Database*: Store video directly in the database where each frame is stored as a jpeg image as BLOBs (Binary Large Objects). Another option is to store the complete video as a binary file in a BLOB.

The later will demand even more out of the database regarding both performance and backup strategy but if the infrastructure is available this it might be applied. The positive effect is store data at one place and therefore lower administration.

5.2.3 *Hardware*

If video is not stored within the database it is recommended to separate the database and video file server in order to configure the hardware individually. The database is CPU and memory demanding as the file server can have less of that and focus on file caching.
As described in 5.1.5 virtualisation is a powerful design for an FOT database and can give the possibility with “CPU and memory on demand”.

An organisation might outsource the system operations but be aware of the cost for network (bandwidth) and backup that can be extensive. The positive effects are to lower start-up cost, professional operations and server and database administration. To validate outsourcing as an alternative four areas could be covered: Cost, Benefit, Flexibility and Risk.

5.2.4 *Distributed system at various locations*

It is strongly recommended that the database is not distributed. For the database, use a single common database. This scenario might happen in a collaborative FOT with stakeholders in different cities and/or countries. The stakeholders must aware the consequences of doing this and network bandwidth will be crucial to be able to share the data. For video storage, also other options can be considered.

When uploading data it is not advisable to do this over network directly to the upload server. FTP or any other file sharing protocol can be used to upload the data to a specific server and from there do the upload of the data to main server.

When analysing data in the database it should work out quite fine accessing the database remote. Although large responds might take a while depending on the network speed between the sites. When extracting a large subset of data it is advisable to store the dataset locally, compress and then download using ftp.

Viewing video will be much more network demanding and one option is to let the organisations store video data locally on their sites (since they will probably be most interested in their own data). If a user at site A wants to view video stored on a different site B, the data could be download with pear to pear and then stored locally at site B. If a third user at site C wants to see the same clip the data can now be downloaded from both site A and B.

5.2.5 *Physical access*

Physical access as well as approval for access to the hardware should be documented in the suggested system overview document (see *Annex A*). If the hardware is located in a shared storage location, detailed documentation on routines from the supplier should be obtained.

5.2.6 *Logical access*

Logical access as well as approval for access to the database should be documented in the suggested system overview document (see *Annex A*).

Role based access is advised where any user to a certain role of the database obtain certain access. This also applies to the operating system. This makes it easier to create a new user account and to replace members within the FOT study. Any FOT must define the roles and permissions of the database. Roles can be:

Table 9: Roles and permissions for an FOT database.

Role	Comment
Database administrator (DBA)	Unrestricted access to the database.
FOT database owner	Unrestricted access to FOT database data and permitted distribute role access to users.
Uploader	Allowed to insert and update data in the FOT database.
Analyst	Allowed to read data from the FOT database and to manipulate data in private user space.
Publisher	Permitted to insert/update/delete data in shared user space.
Web application	Permitted to read data from specific user space containing aggregated data.

5.2.7 *Personal integrity and sensitive data*

Driver data must be stored with access restrictions, defined by the project partners. Also in a collaborative study (with different organisations and/or companies) some data may be classified as sensitive by a partner or even by a supplier of measurement equipment. To solve this issue encryption could be used on specific sensor values. The database could handle this issue, as different vendors have implemented various solutions.

5.2.8 Private vs. public data

Private data should be kept in a private “user space”. Different database vendors use different terminology for this solution. SQL Server and MySQL use the term “database”. Oracle defines “schemas”. Some limitations on the data space (time, amount of data) can be useful and a user log should be enabled to force the user to describe the data. If appropriate, the user could then allow different levels of access to the data using database privileges, or make a copy to a common user space accessible for all users or within an organisation. Public data could be exported to various data formats (DBF, XML, XLS, Matlab) using built-in database tools.

If data should be accessible on a web page with direct access to the database it should probably be aggregated and the amount of data sent from server to client must be considered. It is even more vital that this data is well protected, and that a dedicated database server is used to separate this data from the project data. There are numerous ways of synchronisation to simplify administration.

5.2.9 Backup

An FOT database backup strategy should be based on “acceptable downtime”. Backup implementations can be very costly but the value of the data must be considered. Off site backups are mandatory for managing a disaster scenario. The majority of the data is never edited (video and raw data in the database) and mirroring should be sufficient (could be defined in uploader or as a daily check). For data created in private, organisation, or public user spaces, a daily backup strategy should be applied. From a hardware perspective it is important to monitor disk health to ensure the backup validity.

Video data (file server)
There are different ways to handle this data, of which one or several can be applied:

- *Mirroring*: An identical hardware configuration to the one in production is set up as a mirror and is synchronised once a day. The positive effect is that the FOT database can be recovered quickly even in a disaster recovery, and that it is fairly simple to manage. The downside is investment, to ensure synchronisation, cost for server location and administration.
- *Tape*: Tape backup can be used as in some state-of-the-art FOTs. It is less costly than storing backup data live on a file server. Although when comparing an online disk backup will be faster to restore.
- *USB disks*: The FOT may decide to use the disks from the DAS only once, and then store and secure them at another location. This option might be applied for a small scale FOT but is not preferable for a large scale FOT. The positive effect is low initial investment and very simple to manage. Negative effect: when recovery of the FOT database is needed, the whole upload process must be performed once again. At this time the validity of all disks must be verified once again.

Database data

The backup policy must be based on the time it takes to recover data and the acceptable loss of data. All database vendors have solutions and proof of concepts for this. Mirroring and tape backup can be used as stated above.

Even though some studies may use the original logger data as backup, *any private or published data created afterwards must have valid continuous backups*.

Other

It is up to the project to decide the need for backup on external data (scripts, Matlab, metadata on manipulated material).

5.3 System acceptance

Before an FOT is launched the FOT database architecture should be reviewed by a system valuator to ensure that all requirements are fulfilled and to verify policy documents. All stakeholders in the project must be aware of the possibilities, responsibilities and shortcomings of the system.

5.4 Measures naming guidelines

It is recommended that the FOT project decides on and adheres to a set of naming conventions for measurements. The conventions should preferably be used also for other areas and applications of the project. The strategy used should be well documented and thoroughly enforced.

The motivations for a clear naming convention include (but are not limited to):

- Project-wide consistency,
- Clarity for direct understanding of used measures in analysis,
- Differentiation of non-comparable measurements,
- Avoidance of mix-ups, and
- Aesthetic appearance.

For the naming convention to be successful, it must also be intuitive or easily deductible, useful, providing added value/information, and reasonably easy to remember.

When specific measurements are named (in the matrix, in the database, in tools and scripts, etc.), references to the following are recommended:

- Indicative name (e.g. ground speed),
- Associated source (e.g. GPS, CAN, manually derived, etc.),
- Provider specific abbreviation (vehicle model code, manufacturer name, etc.),
- Sample rate (if different from a common sample rate), and
- Any other FOT specific descriptor.

When constructing the names, the compounds to include should be joined to create a single word. It is important to use a consistent way to divide the compounds. Possible strategies are: “camel case” (SomeSignal), underlines (some_signal), or hyphens (some-signal).

Examples: [GroundSpeed_GPS_1Hz], [GroundSpeedGps1Hz]

Note: In the above example, all recommended references are not used. Depending on context and FOT specific requirements all or only a subset can be used.

The chosen measurement name should not be too long and it should be readable. Abbreviations can (and should) be used only as long as they are widely accepted and well-known. A list of used abbreviations should be made and kept updated in the project.

The aim is to clearly understand what a measurement "is", where it comes from, and how it relates to other measurements. Measurements that are non-comparable should thus be named differently. In this way the risk of making comparisons between measurements that cannot be compared will be avoided.

5.5 Automatic pre-processing and performance indicator calculations in upload

It is recommended to define procedures and implementation schemes on *how* to add calculation of pre-processing and performance indicators in the upload process. These calculations should preferably be read-only for the users. The actual algorithms for the pre-processing and performance indicator calculations in this step have to be well defined and tested (on for example pilot test data), or based on previous experience. It should be noted that there is a risk of creating a false sense of validity for the further use of these calculations. The validation schemes for this data have to be as rigorous as for raw data directly from the DAS. Preferably, the calculations should be performed on a per-individual and vehicle level.

6 Analysis tools

The focus of this section is to describe only the software or different analysis *tools* that can be used in different types of FOTs, including data reduction tools, data filtering tools, and quality management tools. This section *does not* describe the procedures or analysis methods that should be applied to the data, which is described in the Data Analysis and Modelling section of the FESTA handbook.

6.1 Data classification/transcription

This section describes the basic functionality of tools for data classification and transcription of all types of data: video, numerical, and subjective. Please refer to section 3.2.12 for a discussion of types of codings.

First, the process of classification and transcription of FOT data by human analysts has to be supported by software which presents and allows the analyst to explore and manipulate the data (this is described below). Once this requirement is satisfied, there has to be software functionality which allows the analyst to enter codings which are associated with time segments. As the exact coding scheme/syntax will vary widely across FOTs, no detailed recommendations other than common practices can be mentioned here. Important software functionality:

- organising or categorising subjects into groups and subgroups
- defining any set and structure of codes, and associating software buttons and keyboard keys to each category
- editing or updating the coding scheme, even while analysing
- defining events as either a *state* event (e.g. glance left, glance right) or a *point* event (e.g. stop light)
- defining if state events are mutually exclusive or start/stop and set a default state
- defining if codes are nominal (e.g. road types) or rating scales (e.g. observer ratings of drowsiness)
- defining if codes are compulsory or optional
- logging freely written comments created by human analyst (no coding scheme)
- support for inter- and intra-rater reliability analyses

6.2 Time history and event analysis with video

This section describes the basic functionality of tools for viewing numerical time history data and the associated environment sensing data which includes video data, map data (e.g. GPS), and traffic state data (e.g. radar). Some examples of analysis tools are those that have been used for various FOTs can be seen in Figure 10 and Figure 11.

Important functionality for *visualisation and interaction* with data:

- *replay single-participant data* (numerical time-history data, video data, map data, and traffic state data) simultaneously. Multiple windows for different plots and illustrations provide maximal flexibility to arrange and resize is often

spread out on multiple computer screens as in Figure 10 and Figure 11.
Examples of visualisation:

- Video recordings synchronised with other raw data. So, it is possible to plot one or more variables (Data 1 to Data n) at a moment chosen by video image, and of course, it is interesting to make the reverse operation. The measure related to sub system tables (sub systems as LDW or eye tracker) is included in variables (Data 1 to Data n).
- Continuous variables and performance indicators which can be plotted (and zoomed) on graphs. Often plotted data has some moving marker (e.g. a dotted line) to indicate where the synchronised video timestamp matches up in numerical data.
- General information (FOT reference, subject ID, event lists)

- *aggregate and visualise multiple participants data* at once to compare flows of events. Commercial software products such as Matlab or Spotfire are highly flexible and accommodating for these purposes.

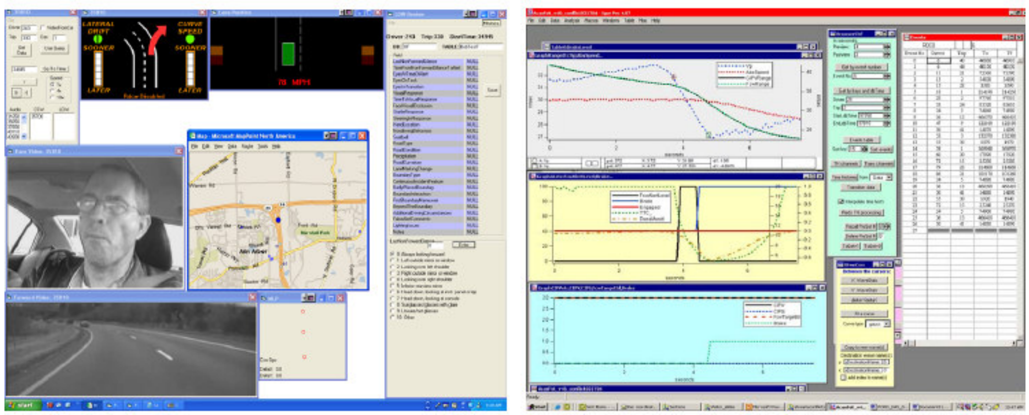


Figure 10: Screenshots of video viewer and time-history browser used in RDCW FOT. (Courtesy of University of Michigan Transportation Research Institute)

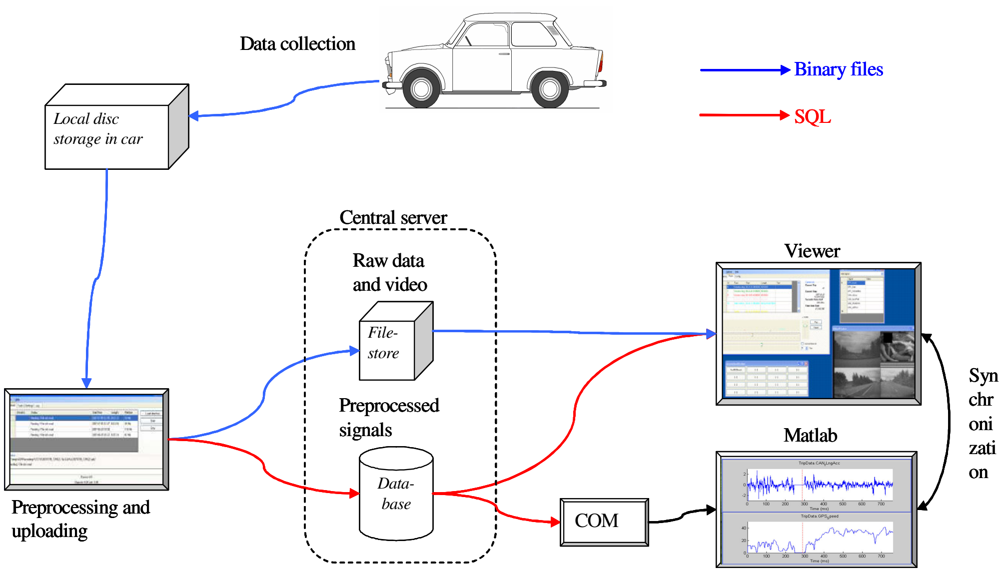


Figure 11: Diagram with screenshots of analysis software.

Interactive video viewer and annotation tool developed for the TSS FOT as SAFER Vehicle and Traffic Safety Centre.

Important functionality for *analysing data* includes:

- database query functionality (e.g. SQL)
- signal processing of numerical data
- fully customisable mathematical computation, analysis, and algorithm development functionality (e.g. Matlab), e.g.
    - Automatic calculation of performance indicators (e.g. frequency of ABS activation or average speed), or semi-automatic calculation (e.g. when the user of data analysis interface has to choose the beginning and the end of an action or a driving situation [e.g. the mean of time headway when following another vehicle])
    - Application of trigger algorithms to find events of interest (e.g. lane changes, near crashes, jerks)
- image processing of video data (e.g. machine vision algorithms to detect traffic signal status)
- grouped analysis of data (e.g. scripts)
- export of results to Excel or statistical packages (e.g. SPSS)

For subjective data (recorded in questionnaires for example), all paper versions must be wholly or partially coded into a computer database. Functionality facilitating analysis of the relationship between objective data and subjective data could be built in. In some cases, when seeing PI values, it is interesting to be able to query about the subjective driver impressions that are associated with that time segment. Alternatively, when analysing a group of drivers, having access to the associated comments, answers to questions, or focus group comments is desirable.

In general, SQL software for database queries, Matlab for computation and common statistical packages like SPSS work well as an analysis package solution. If huge datasets have to be analysed, then more specialised or proprietary solutions may be necessary. SQL and Matlab may require a fairly high level of knowledge to use, so it may be advantageous to either develop proprietary easy-to-use graphical user interfaces.

References

Campbell, K.L., Lepofsky, M., Bittner, A. (2003). Detailed Planning for Research on Making a Significant Improvement in Highway Safety: Study 2—Safety.

Curtis Preston, W. (2002). Using SANs and NAS, O'Reilly Media.

Dingus, T. A., Klauer, S. G. et al. (2006). The 100-Car Naturalistic Driving Study: Phase II – Results of the 100-Car Field Experiment. NHTSA - DOT HS 810 593. Virginia Tech Transportation Institute.

Evrard, Y., Le Maire, P. (1976). Information et décision en Marketing. Dalloz, Paris.

Gibbons Paul, L., To Outsource or Not to Outsource, August 1 2001, (http://www.cfo.com/article.cfm/2999345/c_3036063)

Geraldo, G., Differences Between On-Board Diagnostic Systems (EOBD, OBD-II, OBD-BR1 and OBD-BR2), SAE publication no 2006-01-2671, November 2006

Goldworm, B., Skamarock, A. (2007). Blade Servers and Virtualization: Transforming Enterprise Computing While Cutting Costs. Wiley.

Hampton C. Gabler et al. (2004). Use of Event Data Recorder (EDR) Technology for Highway Crash Data Analysis, Prepared for: Transportation Research Board of the National Academies, NCHRP Wep-Only Document 75

Lebart, L., Piron, M. (2006). Statistique exploratoire multidimensionnelle. 4th edition. Dunod, Paris.

LeBlanc, D., Sayer, J., Winkler, C., Ervin, R., Bogard, S., Devonshire, J., Mefford, M., Hagan, M., Bareket, Z., Goodsell, R. and Gordon, T. (2006). Road Departure Crash Warning System Field Operational Test: Methodology and Results, Volume 1: Technical Reports. UMTRI-2006-9-1. University of Michigan, Ann Arbor, Transportation Research Institute.

LeBlanc, D., Sayer, J., Winkler, C., Ervin, R., Bogard, S., Devonshire, J., Mefford, M., Hagan, M., Bareket, Z., Goodsell, R. and Gordon, T. (2006). Road Departure Crash Warning System Field Operational Test: Methodology and Results, Volume 2: Appendices. UMTRI-2006-9-2. University of Michigan, Ann Arbor, Transportation Research Institute.

Lee, S.E., Klauer, S.G., Olsen, E.C.B., Simons-Morton, B.G., Dingus, T.A., Ramsey, D.J., Ouimet, M-C. (2008). Detection of road hazards by novice teen, Transportation Research Board Annual meeting, Washington DC

Morineau, A. (1995). Aide-mémoire statistique. CISIA - CERESTA Editeur, France.

Neale, V.L., Klauer, S.G., Knipling, R.R., Dingus, T.A., Holbrook,G.T., Petersen, A. (2002). The 100 Car Naturalistic Driving Study Phase I – Experimental Design. NHTSA-DTNH22-00-C-07007. Virginia Tech Transportation Institute.

Vadala, D. (2002). Managing RAID on Linux, O'Reilly media.

Zador, P.L., Krawchuk, S.A., Voas , R.B. (2000). Automotive Collision Avoidance System (ACAS) Program: Final Report. NHTSA - DTNH22-95-H-07162.

Annex A: System overview template

Suggested starting points for system overview tables are presented here.

System overview

Name:	
System owner:	
Parties:	

System valuator:	
Validation date:	

Relation to other system(s):	

Software

1. OS	
Vendor:	
Version:	
Responsible:	
Support contact:	
2. Database	
Vendor:	
Version:	
Responsible:	
Support contact:	
3. Additional	
Name:	
Vendor:	
Version:	
Responsible:	
Support contact:	

Hardware

Responsible:	
Contact:	

Nodes	
Purpose:	
IP address:	
Vendor:	
Model:	
Support contact:	
Location:	

Backup strategy

Responsible:	

Backup routine	
OS:	
Database:	
File server:	
Other:	

Restore

Worst case scenario

Physical access

Approver:	
Other contact:	

Logical access

1. OS		
Role	Permission	Approver

2. Database		
Roles	Permission	Approver
DBA		
Uploader		
Analyst		

3. Spaces	
Description:	
Approver:	

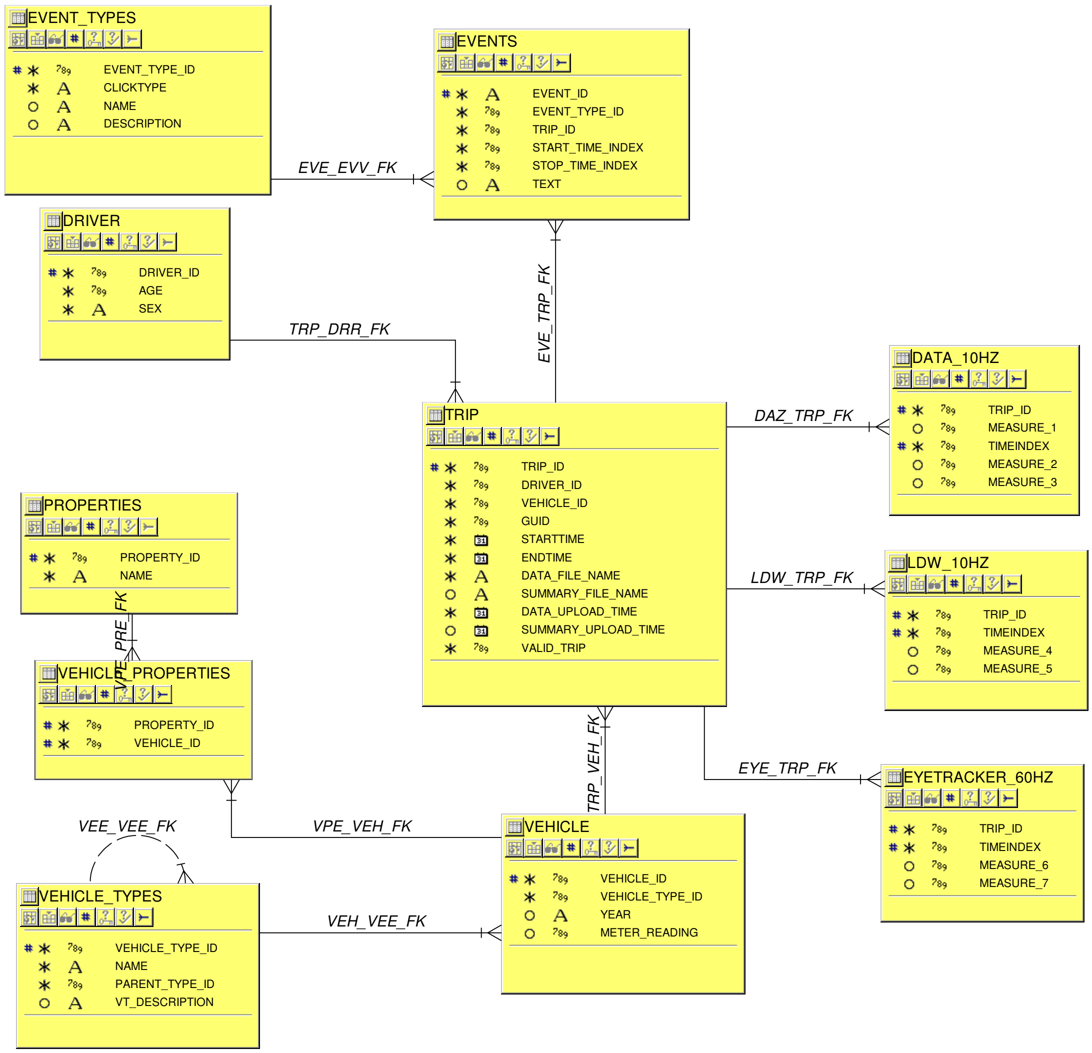
Maintenance

Maintenance manager

Maintenance plan

Annex B: Example database models

FESTA DB Model 1



FESTA DB Model 2