SEVENTH FRAMEWORK PROGRAMME
THEME 7 TRANSPORT

Grant Agreement Number:	233597
Acronym:	PROLOGUE
Full Title:	**PRO**moting real **L**ife **O**bservations for **G**aining **U**nderstanding of road user behaviour in **E**urope
Funding Scheme:	Collaborative project small scale
Project Co-ordinator	SWOV Institute for Road Safety Research
Duration:	01/08/2009 to 31/07/2011
Project Website:	www.prologue-eu.eu

Data collection, analysis methods and equipment for naturalistic studies and requirements for the different application areas

Deliverable D2.1

Authors: Ruth Welsh, Steve Reed, Rachel Talbot, Andrew Morris (Loughborough University)

Status: Final Draft submitted to EC

Date: 31 May 2010

Dissemination Level: Public

Project Consortium

1	SWOV Institute for Road Safety Research (project co-ordinator)	NL
2	CERTH/HIT Hellenic Institute of Transport	GR
3	KfV Kuratorium für Verkehrssicherheit	A
4	Loughborough University	UK
5	Or Yarok	ISR
6	Nederlandse Organisatie voor Toegepast Natuurwetenschappelijk Onderzoek - TNO	NL
7	TØI Institute of Transport Economics	NO
8	Test and Training International Planning and Service GmbH	A
9	Universitat de València	ES

Document Reference

Welsh, R., Reed, S., Talbot, R., Morris, A. (2010). Data collection, analysis methods and equipment for naturalistic studies and requirements for the different application areas. EC FP7 project PROLOGUE Deliverable D2.1. Loughborough University, Loughborough, UK

Acknowledgements

Description
We would like to thank the Virginia Tech Transport Institute for hosting a technical visit and providing valuable information about their experiences of conducting Naturalistic Driving studies

Abstract

Naturalistic driving observation is a relatively new method for studying road safety issues, a method by which one can objectively observe various driver- and accident related behaviour. Typically, participants get their own vehicles equipped with some sort of data logging device that can record various driving behaviours such as speed, braking, lane keeping/variations, acceleration, deceleration etc., as well as one or more video cameras. In this way normal drivers are observed in their normal driving context while driving their own vehicles. Optimally, this allows for observation of the driver, vehicle, road and traffic environments and interaction between these factors.

The main objective of PROLOGUE is to demonstrate the usefulness, value, and feasibility of conducting naturalistic driving observation studies in a European context in order to investigate traffic safety of road users, as well as other traffic related issues such as eco-driving and traffic flow/traffic management.

The current deliverable aims to develop an inventory of the current and appropriate data collection and data analysis equipment for naturalistic observation studies together with a theoretical analysis of the requirements for different application areas. The deliverable also discusses data quality issues and top level data base management requirements. Among the reviewed literature, maximal use is made of the extensive knowledge and experience that comes from the EU projects FESTA and EuroFOT, the 100car study and the SHRP2 preparatory safety.

Table of Contents

Executive Summary

Naturalistic driving observation is a relatively new method for studying road safety issues, a method by which one can objectively observe various driver- and accident related behaviour. Typically, participants get their own vehicles equipped with some sort of data logging device that can record various driving behaviours such as speed, braking, lane keeping/variations, acceleration, deceleration etc., as well as one or more video cameras. In this way normal drivers are observed in their normal driving context while driving their own vehicles. Optimally, this allows for observation of the driver, vehicle, road and traffic environments and interaction between these factors.

The main objective of PROLOGUE is to demonstrate the usefulness, value, and feasibility of conducting naturalistic driving observation studies in a European context in order to investigate traffic safety of road users, as well as other traffic related issues such as eco-driving and traffic flow/traffic management.

The current deliverable aims to develop an inventory of the current and appropriate data collection and data analysis equipment for naturalistic observation studies together with a theoretical analysis of the requirements for different application areas. The deliverable also discusses data quality issues and top level data base management requirements. Among the reviewed literature, maximal use is made of the extensive knowledge and experience that comes from the EU projects FESTA, TeleFOT and EuroFOT, the 100car study and the SHRP2 preparatory safety.

Data Acquisition Equipment

Data acquisition is a hugely complex field requiring knowledge in many different subjects to ensure success in a study; these fields are diverse and range from electronics to fabrications right through to computing and programming. This document provides a base of knowledge gained from previous trials and aimed at meeting the set of proposed variables determined in D2.2. The information included cannot always be prescriptive as a large number of variables will always exist when designing and developing data acquisition equipment, as such and where relevant, experience and good practice has been provided.

Data storage and Management Methods

In-vehicle data storage requires a capability to store adequate amounts of data and compression algorithms could be used to maximise capacity though these may increase the likelihood of error or malfunction. Data storage devices should be inconspicuous but easily accessible for data retrieval. Data has, in previous studies, been retrieved via removable technology (e.g. USB hard drives, Flash cards), via direct download from the vehicle or via wireless technology. Rigorous data **uploading and back-up** procedures are essential to minimise data loss. This also includes data verification.

The literature suggests that resulting **databases** should be relational and support SQL programming. This enables rapid access to the data stored and provides features for more advanced scripts such as evaluating trigger levels of vehicle dynamics. Video data can be stored on a file server or within a database with each frame being stored as a JPEG image or BLOB or the complete video can be stored as a binary file. Careful consideration should be given to the storage capacity of the database in advance to ensure that all relevant data can be stored throughout the course of the study. This should not be underestimated.

The database should be regulated according to **performance, reliability and access.** The fundamental concepts of performance are response time and throughput. The performance requirements guarantee that the database end users can operate as efficiently as possible and that incoming data is handled at feasible rate. Limitations on performance can be, for example, peak hours of use and the number of analysts as well as the amount of incoming data. The availability requirements aim to provide the conditions that will guarantee that the access to the database is as wide and constant as possible. The objective is to try to anticipate possible downtime and service interruptions, in order to predefine recovery strategies. Security requirements for the database environment need to be fulfilled in order to guarantee the common goals of protecting confidentiality, integrity and availability of information. Common security risks are incorrect handling or storing of usernames and passwords, malicious programs and intrusion attempts.

It is recommended that **data quality** should cover the whole flow of the data to include each of the following stages and each of these is discussed in detail in this deliverable;

Data collection

Data transfer

Data storage

Database quality control

A number of functions need to be provided by and software chosen in order to support **data analysis**. These include

- Database query functionality (e.g. SQL)
- Signal processing of numerical data
- Fully customizable mathematical computation, analysis, and algorithm development functionality, automatic or semi-automatic calculation of performance indicators, and 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 results function to tabular format or statistical packages.

Many studies report developing custom software to meet analytical needs. Perhaps most importantly is the need for data reduction in order to derive data to be analysed. This is necessary since the “core” data set itself comprises many hours of video footage. Basically, the **data reduction procedures** lead to the extraction (through data mining) of specific events using the Data Analysis Reduction Tool (DART) which is used to generate queries generated in SQL. The end product (i.e. the data to be analysed) comprise specific signatures that indicate that an event in the driving process has occurred. The signatures themselves are identified through the application of specific criteria (acceleration/deceleration, swerve, etc) that are applied to the data that are derived from the data loggers. The experience that VTTI have with this procedure should be exploited and training opportunities established.

As previously mentioned, naturalistic driving observations represent a relatively new methodology, but one that is continuously evolving. It is therefore likely, that even during the course of the Prologue project new developments will occur. This being the case and such developments becoming aware to the project team, this deliverable will be subsequently updated during within the Prologue life span as appropriate.

In particular it should be noted that further collaboration will be undertaken with the SHPR 2 project as PROLOGUE progresses. Any further insights learnt from SHRP2 will be incorporated into later PROLOGUE deliverables and considered fully in future recommendations fro ND studies.

It is intended that this deliverable be considered later in the project alongside D2.2 within WP5 when guidelines for conducting Naturalistic Driving Studies are developed.

1 Introduction

The term "Naturalistic Driving" refers to a relatively new technique involving the unobtrusive collection of driver behaviour in relation to the driving task in naturalistic settings. The methodology generally involves drivers using their own vehicles on a day-to-day level and their usage and driver behaviour is recorded usually by video camera. The data that is captured using this method can then be analysed in a number of ways. Perhaps the most useful aspect of the data recording is that it allows researchers to analyse driver behaviour in relation to various "critical incidents" that may occur during particular journeys. For example, a driver may need to swerve and decelerate rapidly in order to avoid a collision. The data recordings allow insight into the driver behaviour this preceded the critical incident and issues such as distraction and inattention, which may have been a factor in creating the critical incident, can then be verified. Collection of these data on a wide-scale will ultimately facilitate the development of countermeasures that can prevent crashes where issues such as distraction and inattention but also other causal factors (such as fatigue) are prevalent. For example, driver warning systems could be implemented into vehicles that are capable of recognising when the driver is indulging in inherently unsafe behaviour. Therefore, the results will lead to a better understanding of road safety and help to achieve an intrinsically safe road transport system by improving safety through improved in-vehicle technologies, development of self-explaining roads and advances in driver training techniques.

Naturalistic Driving observations provide information that would be difficult to obtain otherwise. For well known risk factors such as distraction, inattention and fatigue, naturalistic observations would provide reliable information about their prevalence and their true relationship with crashes, i.e. the actual risk level. Other issues for which naturalistic observations would be an ideal method include

- The effect of road design characteristics, or weather conditions on the interaction between driver and vehicle;
- Comparing the driving style of specific road user groups, e.g. novice drivers, elderly;
- The prevalence of mobile phone or other in-car information devices and the relationship with particular behaviour patterns or crashes;
- The effect of passengers on distraction, particular driving behaviour or incidents/crashes;
- The interaction between motorised vehicles and vulnerable road users

The main objective of PROLOGUE is to prove the feasibility and usefulness of a large-scale European naturalistic observation study. The project is aimed at road safety researchers and other stakeholders including car industry, insurance companies, driver training and certification organisations, road authorities, and governments. Whereas road safety is the main motive, the project will also look at the relevance for environmental issues, e.g. CO_2 emissions, and traffic management.

Based on inventory studies, a series of small-scale field trials, and close involvement of user groups and stakeholders, PROLOGUE will result in recommendations and an outline for a large-scale naturalistic study, dealing with research questions, methodology and technology for data collection, data storage, data reduction, data mining and data analysis.

The purpose of task 2.1 of PROLOGUE is to develop an inventory of the current and appropriate data collection and data analysis equipment for naturalistic observation studies as well as a theoretical analysis of requirements for different application areas.

Based upon knowledge of previous and on-going naturalistic observation studies and other studies that have collected large-scale on-road data of road user behaviour, a review of data acquisition, data management and data analysis equipment and techniques has been conducted for this Deliverable. The main input comes from reviews of studies that have already used instrumented vehicles to collect data on both driver and vehicle behaviour (road and other road users) by different types of sensors.

Knowledge about reliable and consistent methods for managing huge amounts of data including data quality and completeness checks and requirements has warranted special attention. Particularly relevant in this regard is the 100-car study which was conducted by the Virginia Tech Transportation Institute (VTTI). In this ground-breaking study, 104 fully equipped cars were driven in Washington DC and the Northern Virginia area for a period of one year. Ultimately, the dataset included around 2 million vehicle miles, 43 thousand hours of data, 241 drivers. The data included 15 police-reported crashes, 67 non-reported crashes, 761 near-crashes and 8295 incidents. The database necessary to store the data recorded from the study was vast.

The 100-car study and the on-going safety studies of the VTTI in the USA including the SHRP2 as well as the Field Operational Tests conducted in EU Member States are especially relevant to this review since between them, they represent the state-of-the-art with regard to the optimal techniques and procedures for capturing the types if data required in a Naturalistic Driving methodology. The review also includes an inventory of which influencing factors and conditions (for example, status of vehicle, type of road, weather conditions, road conditions, traffic conditions, driver status) should be known and how they should be established, which metrics (both longitudinal and lateral) are necessary to collect data for and which sensors are required for obtaining the data for the different application areas.

It should be noted that further collaboration will be undertaken with the SHPR 2 project as PROLOGUE progresses. Any further insights learnt from SHRP2 will be incorporated into later PROLOGUE deliverables and considered fully in future recommendations fro ND studies.

2 State of the Art Review Methodology

A state of the art review of previous and ongoing naturalistic driving studies was undertaken to identify the equipment and techniques which have been used in relation to

- Data acquisition
- Data storage and management
- Data analysis

A number of potential data sources were identified including published papers and reports, naturalistic driving seminars, current EC supported projects involving naturalistic driving and partner expertise. As Field Operational Trials (FOT) use naturalistic driving techniques, some FOTs were included in the review.

A reference list compiled for Task 1.1 was used as a starting point. Each paper was sourced and reviewed however it became apparent that although containing some useful information, the majority of the journal and conference papers focused on the specific methodologies used rather than on technical requirements. It was therefore decided that the focus of the review should be study reports as these contain much more detail. These were identified through references included in the journal and conference papers already identified and by looking at institutions websites that were known to have conducted naturalistic driving studies. In addition to copies of presentations given at The First Human Factors Symposium on Naturalistic Driving Methods & Analyses (2008) and the Fourth Safety Research Symposium (2009), both held in the US, were reviewed in order to identify key studies and institutions.

Detailed publically available reports for the following studies were identified;

- 100 car study (VTTI, USA)
- Naturalistic lane-changes (VTTI, USA)
- Road Departure Crash Warning System FOT (UMTRI, USA)
- Test Site Sweden Field Operational Test (Safer, Sweden)
- TAC SafeCar Project (MONASH, Australia)

The FESTA deliverables were also identified as key data sources, specifically the FESTA handbook.

Other studies that included some element of naturalistic driving that were reported in journal and conference papers included;

- Truck studies – long and short haul (VTTI, USA)
- Eco-driving (Belgium and Greece)
- Teenage driving (University of Iowa, USA; OR YAROK et al, Israel)

As naturalistic driving is a rapidly developing field and studies described in conference papers, journal papers and final reports will have been conducted several years previously, the review aimed to examine current studies. Several studies were identified during the review process and these included:

- SHRP2
- TeleFOT
- EuroFOT
- Interaction

- 2 be Safe
- DaCoTA (start date delayed to Jan 2010 so no outputs available)

The project websites were examined for relevant publications, however none had publically available details about the technical specifications of the studies. Alternative ways of seeking information about current projects had to be sought. Two approaches were used. One was to ask partners to give their expert input about the technical requirements of other studies that they are currently working on as part of the deliverable review process. The other was to organise a technical visit to an institution already expert in conducting naturalistic studies.

Representatives from Loughborough University in the UK and TNO Human Factors in The Netherlands attended meetings and workshops at the Virginia Tech Transportation Institute (VTTI, Blacksburg, Virginia, USA) on 30th November, 1st and 2nd December 2009. VTTI was thought to be the most appropriate institute to visit as they initiated the first ever Naturalistic Driving project – the 100-car Study. This study was undertaken with the primary purpose of collecting large--scale naturalistic (real-world) driving data. This project captured large amounts of data relating to conflicts, near-crash and crash situations which provided vital information about driver behaviour and performance in a naturalistic driving setting. The study involved approximately 2,000,000 vehicle miles and 42,300 hours of driving data (including vehicle data as well as video). Since this ground-breaking study, several follow-on projects have been conducted by VTTI including naturalistic driving studies for truck drivers, older drivers and teenage drivers. VTTI is also involved in the large scale follow up project to the 100-car study known as SHRP2 (Strategic Highway Research Programme) whereby a fleet of over 1,000 vehicles will be instrumented.

The purpose of the visit was to gather information about VTTI's facilities and the equipment used to collect data as well as their capabilities for data storage and analysis. The methodologies, protocols, instrumentation and analytical techniques demonstrated are all relatively new to the research field and were specifically developed by VTTI for the purposes of Naturalistic Driving studies. Details of these were provided to the visitors via a series of workshops, including not only presentations on the studies conducted by VTTI, but also general technical topics: data acquisition and instrumentation, data storage and retrieval, and statistical techniques. Furthermore, the workshops also contained hands-on practical classes (using actual data from the 100-car study and the truck driver study) so that the participants could gain a full understanding of the methods used. These will be directly relevant to the PROLOGUE since the project is aiming to recommend state-of-the-art procedures and data acquisition systems as part of the project outputs.

The technical requirements most likely to be reported in the various literature was that which relates to data acquisition. Reports provided much more detail than the papers but both were more likely to describe data acquisition systems than data storage and data storage was mentioned more frequently than the technical requirements needed for analysis. The FESTA handbook provided some additional detail for data storage and analysis and this was supplemented by the information gained from the visit to VTTI.

The following will list the full bibliography of sources examined as part of this review.

2.1 Bibliography

2.1.1 Reports and Papers

Barr, L. C., Yang, C. Y. D., Ranney, T. A. (2003). An exploratory analysis of truck driver distraction using naturalistic driving data. In Proceedings of the Transportation Research Board 82nd Annual Meeting. Washington, D.C., USA.

Beusen, B., Broekx, S., Denys, T., Beckx, C., Degraeuwe, B., Gijsbers, M., Scheepers, K., Govaerts, L., Torfs, R., Panis, L. (2009). Using on-board logging devices to study the longer-term impact of an eco-driving course. Transportation Research Part D, 14, 514–520.

Dingus, T. A., Klauer, S.G., Neale, V. L., Petersen, A., Lee, S. E., Sudweeks, J., Perez, M. A., Hankey, J., Ramsey, D., Gupta, S., Bucher, C., Doerzaph, Z. R., Jermeland, J., Knipling, R.R. (2006). The 100-Car Naturalistic Driving Study. Phase II – Results of the 100-Car Field Experiment. Report No. DOT HS 810 593. National Highway Traffic Safety Administration. Washington D.C, USA.

Dingus, T. A., Neale, V. L., Klauer, S. G., Petersen, A. D., Carroll, R. J. (2006). The development of a naturalistic data collection system to perform critical incident analysis: An investigation of safety and fatigue issues in long-haul trucking. Accident Analysis and Prevention 38(6), 1127-1136.

FESTA (2008) Data Requirements for FOT Methodology. Deliverable 2.2 of the EC FP7 project Field opErational teST supporT Action.

FESTA (2008) FESTA Handbook Version 2. Deliverable 6.4 of the EC FP7 project Field opErational teST supporT Action.

Hanowski, R. J., Hickman, J. S., Fumero, M. C., Olson, R. L., Dingus, T. A. (2007). The sleep of commercial vehicle drivers under the 2003 hours-of-service regulations. Accident Analysis and Prevention 39(6), 1140-1145.

Hanowski, R. J., Hickman, J. S., Wierwille, W. W., Keisler, A. (2007). A descriptive analysis of light vehicle-heavy vehicle interactions using in situ driving data. Accident Analysis and Prevention 39(1), 169-179.

Hanowski, R. J., Olson, R. L., Hickman, J. S., Dingus, T. A. (2006). The 100-car naturalistic driving study: A descriptive analysis of light vehicle-heavy vehicle interactions from the light vehicle driver's perspective, Data Analysis Results. Report No. FMCSA-RRR-06-004 Virginia Tech Transportation Institute, Blacksburg, VA, USA.

Hanowski, R. J., Perez, M. A., Dingus, T. A. (2005). Driver distraction in long-haul truck drivers. Transportation Research Part F, 8(6), 441-458.

Klauer, S.G., Dingus, T. A., Neale, V. L., Sudweeks, J.D., Ramsey, D.J. (2006). The impact of driver inattention on near-crash/crash risk: An analysis using the 100-car naturalistic driving study. Report No. DOT HS 810 594. Virginia Tech Transportation Institute, Blacksburg, VA, USA.

LeBlanc, D., Sayer, J., Winkler, C., Bogard, S., Devonshire, J. Mefford, M., Hagan, M., Bareket, Z.,Goodsell, R., Gordon, T. (2006). Road departure crash warning system (RDCW) field operational test final report. Technical report 2006-9-1. University of Michigan Transportation Research Institute, Ann Arbor, MI, USA.

Lee, S., Olsen, E., Wierwille, W. (2004). A Comprehensive Examination of Naturalistic Lane-Changes. Final Report DOT HS 809 702. Virginia Tech Transportation Institute, Blacksburg, VA, USA.

Lotan, T., Toledo, T. (2005). Evaluating the safety implications and benefits of an in-vehicle data recorder to young drivers. In: Proceedings of the Third International Driving Symposium on Human Factors in Driver Assessment, Training and Vehicle Design. Rockport, Maine, USA.

McGehee, D. V., Raby, M., Carney, C., Lee, J. D., Reyes, M. L. (2007). Extending parental mentoring using an event-triggered video intervention in rural teen drivers. Journal of Safety Research 38(2), 215-227.

Regan, M., Triggs, A., Young, K., Tomasevic, N., Mitsopoulos, E.. Stephan, K., Tingvall, C. (2006). On-Road Evaluation of Intelligent Speed Adaptation, Following Distance Warning and Seatbelt Reminder Systems: Final Results of the TAC SafeCar Project. Volume 1: Final Report. Monash University Accident Research centre, Melbourne, Australia.

SAFER (2008) Test Site Sweden Field Operational Test Final Report. SAFER Report 2008:1, SAFER, Gothenburg, Sweden.

Sayer, J. R., Devonshire, J., Flannagan, C. (2007). Naturalistic driving performance during secondary tasks. In: Proceedings of the Fourth International Driving Symposium on Human Factors in Driver Assessment, Training and Vehicle Design. Portland, USA.

Sayer, J. R., Mefford, M. L., Shirkey, K., Lantz, J. (2005). Driver distraction: A naturalistic observation of secondary behaviors with the use of driver assistance systems. In: Proceedings of the Third International Driving Symposium on Human Factors in Driver Assessment, Training and Vehicle Design. Maine, USA.

Stutts, J., Feaganes, J., Reinfurt, D., Rodgman, E., Hamlett, C., Gish, K., Staplin, L. (2005). Driver's exposure to distractions in their natural driving environment. Accident Analysis and Prevention 37(6): 1093-1101.

Tijerina, L., Gleckler, M., Stoltzfus, D., Johnston, S., Goodman, M. J., Wierwille, W. W. (1999). A Preliminary Assessment of Algorithms for Drowsy and Inattentive Driver Detection on the Road. National Highway Traffic Safety Administration, Washington, DC, USA.

Toledo, T., Musicant, O., Lotan, T. (2008). In-vehicle data recorders for monitoring and feedback on drivers' behavior. Transportation Research Part C,16(3), 320-331.

Zarkadoula, M., Zoidis, G., Tritopoulou, E. (2007). Training urban bus drivers to promote smart driving: A note on a Greek eco-driving pilot program. Transportation Research Part D,12(6), 449-451.

2.1.2 Other

Presentations from The First Human Factors Symposium on Naturalistic Driving Methods and Analyses, Virginia Tech Transportation Institute, August 27, 2008.

Presentations from the 2009 Fourth Safety Research Symposium - SHRP 2, Washington, D.C, 23-24 July 2009. Available from: http://www.trb.org/StrategicHighwayResearchProgram2SHRP2/Public/Pages/Safety_153.aspx [Accessed December 2009]

2-BE-SAFE [In press] Draft version of Literature review of data analysis for naturalistic driving study. Deliverable 2.4 of the EC 7FP 2-Wheeler BEhaviour and SAFEty.

Technical Visit to Virginia Tech Transportation Institute, USA, 30 November – 2 December 2009.

2.1.3 Project Websites

Naturalistic Driving:

100 Car Study
http://www.access.vtti.vt.edu/

Interaction: Understanding driver interaction with in-vehicle technologies
http://interaction-fp7.eu/index.php

SHRP2
http://www.trb.org/StrategicHighwayResearchProgram2SHRP2/Public/Blank2.aspx

Field Operational Trials:

EuroFOT
http://www.eurofot-ip.eu/

FESTA
http://www.its.leeds.ac.uk/festa/

TeleFOT
http://www.telefot.eu/

It should be noted that further collaboration will be undertaken with the SHPR 2 project as PROLOGUE progresses. Any further insights learnt from SHRP2 will be incorporated into later PROLOGUE deliverables and considered fully in future recommendations fro ND studies.

3 Data Acquisition Equipment

3.1 Introduction to data acquisition equipment

Data logging is a massively important area to the field of science; in fact it could be argued that science is based on the data logged from experiments. Field operational and naturalistic trials are no different in concept but require a more technical approach to gather the data generated. Early systems consisted of mechanical data loggers using ticker tape or punched paper readouts for calculating accelerations and speed. Electronics moved data loggers on and these systems have been successfully used for a wide variety of scientific studies, from weather stations to road traffic monitoring.

More recently data loggers have moved into the electronic age and are now commonly microprocessor based with internal memory capacity, they generally use a combination of internal and/or external sensors to provide the data channels. Historically data loggers were generally stand alone devices with internal battery power. In this form they could record simple channels at low resolution (typically no faster than 1Hz) for long periods without being disturbed. More recently and in line with the demand for more data these loggers have moved into the field of data acquisition and telemetry based systems, these systems are ordinarily powered by an uninterrupted, ‘mains’ type supply and are connected to a more powerful computer (wired or wirelessly) for instant data download at much faster rates (up to 1000Hz is now common).

Data logging systems are becoming more common within vehicles and are now almost at the stage of becoming original equipment (OE) in some vehicles, whether for performance monitoring (Ducati, Nissan) or safety based (Mercedes). This development is in line with other commercial transport operations, such as airlines, where it is mandatory to fit and use data loggers, or Black boxes, in all flights.

The data acquisition equipment section covered here includes a state of the art review for 8 major areas of data collection. These areas combine to provide data for all major vehicle and subject based data. Information is provided on areas such as the loggers themselves along with the sensor groups and are combined with detailed information on vehicle based data, video data and event identification aids.

3.2 Data Loggers

Data loggers provide the backbone to all naturalistic and field operational trials; they provide a means of capturing all the data registered by the array of sensors.

Data logging is in itself nothing new, almost all experiments need a method of recording data, and naturalistic and field operational trials are no different, only perhaps slightly more complicated. Basic data loggers have been used in vehicles for many years from early experiments into crash performance right up to the high technology and high accuracy field of top flight motorsport. Only in more recent years has the technology been easily accessed and the costs reduced.

Data loggers used in previous naturalistic and field operational trials show three distinct groups or approaches to the major piece of equipment.

3.2.1 Supplier

The simplest method of acquiring, installing and using data logging equipment is to purchase ready made equipment. This approach has some significant advantages over the two alternatives described below although, depending on trial design, limitations could become apparent which are less easy to resolve.

A range of different companies supply data loggers in increasing complexities and costs. These companies often work on a country specific basis, meaning little exposure worldwide, although a few of the larger manufacturers are well known throughout the world. The data loggers described in this section are predominantly supplied to the motorsport industry and as such have a model range which spans from entry level (starting at approximately 100) to advanced systems (upwards of 1000).

The equipment is often ‘plug and play’ in that they are supplied ready to install with all the relevant sensors. This makes initial start up and data logging a relatively simple process; there is still a degree of mechanical installation and electronic set up, but a basic system can be installed within a day.

Limitations could arise if the study design changes and additional data channels are required. If these are simple, such as a potentiometer or pressure sensor for example then the system could cope with the additional data, however fundamental restrictions will always remain. For example; the limit of external channels is reached or a higher sampling rate above that of system capabilities is requested.

3.2.2 In-house design

The method which will provide the ideal results for data logging is to design and manufacture the data loggers in-house. This method does present some significant technical challenges as it requires an in-depth knowledge of electronics, mechanical and fabrication techniques and programming, amongst others, to make the individual components work together. This technique works towards an ideal set of requirements as these can be purchased separately and integrated into the system as a whole. For example; the best quality GPS receiver can be specified and purchased rather than relying on a component which is built to a budget – something which could be evident with bought in equipment.

3.2.3 Hybrid system

A method which sits between the two listed above is to integrate bought-in data loggers and sensor components to create a hybrid system. This method addresses the limitations and avoids the major technical work of both the bought-in systems and in-house designs respectively. This method relies on a vehicle mounted computer to synchronise, harmonise and record all the various sensor data streams into, ideally, one data output.

The computer provides the backbone to the system enabling flexibility of component specification and allowing different sensors, loggers and video systems from different manufacturers to be integrated. Expansion of the system is also possible as the computer capabilities can be upgraded more cost effectively that stand alone loggers or sensors.

The technical requirements of synchronising and harmonising the equipment are still relatively advanced, they do however remain almost entirely in the field of software and programming and avoid extremely time consuming and costly processes such as circuit board design and manufacture and mechanical fabrication.

3.3 Specialist sensors

Vehicle based sensors cover all those which record vehicle parameters. These sensors are usually mounted to the vehicle themselves and are linked to the loggers to record information on channels such as pedal position, steering wheel position, road speed, suspension movement etc.

Specialist sensors cover a large and diverse equipment group that varies considerably between trials. The sensors are so varied as to enable each trial to achieve its aim and are often tailored to a particular application, vehicle or installation requirement. It is noted in almost all cases that if you have a requirement to measure something then there is a sensor available to do this. This statement goes someway to illustrating how diverse this equipment group can be.

Due to the large numbers of vehicle sensors used for naturalistic and field operational trials it is not possible to go into every detail here. For example; the requirement to record steering movement may result in different steering position sensors from different manufacturers being used across a number of different trials, these are also often mounted differently and as such have slightly different calibration and setup specifications. To provide an overview of the kinds of sensors that have been used in previous studies see Table 3.1. This details the Sensor type, Make, Model, the study that used this equipment, the units of measurement and sampling rate and also includes a short section of notes that may provide further information.

Table 3.1 Sensor specifications

Variable or Sensor name	Sensor Make/Model	Study	Units or data format	Sampling rate (Hz)	Notes
Brake status	CAN	100 Car		3 - 10	
Configuration	N/A	Safe Car	Integer	1	Indicates enabled Safe Car system
Current speed limit	GPS VicRoad	Safe Car	Km/h	1	
Date	N/A	Safe Car	dd/mm/yy	1	
Following distance warning status	Eaton VORAD radar	Safe Car	Integer	5	0 through 2
Gas pedal position	CAN	100 Car	Integer	3 -10	
GPS latitude	GPS receiver	Safe Car	Degrees	1	
GPS longitude	GPS receiver	Safe Car	Degrees	1	
GPS speed	GPS receiver	Safe Car	Km/h	1	
Heading	GPS	100 Car	deg	1	0-359, 0=North, 90=East, 180= South, 270= West

Lateral acceleration	Internal accelerometers	100 Car	g	10	
Light intensity	Specialist sensor	100 Car		10	
Login	iButton unique identifier	Safe Car	Integer	1	Unique number per participant
Longitudinal acceleration	Internal accelerometers	100 Car	g	10	
Milliseconds since startup	N/A	Safe Car	Ms	1	
Number of Km Travelled	N/A	Safe Car	Km	1	For each participant
Radar, forward, azimuth	Eaton VORAD radar	100 Car	rads	10	
Radar, forward, ID	Eaton VORAD radar	100 Car	Integer	10	Target ID, 0-7
Radar, forward, range	Eaton VORAD radar	100 Car	ft	10	
Radar, forward, range rate	Eaton VORAD radar	100 Car	ft/s	10	
Radar, rearward, azimuth	Eaton VORAD radar	100 Car	rads	10	
Radar, rearward, ID	Eaton VORAD radar	100 Car	Integer	10	Target ID, 0-7
Radar, rearward, range	Eaton VORAD radar	100 Car	ft	10	
Radar, rearward, range rate	Eaton VORAD radar	100 Car	ft/s	10	
Reverse collision warning		Safe Car	Integer	1	0 through 2
Seatbelts warning status		Safe Car	Integer	1	0 through 2
Speed request button status		Safe Car	Integer	1	0 through 1
Speed Warning status		Safe Car	Integer	1	0 through 3
Speed, GPS horizontal	GPS	100 Car	mph	1	
Speed, Vehicle composite		100 Car	mph	3 - 10	
Sync		100 Car	Integer	10	Increas-

					ing integer for each data row
System override status button		Safe Car	Integer	1	0 through 1
System shutdown status		Safe Car	Integer	1	0 through 1
System State	N/A	Safe Car	Integer	1	Indicates logging system status
Time	N/A	Safe Car	hh/mm/ss	1	
Time	N/A	100 Car	s	10	
Time Headway	Eaton VORAD radar	Safe Car	Integer	5	
Trip Identifier	N/A	100 Car			
Turn indicator status	CAN	Safe Car	Integer	1	0 through 2
Turn signal status	CAN	100 Car		3 - 10	
User ID	iButton unique identifier	Safe Car	N3	1	
Yaw rate	Internal accelerometers	100 Car	Deg/s	10	
Lane markings, continuity, left side left line	Machine Vision	100 Car		10	
Lane markings, contunuity, left side right line	Machine Vision	100 Car		10	
Lane markings, continuity, right side, left line	Machine Vision	100 Car		10	
Lane markings, continuity, right side, right line	Machine Vision	100 Car		10	
Lane markings, distance left	Machine Vision	100 Car	in	10	
Lane markings, distance right	Machine Vision	100 Car		10	
Lane markings, type	Machine Vision	100 Car		10	

left					
Lane markings, type right	Machine Vision	100 Car		10	
Lane markings, probability left	Machine Vision	100 Car		10	
Lane markings, probability right	Machine Vision	100 Car		10	

3.3.1 Good practice

It is worth noting the drop out rates of sensors and the data loggers themselves – this can have a marked effect on data quality and reliability. Despite saying that anything measurable will have a sensor available, having too many will cause reliability problems.

As will be discussed in 3.8, off board, data enhancement may be the best way to avoid added complexity while maintaining a high level of data channels.

Another factor which can have a significant effect on sensor reliability and therefore data quality is in the installation of equipment. A review of system installation, whether it is a logging device or particular sensor, might not necessarily constitute state of the art techniques but should be considered good practice. As with all automotive applications there are numerous sources where data quality and reliability can be effected, these are diverse but include; interference (electrical, vibration), environmental (heat, moisture) or contaminants (fluids, dust), all of these can have a degenerative effect on data quality if left unchecked.

Using experience from the field of motorsport, where data loggers and sensor groups have been used extensively, and where, if anything, the strains on such equipment is greatly magnified (think in terms of heat or vibration) there is extensive reporting of good practice. This is also evident from reports of previous naturalistic and field operational trials, most of which reflects the longstanding good practice developed through motorsport

Equipment manufacturers often state how the devices should be fitted into a vehicle however this is often specific to a piece of equipment so a general review of installation good practice is included here under the headlines of the problem.

3.3.2 Vibration

Vibration is an important issue which can affect the long term reliability of electronic components and the short term accuracy of logged data. Different manufacturers will identify different tolerances but good practice suggests that if the item looks (or feels) as though it is vibrating then isolation should be considered. Modern road cars may not require much additional isolation although this depends on the mounting location or trial design. The method of isolation will depend on the situation but in areas of high vibration, very soft rubber mount may offer sufficient isolation, generally in these circumstances the softer the mount the higher the level of vibration isolation. If installation is temporary or the equipment needs to be removed on a regular basis then reusable adhesive putty or Velcro tape have sufficient security and damping qualities.

3.3.3 Temperature

Temperature specifications are, like vibration, equipment specific. However general good practice still exists. Environmental constraints in new car manufacture tends to push engine efficiencies further and as a result greater temperature ranges can be experienced. For example; catalytic converters and exhaust systems can reach temperatures of over 900 ºC under hard use. Most data logging systems can tolerate temperatures between 0 ºC and 80 ºC during use and -20 ºC to 85 ºC at idle. Ensuring a consistent temperature within these ranges is essential. Consideration has to be given to the enclosure material as plastic can deform as high temperatures and LCD screens become slow or even inoperable at lower temperatures. Regards wiring, this should be routed away from extreme heat sources and controlled within the recommended manufacturer limits.

3.3.4 Interference

The design and manufacturer of modern road cars, being predominantly a steel shell containing within it a partially damped reciprocating engine and moving over an uneven surface, induces a degree of resonance. This resonance can be measured at approximately 1Hz and could therefore interfere with sensor readings, particularly high sensitivity accelerometers. As previously mentioned in the vibration section this can be engineered out by careful isolation of components from this resonance or by applying filters to recorded data. Noisy signal traces are often attributable, simply to general vehicle resonance.

Routing cables inside a vehicle can also cause signal problems due to electromagnetic interference or crosstalk. This is caused electric or magnetic fields, generated in parallel cables, coupling and creating ‘crosstalk’ in the signal. A method of controlling this is to use twisted pair cables, these will cancel out the interfering source providing that it is relatively uniform. The result is a ‘cleaner’ signal at the data logger. Another method of reducing noise due to electric or magnetic interference is to reduce the length of the cabling. Careful logger positioning and routing of the cabling over the shortest distance will provide a ‘cleaner’ signal.

Other on-board devices can also cause interference with the loggers and sensors, the effects of these are quite wide ranging but should be explored and controlled for where they occur.

3.3.5 Moisture

Unless the test is to be conducted completely indoors then the vehicle will almost inevitably be exposed to water. This water combined with high sensitivity electronics will cause extensive problems. Protecting the loggers and sensors from moisture ingress is a major issue. Good practice dictates that, as far as possible, all equipment should be contained within the vehicle however this is not always possible when considering wheel speed sensors, external cameras or radar sensors. Good quality electrical connectors (a good rule of thumb is to spend approximately 10% of the equipment budget on these) will help prevent water ingress and will make repairs easier when they do not. Avoiding areas of the vehicle where conditions are harsh is also worthwhile; areas such as wheel housings or chassis components will inevitably be subjected to more water contamination than others. Areas where temperature differentials are apparent, such as the engine bay, may cause condensation to gather in connectors or control boxes as they repeatedly cool or warm during trials.

3.3.6 Fluids

Water is only one type of fluid that is common with vehicle testing; most modern vehicles carry on board approximately 6 different fluid types with some up to 10. These include Engine oil and coolant, brake and power steering fluid, gearbox and differential oils and specialist fluids such as for suspension systems or air conditioning. All equipment needs to be protected from these fluids as they tend to be more toxic than water. Simple guidelines should apply when planning system design or installing equipment. Avoid areas where these contaminants are present, unless of course, it is what needs to be measured and shield wiring and connectors in case of a leak.

3.3.7 Dust and other contaminants

Equally important is to protect equipment from dust and other airborne particulates, these tend to be less invasive but can cause similar problems to water ingress. Sighting loggers fitted with cooling fans away from these particulates will increase the long term reliability while careful routing of cables and connectors (unless absolutely necessary) will improve drop out rates.

3.3.8 Debris

Due to the nature of naturalistic and field operations trails the vehicle will more than likely be used on public roads, these are full of unexpected hazards that test tracks are generally not susceptible to. Debris in the carriageway is one such hazard that can cause sensor loss or damage and as such a complete loss of signal data. Sensors mounted in vulnerable areas such as the vehicle under body, wheel housings or bumpers need to be protected so no damage can be caused to brackets, wiring or the sensors themselves. Sensor replacement, if the unexpected happens, can also be made easier if wiring and cable routing is carefully considered. Fig 1 shows a simple method of routing a concealed cable to, for example; a wheel speed sensor (shown in red). The connector plug (X) is mounted before the cable passes through the bulk head therefore protecting it from contaminants but making replacement of the sensor simpler with less cable to remove and refit.

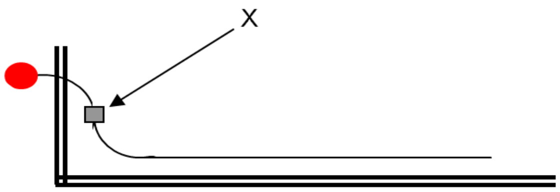

Figure 3.1 Schematic for easy service wiring

3.3.9 Cable routing

Most field operational or naturalistic driving trials require or suggest a level of concealment for the components; this has the advantage of making the driving as natural as possible as the vehicle remains standard looking however a disadvantage lies in the routing and subsequent servicing of wires and components.

With sensor failure being an excepted, but easily mitigated problem, access to equipment installed in the vehicle is of prime importance, this relates to the loggers, sensors or wiring equally.

Wiring should be routed with slack cable at the ends of connection, this relates to the sensor and the logger as these are likely to be frequently accessed. This slack cable acts as a cable full relief, relieving stress on the delicate connectors if the sensor or logger is moved.

On unavoidably long cable runs secondary connectors should be used near the sensor end. This will allow simple maintenance as the majority of the cabling can be left, concealed in the vehicle while the sensor and short cable link is replaced or repaired. This is especially important if the cable routing passes through a bulkhead/firewall or exits the vehicle where servicing complexity is greatly increased.

3.4 Video and imaging equipment

Video data is often vital in understanding or interpreting the vehicle data recorded by the sensors. In almost every way, be it number of cameras, frame rate, resolution, zoom level or storage capacity, camera technology has improved to keep pace with the demands of current day trials. This section introduces video data and other variations of imaging equipment and its use, the specifications for naturalistic and field operational trials and includes examples to demonstrate how camera locations and views can supplement the vehicle sensor data.

The technique of recording and understanding sensor data is well understood and has been used successfully for all types of analysis from industrial based tests through to motorsport. However the need to understand and account for the many different or confounding factors that can occur through a more naturalistic approach will inevitably require video channels. In almost all large scale naturalistic driving studies or field operational trials this video data has been used to evaluate what the vehicle based sensors are recording.

An example of video analysis could be to evaluate distraction or inattention in drivers. This measure is currently complex and expensive to record as it uses high sensitivity eye tracking equipment and/or machine-vision techniques (See 3.5). An alternative technique uses event triggers to highlight areas of interest in the data, such as high longitudinal g, lateral accelerations or yaw, before manual video analysis helps to determine the exact nature of the distraction or inattention at these points of interest. This technique relies on good quality video data in terms of, but not exclusive to, the number of camera angles, video resolution and frame rate.

In addition to the more traditional video channels, imaging technology is being employed both by motor manufacturers and suppliers of data acquisition equipment to build on safety and add detail to the recorded data channels respectively. Technology such as lane departure and lane keeping have been used by motor manufacturers since 2000 and employs machine vision technology to detect the painted lane markings used on major roads. This technology has evolved most recently into after market fitment systems and systems designed specifically for data logging. Another development in video imaging has been eye tracking. Vehicle based studies began approximately fifty years ago but, has until relatively recently, been prohibitively expensive, cumbersome or obtrusive. Recently these systems have been made more automated and have found there way into automotive developments such as Volvo's sleep detection sensor. From a data perspective this piece of equipment can provide valuable information into areas which have been, until relatively recently, difficult if not impossible to record.

3.4.1 Resolution

Camera resolution is often referred to by the number of pixels available on the image sensor; this grid of pixels can also be described in terms of the horizontal and vertical display 'lines'. This system is commonly seen in display types such as televisions, monitors or the size of digital photographs. Generally the greater the number of image sensors (resolution) the sharper the image quality, this however does not guarantee a good image as a number of other factors come into play when dealing with digital video channels.

3.4.2 Frame rate

Frame rate describes the frequency at which the digital video camera samples a unique image, these unique images run consecutively to create a moving image. A low frame rate will be associated with image problems seen as stutter or jumps on the final video, this is best avoided if the video is to be used to suppliment sensor data. DVD quality video signals are becoming common place as camera technology and processing software become more sophisticated, this frame rate represents between 25 and 29.97 frames per second (FPS) depending on whether PAL or NTSC formats are being used.

3.4.3 Encoding

As mentioned above two common types of encoding exist for video channels. These are the Phase Alternating Line (PAL) and National Television System Committee (NTSC) systems. To keep things simple for the purposes of this document the major differences between the two centre around location and frame rates. PAL is used predominantly for Europe, Australasia and some regions of South America and Africa and uses the 25 FPS rate whereas NTSC is used in North America, Japan and some regions of South America using the 29.97 rate. The region where the test is to conducted will normally identify the encoding type to be used.

3.4.4 Positioning

Positioning of the cameras should be considered equally important as getting good picture quality. The diagrams below show some commonly used camera positions and the expected field of view for these. Actual camera locations are not shown as each vehicle interior or exterior location will requiring specialist brackets to be fabricated.

Good practice, installation and positioning advice for loggers and sensors (3.3.1 – 3.3.9) is still relevant when considering video equipment. The failure rates for these devices can be high as they tend to be placed in vulnerable areas to achieve the best field of view; consideration needs to be placed on the installation to avoid these instances.

Sighting the cameras in the vehicle cabin may prove the best view of the driver and other occupants and protect them from an aggressive environment, however, other quality issues could arise in the form of tampering and obscuring. This has been shown to occur in a number of studies, particularly early into the trial period, and unless addressed could lead to missing video channels for the whole trial period. Concealed cameras and/or familiarising the driver with the study aims could prove effective for such occurrences.

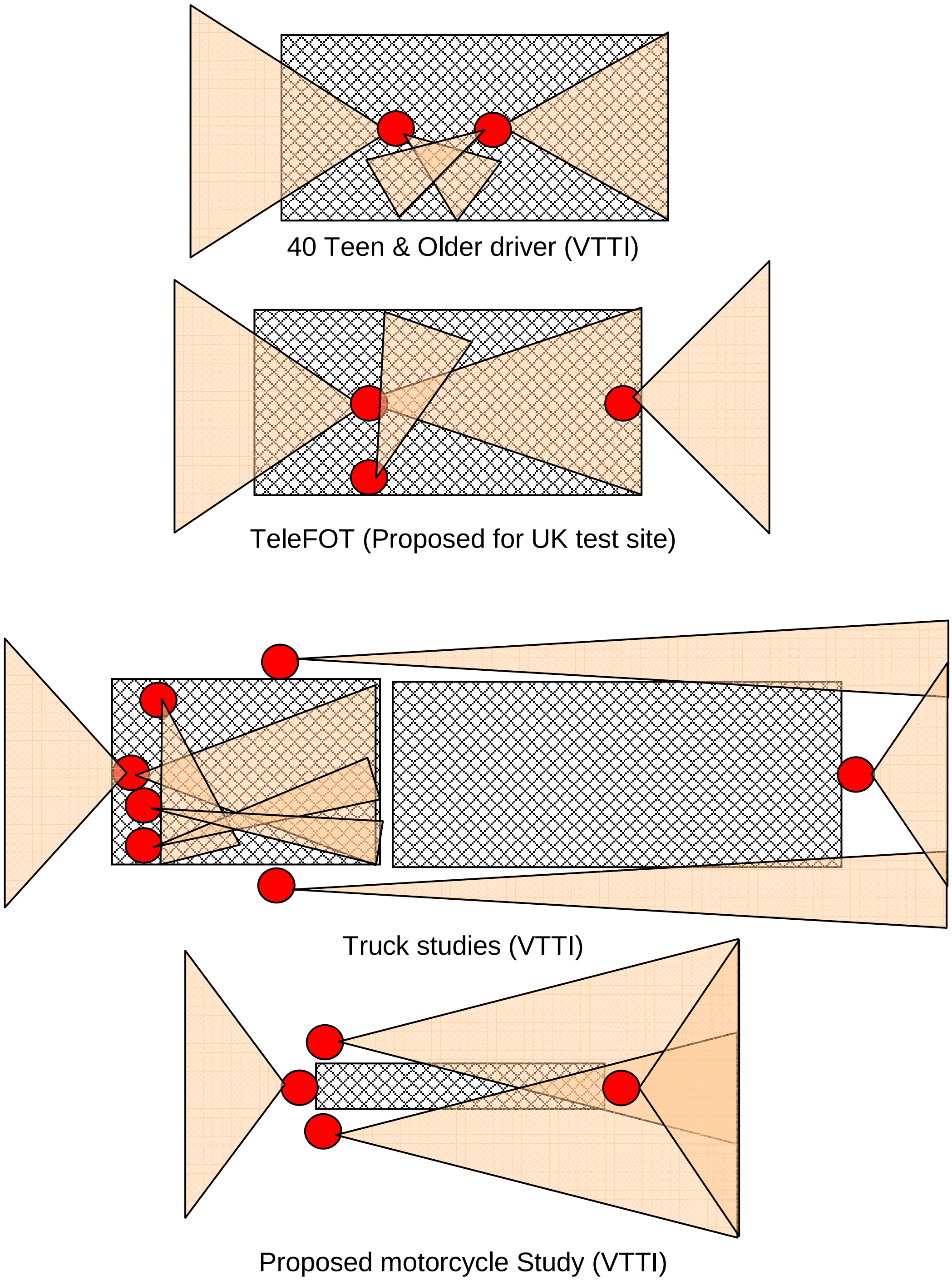

Figure 3.2 Examples of video equipment positioning

In general the greater the video capability the more information will be gathered. Results and data quality, drawn from video footage, from institutions that have completed trials or run a pilot process attest to the value of video data in helping to understand sensor data.

3.5 Eye tracking

Another area which fits in particularly well with the subject of video data is the subject of eye tracking. This field is not solely devoted to the science of eye tracking (it is often simplified to this) as it also includes subject areas such as face recognition, glance analysis and sleep measures amongst others.

One of the simplest methods of analysing behaviour within the vehicle is to carefully monitor and code high quality video footage of the occupants face. This technique, commonly called data reduction, has been used to great effect in previous studies and is particularly relevant if time is more available than funds. Analysis of the footage by trained analysts will allow the coding of a wide range of behaviours. This can include gesture, mood, glance, sleepiness metrics and additional tasks such as eating or mobile phone use. This technique relies on the availability of 'man hours' and good training; the footage will need to be viewed by a trained analyst to extract the best quality data, however the start up costs, beyond that of a good quality camera, are comparatively small.

Beyond the technique detailed above a wide variety of machine vision technologies of varying complexities are available.

Systems that are currently available range, as usual, from the simple to the complex with costs varying respectively. The type of system used will be determined by the research questions and the data quality requirements. For example a simple web cam based device running a relatively simple software package may be all that is required if the research question is equally simple, for example; driver gaze direction on the road ahead?/driver gaze direction not on the road ahead?

More complex (and naturally more expensive) systems use stereo video systems to locate the pupil location in 3D space. From this data all other head measures can be drawn such as head position, head rotations, eye lid measures and blink rates for example. This data is highly accurate (providing careful calibration) but unless exact measures of this type are required, could be above the level of data needed for most trails.

Systems vary in operation and use and will therefore be more suited to certain research questions than others, the major differences occur when considering effects on the driver and evasiveness of the system. Some systems rely on a head or glasses mount (even chin mounts in certain circumstances) to locate the eyes or head and are obviously less suitable for pure naturalistic trials whereas others are non-invasive, relying on remote devices located away from the subject, for example on the dash top or in the instrument binnacle.

Manufacturers of eye tracking equipment, and those that specifically design or optimise for vehicle use, are relatively rare. As such there are only a few suppliers whose equipment is established in full scale trials, one of these is Face lab who provide a range of applications from highly accurate but calibration intensive, through to less accurate almost 'plug and play' devices.

Applied Sciences Laboratories provide head or glasses mount systems which also provide high accuracy measures more suited to laboratory studies due to the small level of physical contact with the systems In both cases the equipment is relatively expensive although it should be noted that high quality data of this type almost certainly will be.

A cheaper option and one which has been demonstrated through the European AIDE project is a web cam based device, the Cockpit Assessment Activity (CAA) developed by VTT technical research centre of Finland and Volvo. This system provides a lower level of data but for a fraction of the Face lab costs, and which could, if deployed correctly with the appropriate research questions, provide invaluable data on distraction.

When considering participant measures such as eye tracking, gaze measures or sleepiness metrics through eye tracking it is also worth mentioning other forms of behavioural data. There are many forms of these, mostly deriving from the fields of sport science or medicine but a few have interesting and useful ties with naturalistic or field operational driving trials.

The two forms described here are not meant as a complete overview of the state of the art in this subject area but as an indication of where human factor data can be introduced into driving trials.

One form of measure that has been used successfully before in VTTI's truck driver drowsiness trials is actigraphy. This is essentially the study of activity and can be used to detect occupant movement over shorter trials or, as has been done before, to log the sleep patterns of long haul drivers. These systems tend to be marginally intrusive - the sensor is worn on a wrist strap and the participant has to continue wearing it - although no more so than a traditional watch.

To explore the subject of driver stress, sleepiness or even the effects of substances, heart rate measurements have been successfully used. These tended to be cumbersome involving the attachment of electrodes to the participant, however recent technical innovations have integrated this technology into a regular t-shirt. This ensures a more consistent data stream as the participant is more likely to wear the equipment allowing much longer data periods to be captured.

3.6 Lane keeping/lane departure

Lane keeping or lane departure measures have been used in previous trials to detect both intentional lane changing manoeuvres, for example their duration and frequency, or for unintentional lane changes, for example, the effects on lane keeping through distraction.

The technology involved in the processing and recording of this measure is well established having been used in production vehicles since 2000. The technique uses machine vision based processing to detect the white painted lines on a road surface, this is monitored in real time and infringements recorded as unintentional (usually a lane change without turn signal use) notified with an alarm, vibrating seat base or in some cases an automated, corrective steering input.

Third party lane detection systems, those available for retro fit, are now becoming available and as such could, depending on access and synchronisation issues, provide a useful source of data. An alternative if more complex route could be the development or use of existing machine vision software. Using this in combination with a adequate forward view camera could provide the data needed for assessing lane keeping and lane change manoeuvres.

As mentioned previously the technology is commonly fitted by the OEM, particularly in higher specification vehicles. If CAN access is sufficient then the data could be drawn from this although limitations and restrictions could apply as discussed in 3.9. Previous trial experience from a number of organisations have shown good data from all three of the sources listed here however replication of this could be very dependant on a number of issues such as vehicle selection, OEM support and software development capabilities to name a few.

3.7 Event identification aids

Huge amounts of data are generated through both naturalistic studies and field operational trials. There is almost no way or, in most cases, a need to analyse all the data. Depending on the study aims and the research questions there will be a requirement to develop event identification aids to help manage the vast amount of sensor and video data. This data can then be categorised in terms of relevance to the project. For example; if the study aim is to identify instances of speeding and events associated with this then a series of identification aids will need to be specifically designed to pinpoint these instances.

This simple example will result in the development of a mathematical ‘filter’ that can be applied to the data. For the example above this could use both the legal speed of the road and the road speed of the vehicle to determine the moments when the vehicle is above (either exactly or by a margin) the legal speed limit.

Event identification aids can be used to great effect when dealing with the large amounts of video data. It is generally very time consuming and unproductive, in terms of identifying points of interest, to look through all the video. The same technique used to identify events as detailed above can be used to reduce the amount of video analysis as it automatically highlights these areas.

Mathematical event identification aids should not be accepted as a constant. Differences in vehicles, drivers and trials will cause the identification filter to become more or less sensitive depending on the logged data. For example; a light vehicle or road car will show a significantly different dynamic trace in the data compared to heavy trucks. Variables such as negative longitudinal g (deceleration) could vary by as much as 0.3g between a braking event with a truck (approx’ -0.3g) and road cars (approx’ -0.6g). This kind of difference could cause a massive increase in ‘events’ which in turn increases data and video analysis time.

3.7.1 Event triggered logging

Using mathematical event identification while conducting trials provides a method to focus the analysis; this technique is often referred to as event triggered logging. This technique uses similar logging equipment but provides a buffering system to only log data for a defined period before and after certain ‘event’ criteria is reached. This technique can dramatically reduce the amount of data recorded by only returning relevant ‘snippets’ as and when certain data values are reached. Analysing data generated from this method will ensure that all the records are within the specified event criteria, there is no need to post-process the data or look through vast quantities of numerical or video channels in order to identify instances of interest. Another advantage of using this method is that long duration trials can be conducted without subject interference and the system needs much less data storage space.

This technique does however have a number of significant disadvantages. Due to the extended periods between data download and therefore less contact with the equipment, a data channel or sensor could drift out of calibration or stop working completely without being picked up. This could have a significant knock-on effect on data quality. There could also be a marginal increase in the risk of missed events. This could occur as more and more automated systems are applied, removing the event classification from a labour intensive but accurate method and relying more heavily on the mathematical event triggers to filter the data – an issue related to this point is that there is no additional data recorded for subsequent analysis. For example; if the project aims change or the mathematical filter needs modifying, there is no way of post processing data which wasn’t collected in the first instance.

Some good practice for using event triggered logging, or any longer duration logging trails, is to provide some form of data quality checks. A system of this type should allow the logging system to be interrogated remotely for routine 'health' checks or to send a failure report when a system or sensor fails. This will provide early notification for servicing or repair and will give an indication of expected data quality after the trial or data download is complete.

3.7.2 Event button/Critical incident button

Another method of determining where instances of interest occur is to install an event or Critical Incident button in the vehicle for self reporting by the driver or passengers. This piece of equipment, whether used for continuous logging or event triggered trials, provides a signal or 'flag' on the data trace to identify where a driver/passenger assessed event has occurred. In event triggered trials this usually means that the logging equipment is triggered and video, data and audio channels are opened for a defined period (roughly a minute or two) around the button press time.

This technique can provide an insight into near misses and other less significant events that event triggered logging would normally miss. For example; an impact with an animal or slight run off road incident could be flagged for later analysis whereas it could be missed on the data trace as it remains well within the set trigger limits. This technique does have some disadvantages in terms of misuse. It is common from previous studies to receive numerous flagged events through the event button for the reporting of, for example; other driver behaviour, reports of logging system failure or events that in hindsight would not normally have been classified an event. This can significantly increase analysis time as these tend to be less severe and therefore need to be removed or recoded for subsequent analyses.

The recorded button press and data that applies to this being operated is largely unnecessary in terms of event notification. This is due to a driver's natural reaction to unexpected events occurring on the roadway. For instance; the driver will normally, if not always knowingly, brake or swerve to an extent which is flagged as an event through the mathematical triggers, only after the event has passed does the driver normally press the button. This overlap in event triggers shows that, if the mathematical triggers are correct, then the data logger will react and record information faster than the driver could, therefore largely negating the use of the incident button for more severe incidents.

3.8 Data enhancement provision

No number of sensors, channels of video data or participant questionnaire can capture and record all data. There are certain areas where data needs to be added into the collection process in order to record, for example; the environment around the vehicle. There exists an array of data enhancement techniques which will ensure that the most complete picture of the trial is recorded, some of these are described in this section.

3.8.1 Environmental factors

There are vehicle based methods of recording weather conditions and environmental factors present during a trial, however these tend to be either expensive, cumbersome or time/labour intensive. For example; the presence of rain can be detected and coded by scanning the video data for instances of wiper usage (discounting instances of wash/wipe of course) or by monitoring the wiper switch electronics. This method is not always practicable, a much neater (if slightly less accurate) method of obtaining this information could be through weather and environment reports.

Weather forecasts are, like most other predictions, susceptible to change. This, depending on the trial aims, may or may not provide an acceptable level of accuracy; it will however provide a much simpler method of determining the prevailing weather conditions for the specific period of time and/or location. Another factor to consider is the resolution for weather forecasts, this is in general quite low, for instance there may only be an update every hour or so. Consideration should be given to this and the trial aims to ensure that adequate data is collected.

Another record of the driving environment would be instances of daylight and darkness. This could be done through an additional sensor, for example; a photodiode light detector, however as mentioned previously this will only add to the already complex on board data acquisition equipment.

The hours of darkness and daylight for each day and for each location on the globe are accurately recorded, this allows for post data collection calculation and data enhancement. Tables are available showing the exact time where daylight (sunrise) occurs and darkness (sunset) begins.

This method also avoids ‘false positives’ which a photodiode will be susceptible too. Instances of the vehicle entering a tunnel or an underground car park could result in a false darkness/night time reading; likewise, parking the vehicle in a brightly lit area at night could give the opposite reading.

3.8.2 Road type (map matching)

Another area of data enhancement which has been used to record the roadway environment is map matching. This technique allows the coding of roadway based data such as road type/classification, nodes/junctions, area description and speed limit in addition to the existing GPS based data stream. No sensor or video coding will achieve the results that map matching can provide. Map matching also allows for finer analysis of the data as road types and speed limit bands can be selected to determine, for example; over speed in urban 50kph zones or the percentage of a journey spent in urban zones.

Infrastructure data can, in some circumstances, be used to enhance the vehicle based data stream. This data is recorded routinely on selected roads and could include information on Traffic flow or throughput for each section or lane, pollution monitoring and also visibility sensing. This data is usually recorded by a third party so access and synchronisation issues will arise. However, good integration of this data, depending on trial design, could be vital in understanding the vehicle data.

3.9 CAN Data

Another large area of data acquisition is accessing the internal vehicle network. This is commonly referred to as the ‘controller area network’ data or CAN data although a number of other protocols exist. These include ‘Local Interconnect Network’ (LIN) used to control low speed networks for comfort and convenience components such as air conditioning or electric seat motors, and the more recent ‘Flexray’ developed to use much higher data rates for components such as adaptive dampers and advanced vehicle control units.

CAN, LIN and Flexray can be summarised as message protocols used by manufactures for the communication of electronic microcontroller systems within the vehicle.

This section covers the accessing of CAN data as this is both the most common and relevant for current FOT and Naturalistic activities. CAN development started in the early 1980s and was available to automotive manufactures by 1987. The CAN protocol was evolved over the years and is now mandatory for all vehicles sold in the European Union.

3.9.1 Vehicles Types

Not all vehicle controller area network (CAN) are created equal, there is little commonality between different automotive manufacturers or even between makes and models from the same manufacturer. These differences between CAN protocols can make vehicle selection problematic. Historically, trials, due to their large initial investments and complex technical nature, have tended to use manufacturer support in the form of sponsorship or project partners to help simplify and speed up the access to CAN data. This technique will always provide the best access to the data as the manufacturer will have a vested interest in the resulting data and analysis. However it is becoming increasingly common, due to the cost reduction in technical equipment, not to have major manufacturer support. This presents a significant technical challenge to organisations and institutions. There are however some techniques and good practice guidelines to help speed up this area of data analysis.

At present a number of data logger manufacturers will provide CAN access hardware and software. This is often good at getting hold of the more basic parameters and, unless the proposed trial needs additional channels, could provide adequate data. The types of channels that can be accessed through this technique are not easy to specify here as they can vary dramatically between logger manufacture and vehicle make or model, although it is safe to say that the total numbers will be significantly reduced from a more technical approach.

Using anecdotal information from previous trials some good practice on vehicle selection can be drawn.

<u>Test</u>

Testing on as many different vehicles, different models and different variants will provide an idea of the level of data availability between different vehicles. This can be used as an initial filter for vehicle selection if they are to be bought or will help develop the sampling plan for participants.

<u>Age and condition</u>

The age of the vehicle, and therefore its year of manufacture, will have a significant effect on the availability of data from the CAN. Early vehicles have in general much fewer CAN variables than modern cars, due mainly to increased electronic complexity. Depending on the study design this may have a bearing on vehicle selection as specific channels may be needed and only a more modern vehicle can provide this – however, earlier vehicles tend to be easier to access or interpret.

<u>Skew to participant</u>

In a number of studies the CAN data has become secondary to the securing of participants, this becomes more of a concern if the study is based on relatively restricted driver types. A technical solution around the problem of missing CAN data may be easier than recruiting another participant. Sensors are readily available to log parameters not recorded by CAN, Equipment such as positions sensors (pedal, steering wheel), engine sensors (RPM, temperature, pressure) and vehicle dynamic data (road speed, suspension movement) amongst many others are easily obtained.

Other factors to consider before accessing the CAN data focus on the responsibility of the organisation and liability. One area where the CAN data is commonly used is to control vehicle safety systems. These systems may include traction control, ABS or stability control, all of which could cause critical events should system failure occur.

Measures need to be taken to ensure that accessing and reading the CAN data will not interfere with the electronic vehicle systems. In some cases where after market or third party devices are used to access the CAN network the vehicle warranty may become invalid, this is especially important if the vehicle is owned by the participant.

3.9.2 CAN data quality

A great deal of effort is placed in acquiring good quality data, whether from GPS, accelerometers or auxiliary sensors throughout the vehicle, this should also be true for CAN data. Determining the quality of CAN data is particularly difficult as the proprietary filtering and signal conversions applied by the original equipment manufacturer (OEM) will all have an effect on data quality. Just understanding the purpose of the CAN signal may give an indication of the expected accuracy. The vehicle and electronic control modules do not need large amounts of highly accurate or frequently sampled data to operate, a step change or yes/no response may be enough for component interaction. It can therefore be seen that, from a research viewpoint and dependent on study, this data could contain insufficient quality.

This is not to say that CAN data is not worth collecting, just that the extra effort that could be required to access this data may outweigh the value of the data gathered.

3.10 Interactions and synchronisations

Recording of the data as described in section 3.1 through to 3.9 is only part of the story. A large part of the work in terms of the usability of the data and therefore, in some ways, the success of the study is in data synchronisation.

Fundamentally ‘data synchronisation’ describes how the equipment installed in the vehicle, whether data loggers, video equipment or eye tracking, interact with each other to provide a complete data set. This is sometimes simplified and referred to as ‘time stamping’ – the process of synchronising the data with respect to a common time.

Common time within the data logging spectrum varies considerably. The vehicle for instance will have a common time from CAN or the CAN logger. There may also be, depending on equipment used, different times for eye tracking data streams, other computer hardware based within the car such as a central PC or from GPS based loggers, derived from the GPS time signal.

3.10.1 GPS time

Most GPS based loggers use GPS time to synchronise the data. Typically for after market systems the data recorded by these devices contains variables such as vehicle speed, location, altitude, some degree of acceleration and system data all synchronised to a common GPS time. This time is generated by satellites as part of their positioning message and as such can be considered a constant, although in reality a small degree of drift will be present. After market data logging systems based on GPS will use this time and may also synchronise video logging streams to the sensor data stream based on this. Problems occur when equipment from other suppliers needs integrating; this is where using an alternative synchronising time source is needed.

3.10.2 PC time

An alternative method of synchronising data that will allow greater flexibility when choosing equipment is to refer the data to an alternative time source, this is typically an on board PC. This technique has been used successfully in a number of projects as it allows a wide variety of equipment from different manufacturers to be merged into one data stream with a common time reference.

By using this system data can be streamed constantly from the different systems and recorded on an onboard computer, the computer uses a time standard (Usually the set computer time) and references all the data to this. Difficulties may arise where systems have different start up times, for instance a GPS receiver. This piece of equipment may take between a few seconds and a few minutes to detect satellites and begin logging which could create latency in the data stream. This should be tested through a pilot study to understand or correct for the effect.

The advantages of using the PC based system lie in the relative ease with which it can be expanded or modified. Compared to an 'off the shelf' logging system, which is usually designed to work as a unit using proprietary software, the modifications and upgrades can be completed with relative ease. Another advantage that has become more relevant recently is that computer hardware and software is cheap. This is of course 'cheap' when compared to designing and developing a GPS based logger system from a blank sheet, but items such as expanding the memory or the increasing the processing power are now well within both the technical and monetary limits of most institutions undertaking this kind of work.

Much like CAN data it is particularly difficult to be prescriptive when considering a final data set or recommended equipment. A great number of variables exist which cannot be fully covered in this document, for example; the PC and associated software/operating system to be used, the different types of logger, the extent of the instrumentation in terms of the requirement for eye tracking/lane departure.

3.11 List of equipment specific to metrics in D2.2

This document goes some way to determining a complete equipment list to meet the set of variables determined in D2.2. The information included has been derived from a vast range of organisations and previous experience from running trials. Where relevant, experience and good practice has also been drawn from other data intensive sectors such as the motorsport industry or industrial vehicle testing.

It is not possible to determine every problem that could be encountered or provide a solution to this. This is due to the varying nature of trial type, vehicle selection, trial environment, available budget and time, and institution expertise amongst others, all of which could significantly alter the approach taken.

Good practice, advice and experience are included where no specified approach can be given. This unfortunately does not make the report a complete guide to field operation trials or naturalistic driving. It should however help guide the decision making process through these areas and provide an alternative approach or compromise.

A summary of the report with reference to D2.2 metrics is shown below. Those identified as essential are addressed in **bold** with associated notes, however more information can be found in the body of the report. Those not in bold are included for completeness and although some are addressed in the report no further details are included. Indeed for some of the variables there may be, to date, no practicable way of measuring these in a naturalistic or field operational test.

.

	Variable	Equipment needs	Ref
	Driver traits/states		
1	**Age**	**Questionnaire - Unchanging.** Recorded pre-trial through Initial questionnaire or screening process	
2	**Gender**	**Questionnaire - Unchanging.** Recorded pre-trial through Initial questionnaire or screening process	
3	**Country of living**	**Questionnaire - Unchanging.** Recorded pre-trial through Initial questionnaire or screening process	
4	**Education**	**Questionnaire - Unchanging..** Recorded pre-trial through Initial questionnaire or screening process	
5	Income		
6	**Professional driver**	**Questionnaire - Unchanging.** Recorded pre-trial through Initial questionnaire or screening process	
7	**Driving experience (years, total kilometres, kilometres per year)**	**Questionnaire - Pre defined intervals depending on trial design and aims.** Recorded pre-trial through Initial questionnaire or screening process	
8	**Sensation seeking scale**	**Questionnaire - Unchanging.** Recorded pre-trial through Initial questionnaire or screening process	
9	**Physical condition**	**Questionnaire - Unchanging.** Recorded pre-trial through Initial questionnaire or screening process	
10	Aggressiveness		
11	Cognitive skills		
12	Risk perception		
13	Locus of control		
14	Self-reported driving behaviour (DBQ)		
15	Attitudes/intentions towards speeding, safety, environment		
	Driving behaviour metrics		
16	**Frequency of performed left and right lane changes (number per kilometre and hour)**	**Logged - Lane detection equipment.** Third party radar sensors or machine vision technology. Some OEM fitment with appropriate CAN access. *Also need accumu-*	**3.6**

		lated driving distance (73) or accumulated driving time (72) for completion of metric	
17	Frequency of active overtaking (number per kilometre and hour)		
18	Frequency of passive overtaking (number per kilometre and hour)		
19	Deviation from desired lane		
20	Frequency of route changes (number per kilometre and hour)		
21	Travel time uncertainty		
22	Delay		
23	Following/free state profile		
24	**Speed profile**	**Logged.** Normally from GPS based logger system – alternatives include wheel speed sensor, optical road speed sensor or CAN from instruments. *Also need time(72,77) for speed profile*	**3.3**
25	Intended speed		
26	Desired lane		
27	Intended route		
28	Actual route		
29	Use of car horn		
30	System interaction and driving behaviour related responses to alarm/warning		
31	Reaction time to alarm/warning		
	Driver distraction and state metrics		
32	**Mental workload**	**Logged – eye tracker.** Normally from eye tracking equipment - alternatives include manual data reduction techniques from video	**3.5**
33	Stress		
34	**Fatigue/drowsiness**	**Logged – eye tracker.** Normally from eye tracking equipment (automated) - alternatives include manual data reduction techniques from video	**3.5**
35	**Distraction from primary driving task (eye-tracking, glance duration, fixation)**	**Logged – eye tracker.** Normally from eye tracking equipment (automated) - alternatives include manual data reduction techniques from video	**3.5**
36	Head-tracking		
37	Number and position of hands on steering wheel		
38	**Presence and use of in-car devices (e.g.**	**Video.** Data derived from	**3.4**

	mobile phone, navigation system etc)	manual data reduction techniques.	
39	Driver identified events		
40	**Presence of passengers**	**Video.** Data derived from manual data reduction techniques.	**3.4**
41	(moving) object inside vehicle		
42	**Performing tasks other than the primary driving task:** - **eating/drinking** - **adjusting radio or other in-car device (e.g. climate controls, CD etc)** - **dialling or texting on mobile phone**	**Video.** Data derived from manual data reduction techniques.	**3.4**
	Vehicle condition metrics		
43	**Vehicle type (manufacturer, model, vehicle age)**	**Questionnaire - Unchanging.** Recorded pre-trial through Initial questionnaire or screening process	
44	**Vehicle mass: driver + passengers, load besides driver and passengers, trailer connected or not, amount of fuel in the tank**	**Questionnaires, Video, Logged.** Basic vehicle specification (questionnaires) with Fuel load (logged, CAN) and other load information (video)	
45	Air conditioning: use/not use		
46	Wiper status: use/not use		
47	Other auxiliaries: use/not use		
48	Cooling fan: operating/ not operating		
49	**Type of transmission**	**Questionnaire - Unchanging..** Recorded pre-trial through Initial questionnaire or screening process	
50	**Type and amount of in-vehicle systems**	**Questionnaire - Unchanging.** Recorded pre-trial through Initial questionnaire or screening process	
	Vehicle parameters		
51	**Speed**	**Logged.** Normally from GPS based logger system – alternatives include wheel speed sensor, optical road speed sensor or CAN from instruments.	**3.3**
52	**Acceleration (longitudinal, lateral & gyro)**	**Logged.** Normally from vehicle based accelerometers, alternatively longitudinal through speed trace if sam-	**3.3**

		pling is high enough	
53	**Deceleration (incl. sudden braking)**	**Logged.** Normally from vehicle based accelerometers, alternatively longitudinal through speed trace if sampling is high enough. *Optional brake status to reduce non-brake events (56,57)*	**3.3**
54	Percentage throttle		
55	Percentage clutch		
56	Percentage brake		
57	**Brake force**	**Logged.** Normally from brake force sensor (brake fluid pressure), alternatively pressure sensor on pedal.	**3.3**
58	Gear position		
59	**Steering wheel angle**	**Logged.** Commonly CAN alternatively string or rotational potentiometers	**3.3**
60	Turn signal		
61	**Lateral position**	**Logged – Lane detection equipment.** Third party radar sensors or machine vision technology. Some OEM fitment with appropriate CAN access.	**3.6**
62	**Lane departure**	**Logged – Lane detection equipment.** Third party radar sensors or machine vision technology. Some OEM fitment with appropriate CAN access.	**3.6**
63	**Time to line crossing (TLC)**	**Logged – Lane detection equipment.** Third party radar sensors or machine vision technology. Some OEM fitment with appropriate CAN access.	
64	**Distance to vehicle in front**	**Logged – headway sensor.** Third party radar sensors or machine vision technology. Some OEM fitment with appropriate CAN access.	**3.3**
65	**Distance to vehicle behind**	**Logged – headway sensor.** Third party radar sensors or machine vision technology. Some OEM fitment with appropriate CAN access.	**3.3**
66	Distance to other surrounding vehicles		
67	Side vehicle detection		
68	**Time headway (forward & rear headway detection)**	**Logged – headway sensor.** Third party radar sensors or	**3.3**

		machine vision technology. Some OEM fitment with appropriate CAN access.	
69	Space headway (forward & rear headway detection)		
70	**Time to collision (forward & rear TTC)**	**Logged – headway sensor.** Third party radar sensors or machine vision technology. Some OEM fitment with appropriate CAN access.	**3.3**
71	**Post encroachment time (PET)**	**Logged – headway sensor.** Third party radar sensors or machine vision technology. Some OEM fitment with appropriate CAN access.	**3.3**
72	**Travel time (including stop time)**	**Logged.** Common trial time, usually from key on to key off. *Possible post processing to determine key off stop length for instance*	**3.3**
73	Travel distance (mileage)		
74	Waiting time at intersections		
75	Friction		
	Environmental metrics		
76	**Precipitation (i.e. snow, rain, fog)**	**Data enhancement – Weather conditions** Records of weather conditions on trial date. Alternatively video data reduction or wiper/light use through CAN data.	**3.8**
77	**Time: date; time of the day**	**Logged.** Common trial time. Based on PC, Car or, more commonly, GPS time	**3.3**
78	**Daylight/dark conditions**	**Data enhancement – light conditions** Records of light conditions on trial date. Alternatively video data reduction, light sensor or light use through CAN data.	**3.8**
79	Air pressure (measured with vehicle sensor)		
80	Air temperature (measured with vehicle sensor)		
81	Humidity (measured with vehicle sensor)		
82	Wind speed		
	Road condition metrics		
83	Road surface conditions (in Sweden by use of the winter model)		
84	Road distance (a GPS will probably not have accuracy enough)		
85	**Road and traffic conditions based on GPS**	**Video.** Data derived from	**3.4**

	and time:	manual data reduction techniques. Alternatively location and time matched traffic data *(if available for trial location)*	
86	**Gradient; horizontal curve; junction; roughness;**	**Data enhancement - Map Matching.** Junction information available through map matched data. Gradient possible through GPS data	**3.8**
87	Macro texture		
88	**Road type**	**Data enhancement - Map Matching.** Road type information available through map matched data.	**3.8**
89	**Environment (Urban/interurban/rural)**	**Data enhancement - Map Matching.** Road Environment information available through map matched data.	**3.8**
90	**Number of lanes**	**Data enhancement - Map Matching.** Lane information available through map matched data.	**3.8**
91	**Width of lanes**	**Data enhancement - Map Matching.** Lane information available through map matched data.	**3.8**
92	Base capacity and saturation flows		
93	Central barrier		
94	Sight distance		
95	**Speed limit**	**Data enhancement - Map Matching.** Speed limit information available through map matched data.	**3.8**
96	Location of speed cameras		
97	**Current traffic management: road markings, signs, rumble stripes, etc**	**Data enhancement - Map Matching.** Traffic management information possible through map matched data. *May require data enhancement*	**3.8**
98	Bus stops or parked cars along the street		
99	Hard shoulder		
100	**Intersections:** - **frequencies** - **intersections types (signals/roundabouts/yield/stop)** - **exit roads**	**Data enhancement - Map Matching.** Intersection (node) information available through map matched data.	**3.8**
101	Number of stops on route		
	Traffic condition metrics		

102	**Traffic density**	**Video.** Data derived from manual data reduction techniques. Alternatively location and time matched traffic data *(if available for trial location.* Congestion measures calculable depending on speed limit info (95) and vehicle speed (51)	**3.4**
103	**Speed distribution, average speed and standard deviation**	**Post processing.** Calculable through vehicle speed (51) and time (72) data.	
104	**Traffic composition**	**Video.** Data derived from manual data reduction techniques. Alternatively location and time matched traffic data *(If available for trial location.*	**3.4**
105	Traffic signal picture		
106	Traffic flow		
107	Other (unrelated) incidents that may affect traffic flow		
108	**Category of road users in vicinity (pedestrian, cyclist, light/heavy vehicle, etc)**	**Video.** Data derived from manual data reduction techniques. Alternatively location and time matched traffic data *(If available for trial location.*	**3.4**
109	**Speed/acceleration of road users in vicinity**	**Video.** Data derived from manual data reduction techniques. Alternatively location and time matched traffic data *(If available for trial location.*	**3.4**
110	Behaviour of road users in vicinity		

3.12 SHRP2 DAS

With the advent of the SHRP 2 project, the natural successor to the highly successful 100 car study, a development of the data acquisition system (DAS) architecture was sought. The 100 car study used high sensitivity and accuracy sensors coupled to a powerful DAS system to record hundreds of hours of video and alpha numeric data leading to a successful project, however, the system as detailed previously was by now outdated and cumbersome for the purposes of the more ambitious SHRP 2 project.

VTTI developed a much more streamlined system to capture all the required data while ensuring the participants vehicle remained as standard as possible. This was achieved by integrating a number of the sensor units and developing where needed unobtrusive wire free communications.

The schematic of the system detailed below is shown in Figure 3.3.

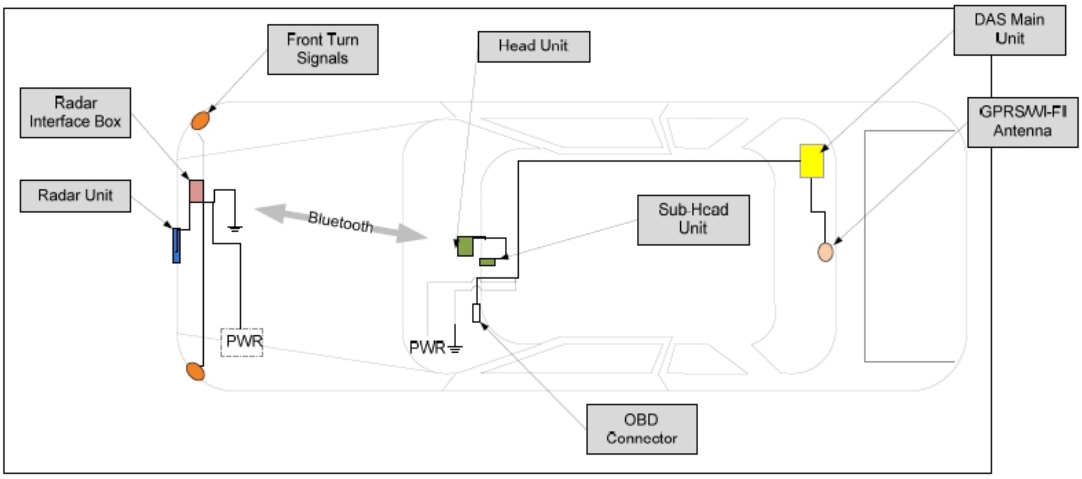

Figure 3.3 Schematic of planned SHRP2 DAS system

The major part of the DAS remains the main unit. This, like the 100 car study, remains in the luggage compartment out of sight and away from damage. From here there is direct communication with all sensor and video units. Vehicle positioning data is provided to the DAS via a GPRS/WI-FI antenna located towards the rear of the vehicle.

From here sensor data is provided from 3 sources. Initially this is physically wired to both the head unit and an OBD connector and additionally from a wireless transfer to radar and auxiliary data.

The head unit is designed in house by VTTI and is installed behind the rear view mirror on the vehicle wind screen. This unit contains an array of camera views and built in accelerometers used to detect occupant/environment data and vehicle dynamic behaviour respectively. The unit itself is relatively small being approximately 100mm by 100mm, this could ensure that subjects are not influenced by obviously positioned data acquisition equipment,

The OBD connector, wired to the main DAS unit, is connected to the vehicle communication network and as such reads information on both engine and chassis parameters

The final part of the system is not physically connected to the main DAS unit but sends data via Bluetooth. This method avoids the need to breach the vehicles fire wall separating the engine compartment from the passenger compartment. The data on this system is based around a forward facing radar and the vehicle turn signals

4 Data Storage and Management Methods

This section details the available knowledge and experience relating to data storage and management methods for ND studies and FOTs. The information has been divided into the following subsections;

- In-vehicle data storage and retrieval
- Data storage – uploading and back up
- Data storage – database creation
- Data base Requirements – performance, reliability and access
- Data Quality

4.1 In-vehicle data storage and retrieval

There is a need for in-vehicle data storage in all studies which involve naturalistic driving. The amount of storage necessary and technologies chosen will vary according to the data acquisition methodology.

The FESTA Handbook (FESTA, 2008) suggests that compression algorithms could be used to maximise storage capacity by reducing the size of data. It suggests that lossless compression, for example ZIP, should be used for CAN and/or sensor data but loss compression can be used for voice and video, for example MP3 or MPEG-4. Although the Handbook also warns due to the additional complexity of including compression in a system, the likelihood of error and/or malfunctioning increases.

It is also necessary to transfer the data stored in the vehicles to a central database for analysis in a way that minimises data loss. Previous studies have used a variety of technologies to both store and retrieve data from vehicles. The data storage devices were usually stored in the vehicle boot (cars) where they were unobtrusive to the driver but allowed access where necessary for data retrieval.

One technology is removable storage devices. The Test Site Sweden FOT (SAFER, 2008) used USB hard drives that were replaced periodically and taken to the central database for upload. Regan et al (2006) report that Flash cards were used to store data within the vehicles used for an Australian FOT. Participants were sent a 64Mb new card every month. They replaced their old cards with the new Flash Cards themselves and then sent the old cards back to the researchers.

Other studies use technology that allows the researcher to upload data directly from the vehicles without involvement of the participants. Lee et al (2004) state that VTTI's lane-change study used an in vehicle data storage system. This consisted of;

- two high-8 video cassette recording devices
- a central processing unit which combined data from sensors & radar
- a titler which added a time stamp to each frame of video
- zip drive
- harness bundle
- keyboard
- quad splitters

The unit was stored unobtrusively in the luggage compartment and allowed the researcher to gain access to change tapes and download computer data files.

UMTRI's Road departure crash warning system FOT used external technology which was connected to the vehicles to upload data when the participants returned to UMTRI following their period of naturalistic driving (LeBlanc et al, 2006). Technicians used a 'cart' equipped with a 13.8 volt DC power supply, network switch, mode control switch, keyboard, mouse and LCD monitor to maintain and download from the in vehicle devices before uploading to the central server. The 100-car study (Dingus et al, 2006) also used a physical connection to retrieve data from the involved vehicles. However this study used additional technology which allowed the researchers to locate and travel to the cars involved in the study. 'Chase' cars located study vehicles by using a mobile phone and laptop configuration which pinpointed both the chase car and the study vehicle locations on a map. Once the chase car had travelled to the study vehicle location, a data transfer cable was attached to an outlet near the license plate of the study vehicle. This allowed data to be downloaded.

Another method used in some more recent studies is to use wireless technology to retrieve data. Data from the cars in McGehee et al's (2007) study were automatically downloaded via a secure wireless connection when the teenage participants parked in their school car park. Encrypted data was then sent to the research laboratory. Wireless technology was also used by the Belgium eco-driving study reported in Beusen et al (2009). CAN data was stored on the data logger using an internal memory card and was transmitted daily to the central server via an on-board GPRS-modem.

4.2 Data storage – uploading and back-up

The FESTA handbook (FESTA, 2008) states that data retrieval and upload procedures are essential to reduce the likelihood of data loss. The handbook states that the verification of data completeness and the backup of data should be part of this process. These processes should occur before the deletion of data stored within the vehicles. The actual procedures and technologies used to achieve this are reported less often in the literature. Those that do report this also report a certain amount of data reduction or organisation occurring as part of the upload/transfer to a central database.

The procedures for uploading data in the UMTRI FOT are described in LeBlanc et al (2006). The Data Acquisition System included a computer that maintained a database of data collected. This included a table that catalogued the names and sizes of data files which assessed in the tracking of data. Once data was downloaded from the vehicles, a program copied all files to the project file server and then inserted the numeric data into a Microsoft SQL Server database. A number of secondary processing tasks were run on the database server during low-usage periods. All files were backed up and moved off site for safe storage. In addition to this raw data from the vehicle and subjective data was sent to an independent evaluator.

The Test Site Sweden FOT (SAFER, 2008) also used specific software to assist with the upload process. This software was developed for the project with the aim of matching the formats of the raw (in-vehicle) data to the database structures. During upload, video was separated and signal data was interpolated to a base frequency of 10Hz. Driver ID and subject data provided by the driver in a 'driver form' was attached manually. In contrast to the studies described above, all the raw data from the 100-car study (Dingus et al, 2006) was copied and stored on DVDs following data integrity verification. Two sets of DVDs were created and stored at 2 different locations. 1 copy was kept in Northern Virginia where data downloading took place and 1 was sent to VTTI. 'Triggering software' was then run on each DVD that was sent to VTTI in order to capture event data. This selected data was then copied to the Networked Attached Storage (NAS) server. Once the event data had been copied to the NAS a further copy was made and then custom software was used to delete the data from the instrumented vehicles. In this way the 100-car study was able to maintain 2 copies of all data in order to minimise data loss.

4.3 Data storage – database creation

The precise organisation of data in a database will relate to the research questions and methods. However a database of some sort is essential for data analysis therefore each study will have to have selected appropriate software and hardware to meet their data storage needs – especially as naturalistic driving studies produce vast quantities of data. Some study reports give an indication of this although as many use custom software, this information is not detailed.

The FESTA handbook suggests that all data except video should be stored in a relational database, supporting ANSI SQL. It identifies two ways of storing video data. The first is on a file server. The second is within a database with each frame being stored as a JPEG image or BLOB (Binary Large Object). Alternately the complete video could be stored as a binary file.

The data generated by the UMTRI FOT study was stored in a data archive. This contained all types of data including raw data from the vehicles (numerical, video and audio); supplementary data (biographical, highway performance; system data including secondary processed data and corrections as well as subjective data collected from the drivers before and after their period of naturalistic driving. Excluding the secondary processed data, the total size of data stored in the archive was approximately 350GB. The database (Microsoft SQL) containing the numerical data was just over 204 GB and contained 89 tables with 54 billion data elements. The video data was reduced to 135GB by a combination of sub-sampling images and efficient compression (90%). The database produced by the Test Site Sweden FOT (SAFER, 2008) was based on the UMTRI structure and design with a few modifications according to the differences in the studies. As previously stated the Test Site Sweden study separated video from the other data from the study vehicles. Video data (binaries) were stored on the same hardware as the database tables but they constituted virtually two separate systems.

The 100 car study used similar methods. The study collected over 6 terabytes of data and these were stored on over 1,300 DVDs. However in order to analyse the data, several MySQL databases were created on the NAS server. Analyses were based on 3 types of data; driving performance derived from the raw data collected in the vehicles; reduced data from event analysis; and subjective pre and post data collection questionnaires. These were copied, created, or edited into the MySQL databases and linked using identification codes

The Test Site Sweden FOT also used the SQL language when they developed a database search interface specifically for the project. SQL was chosen as it '*enables rapid access to the data stored and provides features for more advanced scripts, e.g. for evaluating trigger levels of vehicle dynamics*' [p7].

4.4 Data base Requirements – performance, reliability and access.

The following is summarised from the TeleFOT D2.3.1 Data Specification and Quality

4.4.1 Performance Requirements

The fundamental concepts of performance are response time (how quickly a system responds to a request) and throughput (how much work can be done in a specified amount of time). Performance requirements guarantee that the database end users can operate as efficiently as possible and that incoming data is handled at a feasible rate. Limitations on performance can be for example peak hours of use and the number of analysts as well as the amount of incoming data.

The following should be considered in relation to performance:

Table 4.1 Database Performance Requirements

Requirement	Comment
Business level performance requirements	The Data Management Centre must be able to process more data in a day than can be generated in a day. Data Management Centre must also be able to provide the "raw" input data it downloaded from data loggers or local data centres
Data logger related performance requirements	The data logger must be able to record data using sufficient sampling frequency. For GPS this stands for a sampling frequency of 1 Hz or higher and for acceleration data a sampling frequency of 50 Hz. The Data Management Centre must be able to process the amount of data recorded by data loggers.
Monitoring performance requirements	Data Management Centre must have a function for monitoring the data processing. In particular, the amount of records processed per unit of time.
Database performance requirements	In order to be useful when conducting analyses, the database has to be able to provide required data within a reasonable amount of time. The maximum time that an advanced and optimized database query can take up to is determined to be one hour.
Multiple transfers requirements	Data Management Centre must be able to open multiple concurrent connections to receive and process all the data from the amount of around 3000 data loggers within the project. Data Management Centre must be able to handle the estimated amount of traffic generated by the data loggers.
Maximum throughput time requirements	The requirement for maximum throughput time depends on how quickly the data is needed for analysis. A reasonable time for the data to be ready for analysis purposes has been determined to be 24 hours after it has been collected. The same reasoning applies to throughput time for data processing and fusion purposes as well.
Maximum load and response time requirements	The number of different users and user interfaces will be rather limited which brings down the maximum load and response time requirements. However, the data collection, processing, storing into a database and database queries for analysis purposes will all happen concurrently at times. As a result, the data collecting, processing as well as the database have to be developed to be efficient as the amount of data will be large.
Minimal functional coverage	Data Management Centre must be able to create all the data entries (and structures) reflecting the current business requests. The same is valid about the deletion requests, to be accompanied by the reference integrity constrains (no orphans, cascade etc.).

	The data migration routines if any must be provided.
Data accessibility	Data Management Centre should make accessible the business the contents: the data must be readable and updatable, according to the rights of the user. Being a B2B service, real time access should be supported.

4.4.2 Availability Requirements

The availability requirements aim to provide the conditions that will guarantee that the access to the database is as wide and constant as possible. The objective is to try to anticipate possible downtime and service interruptions, in order to predefine recovery strategies.

The most common sources of availability limitation can be, for instance, electric power down, absence of connectivity, denial of service by hackers. Planned and unplanned downtime sources should be evaluated and ranked, in order to prepare mitigation strategies.

The following should be considered in relation to Availability:

Table 4.2 Database Availability Requirements

Requirement	Comment
Redundancy	Data Management Centre must provide solutions to guarantee locally and/or remotely the back-up version of the DB requiring also some added storage capacity.
Back-up frequency	Based on the amount of data and of users, the periodic back-up must be planned. The synchronization with the remote devices should be considered as well.
Mean Time of repair	Based on the determination of the criticity, the time of repair should be calculated and communicated to stakeholders.
Mitigation strategies	In order to guarantee and make the access to the DB contents faster, possible strategies can be implemented as precalculated queries.
Recovery time objective	If the local DB will prevent data loss, implementing redundancy and mitigation strategies, the recovery time in case of service interruptions would not be critical. The evaluation of the maximum time has to be defined according to the business scenarios. For instance the maximum time would be in the order of seconds, to meet eCall scenario requirements. A downtime of one week could be tolerable as insurance companies requirement.

4.4.3 Security Requirements

This section provides security requirements for the database environment which need to be fulfilled in order to guarantee the common goals of protecting confidentiality, integrity and availability of information.

Common security risks are incorrect handling or storing of usernames and passwords, malicious programs and intrusion attempts.

The following should be considered in relation to security:

Table 4.3 Database Security Requirements

Requirement	Comment
Identification	Every entity, device or user, using database environment must be identified before granting access to database environment. Normally username is used to identify the entity.
Authentication	Every entity using database environment must be able to prove its identity for database environment. The entity must have a credential issued by the database administrator for proving the identity. Password is the most common credential for accessing some system.
Authorisation	Entities can only access data and resources for which they have been properly authorized. Access rights to database resources and data must be defined carefully and by using principle of least privilege. The entity should have access only to the resources and data that it is necessary and can be granted without breaking privacy policy of the database environment.
Immunity	Database environment must protect itself from infection by unauthorized undesirable programs, e.g. viruses and Trojan horses.
Integrity	Database environment must ensure that any data cannot be modified without authorization and all changes to data have to be legal, accountable and correct. The typical objectives of an integrity requirement are to ensure that communications and data can be trusted.
Intrusion detection	Database environment must detect and record attempted access or modification by unauthorized entities. Intrusion detection system should inform security personnel on intrusion attempts so that they can handle them.
Non-repudiation	Database environment must prevent an entity to one of its interactions (e.g., message, transaction) from denying having participated in all or part of the interaction.
Privacy	Database environment must secure the confidentiality of the sensitive information, i.e. ensure that unauthorized individuals and programs do not gain access to sensitive data and communications. Privacy is both political and technical issue. Policy defines what data is sensitive and technologies and processes implement the privacy.
Security auditing	Database environment must enable security personnel to audit security mechanisms and their status.
Survivability	Database environment must gracefully survive from unintentional loss or destruction of a component.
Physical protection	Database environment must protect itself from physical assault. The typical objectives of physical protection requirements are to ensure that an application or centre are protected against the physical damage, destruction, theft, or replacement of hardware, software, or personnel components due to vandalism, sabotage, or terrorism.
System maintenance	Database environment should prevent authorized modifications (e.g., defect fixes, enhancements, updates) from accidentally defeating its security mechanisms.

4.5 Data Quality

Ensuring the quality of the data obtained through a Naturalistic Driving study is paramount for robust and conclusive subsequent analysis. Data quality procedures should not merely relate to database cleansing but follow the whole chain of the data from acquisition to analysis. Thus, consideration is given to the data quality requirements under the following headings

1. Data collection
2. Data transfer
3. Data storage
4. Database quality control

The majority of the information presented in this section is drawn from the FESTA handbook and deliverables but is equally applicable to ND studies. In addition the lessons learned, relating to quality issues, from the 100 car study, have been reviewed. These lessons and recommendations for future ND studies are grouped according to the following categories in the 100 car study report

a) DAS Installation

b) Hardware and Software Maintenance

c) Data Downloading

However, the comments have been re-categorised and referenced terms of the list of headings 1-4 above for consistency of the reporting.

Any additional information from the state of the art review, not included either in FESTA or the 100 car study, is referenced in the appropriate category. It should be noted however that little new material was found in the review.

4.5.1 Data Acquisition

Data acquisition refers to the installation of equipment and the collection of data.

For the installation, the 100 car study found that customised mounting brackets were required that had sufficient adjustability since it could not be assumed that one mounting system would be suitable for all vehicles. It also recommends that, in order to manage quality control of installation, the selection process should carefully consider the capabilities of the contractor, their willingness to receive specialised training and their typical level of customer service. Strict guidelines for the responsibility for repair and payment for installation problems need to be established together with explicit time frames. The use of participant satisfaction surveys was found to improve the performance level of subcontractors and is advised as are random inspections for quality and consistency. Any personnel responsible for maintenance of equipment should be local to the study area and data down-loaders should be trained to carry out minor repairs.

The 100 car study employed a DAS with a unique remote tracking capability that proved essential for the data down-loaders to locate vehicles. The system also had the capability to transit limited amounts of data directly from the vehicle which could be used for fault detection. The DAS was unobtrusive and difficult to tamper with. Beusen et al (2009) also advocated installing the device out of sight of the driver. This approach however made repairs awkward, sometimes requiring the removal of the DAS which was inherently tricky. The study recommends a more simple removal process in future studies or to ‘modularize’ the DAS so that components can be individually removed for repair.

The DAS used in the 100 car study drained the batteries of several of the cars in which they were installed. Safeguards to prevent this problem included the provision of an internal battery backup system that could be used to operate the system while the vehicle was turned off (e.g., when data were being downloaded), the inclusion of a software switch that turned the DAS off if the voltage of the car battery dropped below 11 V, and the inclusion of a "suicide" feature that automatically shut the DAS down when the vehicle was turned off (except for data download purposes). However, these safeguards failed in some cases. For example, the system would keep running after the car was turned off, which occurred when the operating system was not working properly. In those instances, the system kept resetting the CPU to restart the operating system, thereby draining the vehicle battery in the process. These incidents inconvenienced the participants, and should be avoided to the largest extent possible in future efforts. Better sensors and more robust system shutdown algorithms can be created to address the majority of these issues, and should be implemented in the future.

Installation of all of the sensor equipment should be monitored. When necessary, installation verification tools should be created and used to ensure consistency between vehicles.

The 100 car study found that Window98 was not the best operating system for the DAS and recommends that future systems use a Linux-based DAS.

FESTA makes the following recommendations;

- The minimum requirement for robustness is that the entire system should operate under the normal driving conditions for the specific FOT including the harsher situations of normal driving.
- When controlling the power supply to the DAS, the start-up and shutdown speeds must be optimised to reduce the 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. As much as 80% of the DAS hardware problems can be deduced to physical connector issues. Too high and too low temperatures (both static and transient) do affect the DAS. Components with moving parts need special attention.
- Attaching any equipment to the in-vehicle CAN-bus systems has to be done very carefully. Failure to adhere to this may be dangerous and result in vehicle operational malfunction that may result in significant cost, injury or death, or produce other very unwanted results. Beusen et al (2009) also note the importance that the DAS does not interfere with the engine management system.
- When acquiring sensor data from a vehicle CAN-bus the information is passed through several stages before it can be read from the CAN-bus. These stages are likely to affect the signal value both in terms of amplitude and need to be carefully observed.
- To ensure data validity and quality, a calibration and verification scheme is recommended. For data quality aspects it is important that all installed systems of the same category are calibrated and verified using the same procedures. During the verification process, a full dataset should be recorded for the analysts and quality management team in order for them to verify that the installation adheres to the analysis requirements.

Where video footage is being collected, the following should also be observed.

- Direct real time observations should be carried out with great care and as unobtrusively as possible to minimise the risk of the driver modifying his/her driving behaviour.

- The number and resolution and views captured by the cameras should be sufficient to address the hypotheses.
- Pre evaluation of video image quality should be undertaken.

4.5.2 Data Downloading / Transfer

Data transfer is the physical transfer of data usually over a point-to-point or point-to-multi-point communication channel. When data have been collected by the vehicle or system DAS, there is an obvious need to transfer these data from the respective systems to a data storage facility or location.

Essential to successful downloading is access to the vehicle and the 100 car study found that some participants were not fully co-operative. Obtaining participants regular schedules aided the process. Detailed logs should be kept of when downloads occurred for each vehicle allowing prioritisation based upon remaining DAS storage capacity for each vehicle. The data down loaders thus have to be prepared to work flexible hours in order to optimise convenience for the participant along with remaining storage capacity.

The downloading process requires consideration of the of data security, archiving and storage issues (see separate section). Careful logging of data received by the servers should be compared to similar logs of data sent to ensure that no data are missing. Backups were stored in more than one location. The 100 car study stresses the importance of close communication and interaction between data managers and server managers to ensure a smooth and complete data flow.

FESTA provides the following general guidance;

- When the point is reached where data transfer is required, checking procedures should be implemented to ensure that all collected data is backed up and stored in a safe place in order to minimise data loss. The aim is to prevent data loss, verify data completeness and to prevent data storage waste.
- Data back-up should be initiated to prevent data loss by having multiple sets of the data stored in different places. Beusen et al (2009) report that CAN data was stored on internal memory card of the on-board logger and transferred to central sever by the GPRS-modem on a daily basis
- Data verification is aimed to assure that no data is lost during data transfer and data back-up.
- Once data transfer has taken place and suitable data back-ups have been created, the test-site should consider data deletion to ensure that storage space is newly available in the vehicle.
- Experience from previous studies suggests that data loss at the retrieval/upload stage is common even if it could be almost totally avoided with a robust and well-tested procedure. To prevent data loss during the data upload/retrieval procedure it is important to use a verification process to check that the data are consistent before deleting it from the vehicle. In the case that the verification process reveals that the data are not consistent, the vehicle data logger should be checked as soon as possible.
- A process should be considered whereby synchronisation with subjective data can be ensured. Previous experience suggests that the accuracy needed in most cases is less than 5 seconds. For post-hoc structured comments or questionnaires on video or events, it is important to define a process of linking these events to the time.

4.5.3 Data Storage

It is expected that ND studies will generate thousands of hours of raw data during the collection phase. Therefore the quantity of data needs to be handled by an appropriate data management process to ensure data quality, to avoid data loss and to provide ease of access to the data analysts. Data storage has implications for data quality and the following points, from the FESTA guidelines should be considered;

1. The main aim of the storage capacity estimation is to guarantee the availability of free space for recording the vehicle data. Ideally the sample rate for each signal should be the lowest possible able to guarantee no information (relevant for answering the research questions and hypotheses) is lost in the sampling process.
2. If there is no space available on the storage device this would inevitably result in data loss. Therefore, a 20% to 50% on storage size tolerance is recommended.
3. A safe data deletion procedure implies that no data should be deleted in the vehicle until a copy of the data has been backed-up, verified and stored in a safe place.
4. Storage of all data but in particular video data should be in a relational database. Implementation should consider what to do in the event of a data loss from a sensor (for example, a null value could be inserted).

4.5.4 Database quality control

This section is based predominantly on the experience of the UK in-depth accident database manager at the VSRC in Loughborough. As database materials are completed, data will be entered and stored on a central database. The quality of this database depends on good management of the case materials and a well-designed user interface for data input and data downloads. Oracle, SQL Server and MS-Access are examples of database applications available to serve as the central database. A logical hierarchy of relationships between the component data tables, if applicable, is an important foundation.

The study may generate digital case materials that cannot be incorporated into the central database. This could include images, video, sensor output, and time-location datastreams. It is important that these files are systematically named according to rigorous protocols so that they can be identified by computer logic. This also applies to case directories and folders.

A user-friendly data input system with the capability for validation checks is necessary to create a good quality database.

The management of data records - creation (especially), modification and deletion - is the first general requirement. The system should also respond interactively to data input by only showing relevant sections of the forms, hiding those that are irrelevant or not applicable to the case at hand.

A very simple but critical aspect of the design of the database is to ensure that a value has been entered for each field (even if is ‘Not Known’ or ‘Not Applicable’) and the data that are entered are valid. A warning should be issued at data entry for values that are valid but extreme, rare or otherwise improbable.

At certain stages of data input it is recommended to have the user "sign-off". This signals the completion of part or all of the data input stage

Figure 4.1 below illustrates the database quality control.

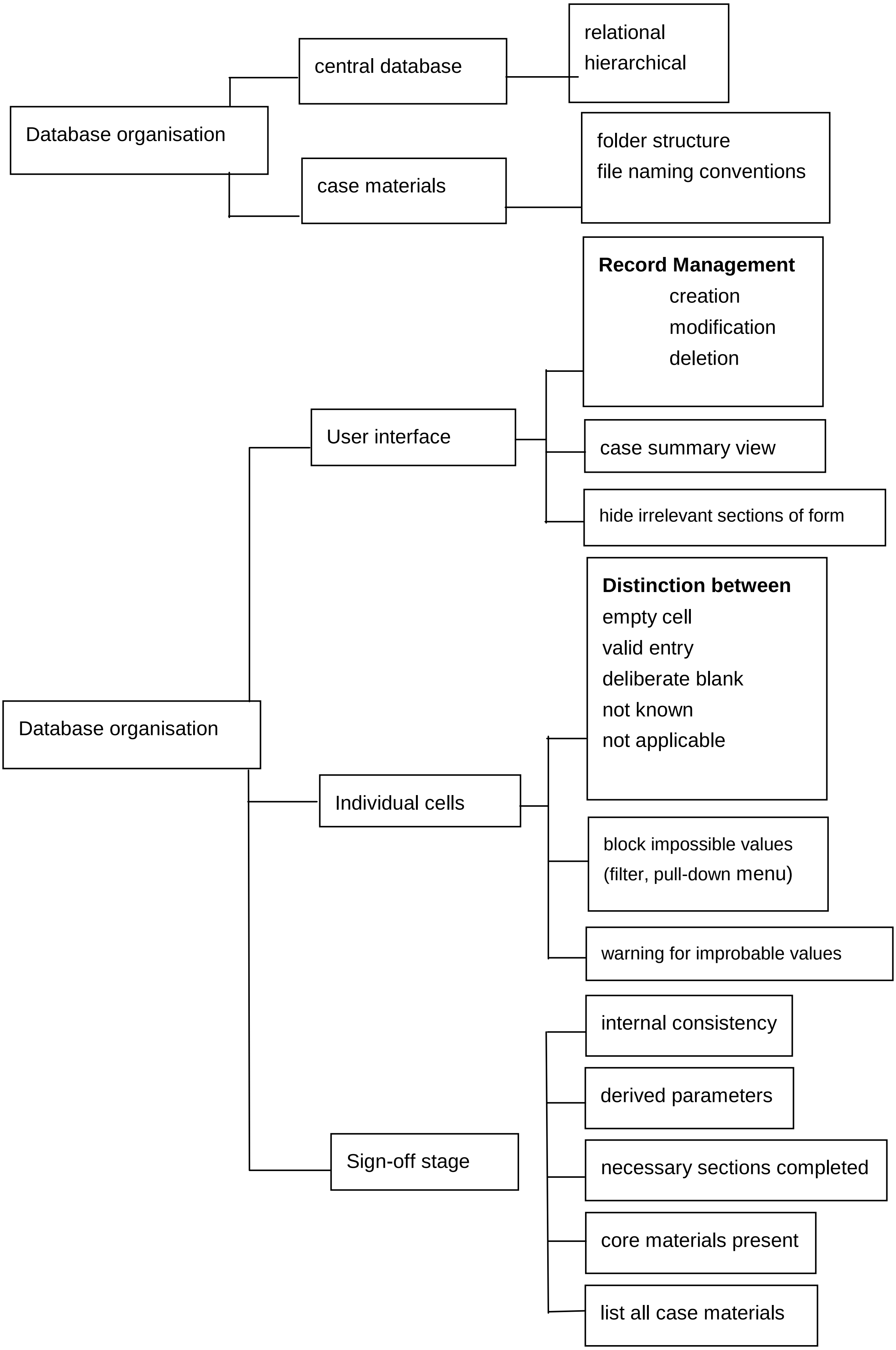

Figure 4.1 Database quality control

5 Data analysis tools including Data Reduction

5.1 Software functionality

The FESTA Handbook lists a number of functions that need to be provided by the chosen software and hardware to support data analysis [p84]:

- Database query functionality (e.g. SQL)
- Signal processing of numerical data
- Fully customizable mathematical computation, analysis, and algorithm development functionality, automatic or semi-automatic calculation of performance indicators, and 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 results function to tabular format or statistical packages.

The Handbook recommends that a number of software packages should be utilised to form an analysis package. Examples given were using SQL software for database queries; for computation, mathematical software such as Matlab could be used and for statistical analysis, widely used statistical analysis applications would be appropriate (e.g. SPSS). It also suggests that custom solutions (proprietary) with user friendly graphical interfaces may be most appropriate for the analysis of large and complex datasets.

The Test Site Sweden FOT used a similar variety of tools in their analysis software package. Database query, video and annotation tools were developed separately from graph functions and calculation functions. As previously mentioned, a database search interface was developed for the project using the SQL. In addition to this a Component Object Model (COM) was designed to enable researchers to analyse the database using their preferred software (in this case Matlab). The developed COM enabled applications to retrieve data by specifying either a complete SQL query or a trip's identification number. This produced data in a suitable format to enable further data analysis. The processing speed of the COM was trialled using Matlab using a fairly complex script that involved the repeated loading of entire trips and processing them according to basic trigger levels using a data from a number of signals (vehicle data such as that from sensors). In this case the loading and processing 10 hours of signal data from the server (30 signals sampled with 10 Hz) took about 7 seconds. This processing speed was concluded to be adequate.

Custom software was developed to create an interactive viewer and annotation tool that allowed the viewing of video and numerical values and manually adding annotations to data. The aim was to utilise, wherever possible, existing applications and Matlab was chosen to be a basis of the analysis tool as it was thought to be the most appropriate calculation/quantitative analysis tool due to its flexible and modular analysis environments. In addition, Matlab was considered to be user friendly and have a thorough library of mathematical functions and algorithms. Graph functions were created as an add-on to the standard Matlab package.

The different tools were designed to interact to appear as one analysis package. The COM interface described above was designed to obtain a real-time connection between the tools and Matlab's automation functions were used to receive information from the video tool which allowed continuously updated plots in Matlab. A graphical user interface was developed within Matlab to allow user interaction in this process.

The Test Site Sweden FOT report gave the most detail about the software they used for data analysis as most of the literature concentrates on the analysis methodologies. However other study reports give some details of some aspects of their analysis. Data reduction/event identification and conversion of that data into other formats for more detailed data analysis was the most common processors to be documented. Stuttes et al. (2005) report that video data was coded manually by researchers viewing the video. Once this was finalised it was converted into a SAS data file to allow in-depth analyses including statistical testing. Barr et al. (2003) state that a specific software program was used to perform eye glance data reduction. The program was called Micro DAS MPEG Video Player and was developed by the NHTSA Vehicle Research and Test Centre.

Lee et al (2004) report that a lane change data integration and analysis program was developed to support data analysis. This program combined data from the many data sources including video, radar, sensor and additional data that had been entered in an excel spreadsheet. Data packets were available for statistical analysis once events had been identified, categorised and entered into the program.

The team at VTTI used basic statistical packages to analyse their large datasets (such as SAS and SPSS) although the derivation of the data to be analysed is made through a data reduction process since the “core” data set itself comprises several hours of video footage. Basically, the data reduction procedures lead to the extraction (through data mining) of specific events using the Data Analysis Reduction Tool (DART) which is used to generate queries generated in SQL. The end product (i.e. the data to be analysed) comprise specific signatures that indicate that an event in the driving process has occurred. The signatures themselves are identified through the application of specific criteria (acceleration/deceleration, swerve, etc) that are applied to the data that are derived from the data loggers.

5.1.1 Data reduction

The 100 car study gives the following advice regarding data reduction;

The data reduction process for this study was developed to record epidemiological data, similar to the GES crash database, as well as record data that has typically been collected in other instrumented vehicle studies, thus greatly augmenting both types of data collection. The 5 channels of video were primarily used to record these variables. However, the data reduction software, developed in-house, allowed the data reductionists to access time plots of the various vehicle sensors (i.e., longitudinal deceleration, vehicle speed) and could be used to record certain other variables as well.

Even with driving performance data and video greatly enhancing the data reduction process, many reduction variables still required a judgment call or subjective analysis on the part of the data reductionist. Many steps were taken to ensure inter-rater reliability and reduce subjectivity among the data reductionists for these types of variables.

First, a two-week training process was provided for each reductionist to allow them to:

- Learn the data reduction software,
- Practice viewing all 5 channels of video,
- Understand the trade-offs of using the video versus using the driving performance time plots, and

- Work with both the lab manager and other trained reductionists to develop a broad understanding of the types of judgments that needed to be made.

Second, all data reductionists were expected to attend weekly meetings in which questions and issues about various data reduction topics were discussed. Third, the lab manager(s) performed spot-checks of all reductionists' work and provided individual feedback to the reductionists. Reductionists were also required to spend 30 minutes each week spot-checking other reductionists' work and providing feedback/discussions to these reductionists. This step was useful for two reasons: (1) it improved accuracy in the database, and (2) it allowed the reductionists to observe other's work and conduct a comparison to their own work, thereby increasing consistency among all reductionists.

Finally, three inter-rater reliability tests were conducted in which the reductionists were all required to validate the same 20 events (per test) and fully reduce two of the twenty events. The test results indicated that there was 88 percent inter-rater reliability for validation of events and 99 percent intra-rater reliability for recording all of the reduction variables. An interesting anecdote is that the inter-rater reliability tests proved to be a very beneficial training tool and will be used from the earliest stages of future data reduction efforts.

Because more information was available to the data reductionists than to the GES analysts who enter information from police-accident reports into the GES database, many of the GES variables were expanded for this study. The GES database is for crashes only, so some of the GES variables were not included in the 100-Car Study database because they were not applicable (e.g., occupant injury, EMS response times). As the reduction process began, the high variability among the events and among the drivers became more apparent. Nevertheless, coding a pre-incident manoeuvre, precipitating factors, contributing factors, and evasive manoeuvres for each event, as well as coding a pre-incident manoeuvre and evasive manoeuvre for each vehicle involved and surrounding the event, appeared to adequately capture the pertinent information for the vast majority of the events. Having the data reductionists write a narrative, or written description of each event, allowed other useful information to be recorded and used for future analyses.

Incorporating 5 video channels into the 100-Car Study DAS was done to ensure the capture of as much of the drivers' view surrounding the vehicle as possible (forward view, rear-view, rear-facing passenger window, and outside the driver's window, via the angled face camera), as well as driver behaviour (face view and over-the-shoulder view). There are trade-offs associated with these 5 camera views, which include size of video files and resolution of the video. Five channels of video increased the bandwidth of the video data, which forced VTTI engineers to decrease the level of resolution of the video so that storage issues would not become problematic. However, the resolution level provided by the system still allowed eye glance reduction to be performed. The resolution levels had a higher effect on discriminating objects and obstacles outside the vehicle. Potholes, for example, were very difficult to identify. Street signs (i.e., speed limit signs) were not readable. Objects inside the vehicle were also sometimes difficult to identify in the camera views. Any problems due to resolution were compounded by night time hours (in which visibility is lower) and sunlight glare (which "washes out" the camera). These aspects also made eye glance reduction much more difficult, although still possible in most cases. While technological advancements in video have already addressed many of these problems, the usefulness of all 5 video channels should be addressed prior to a large-scale study and trade-offs between video resolution and additional channels of video should be weighed carefully.

References

Barr, L. C., Yang, C. Y. D., Ranney, T. A. (2003). An exploratory analysis of truck driver distraction using naturalistic driving data. In Proceedings of the Transportation Research Board 82nd Annual Meeting. Washington, D.C., USA.

Beusen, B., Broekx, S., Denys, T., Beckx, C., Degraeuwe, B., Gijsbers, M., Scheepers, K., Govaerts, L., Torfs, R., Panis, L. (2009). Using on-board logging devices to study the longer-term impact of an eco-driving course. Transportation Research Part D, 14, 514–520.

Dingus, T. A., Klauer, S.G., Neale, V. L., Petersen, A., Lee, S. E., Sudweeks, J., Perez, M. A., Hankey, J., Ramsey, D., Gupta, S., Bucher, C., Doerzaph, Z. R., Jermeland, J., Knipling, R.R. (2006). The 100-Car Naturalistic Driving Study. Phase II – Results of the 100-Car Field Experiment. Report No. DOT HS 810 593. National Highway Traffic Safety Administration. Washington D.C, USA.

FESTA (2008) FESTA Handbook Version 2. Deliverable 6.4 of the EC FP7 project Field opErational teST supporT Action.

LeBlanc, D., Sayer, J., Winkler, C., Bogard, S., Devonshire, J. Mefford, M., Hagan, M., Bareket, Z.,Goodsell, R., Gordon, T. (2006). Road departure crash warning system (RDCW) field operational test final report. Technical report 2006-9-1. University of Michigan Transportation Research Institute, Ann Arbor, MI, USA.

Lee, S., Olsen, E., Wierwille, W. (2004). A Comprehensive Examination of Naturalistic Lane-Changes. Final Report DOT HS 809 702. Virginia Tech Transportation Institute, Blacksburg, VA, USA.

McGehee, D. V., Raby, M., Carney, C., Lee, J. D., Reyes, M. L. (2007). Extending parental mentoring using an event-triggered video intervention in rural teen drivers. Journal of Safety Research 38(2), 215-227.

Regan, M., Triggs, A., Young, K., Tomasevic, N., Mitsopoulos, E.. Stephan, K., Tingvall, C. (2006). On-Road Evaluation of Intelligent Speed Adaptation, Following Distance Warning and Seatbelt Reminder Systems: Final Results of the TAC SafeCar Project. Volume 1: Final Report. Monash University Accident Research centre, Melbourne, Australia.

SAFER (2008) Test Site Sweden Field Operational Test Final Report. SAFER Report 2008:1, SAFER, Gothenburg, Sweden

Stutts, J., Feaganes, J., Reinfurt, D., Rodgman, E., Hamlett, C., Gish, K., Staplin, L. (2005). Driver's exposure to distractions in their natural driving environment. Accident Analysis and Prevention 37(6): 1093-1101.

List of Figures

List of Tables

List of Abbreviations

ABS	Anti-lock Braking System
ANSI	American National Standards Institute
BLOB	Binary Large Object
CAA	Cockpit Assessment Activity
CAN	Controller Area Network
COM	Component Object Model
CPU	Central Processing Unit
DART	Data Analysis Reduction Tool
DAS	Data Acquisition System
DB	Data Base
DVD	Digital Versatile Disc
EU	European Union
FOT	Field Operational Test
FPS	Frames per Second
GPS	Global Positioning System
LIN	Local Interconnect Network
ND	Naturalistic Driving
NTSC	National Television System Committee
OBD	On Board Diagnostic
OE	Original Equipment
OEM	Original Equipment Manufacturer
PAL	Phase Alternating Line
SAS	Statistical Analysis System
SHRP	Strategic Highway Research Programme
SPSS	Statistical Package for the Social Sciences
SQL	Structured Query Language
USB	Universal Serial Bus