



FESTA Handbook

Version 8

September 2021

Updated and maintained by
FOT-Net, CARTRE

and

ARCADE

EU Coordination and Support Actions



Table of Contents

Table of Contents	ii
Acknowledgements.....	ix
List of Abbreviations	xii
Glossary	xv
List of Figures	xviii
List of Tables.....	xix
1 Introduction.....	20
1.1 FESTA project and handbook	20
1.2 FESTA V.....	21
1.3 Data Sharing.....	23
1.4 Cooperative, Connected and Automated Mobility.....	24
2 Planning and Running a Field Operational Test	25
2.1 Introduction	25
2.2 The FOT Implementation Plan	25
2.2.1 Purpose.....	25
2.2.2 Description of the FOT Implementation Plan	26
2.2.3 Development of the FOT Implementation Plan	27
2.2.4 Assumptions underlying the FOT Implementation Plan	27
2.2.5 Using the FOT Implementation Plan	28
3 Legal and Ethical Issues	30
3.1 Introduction.....	30
3.2 Ethics	30
3.3 Legal permissions and cross-border testing	31
3.4 Participant recruitment.....	31
3.5 Participant agreement	31
3.6 Protection of personal data	32
3.7 Proprietary data and stakeholder confidentiality	34
3.8 Risk assessment and safety.....	35
3.9 Insurance	36
3.10 Responsibilities and legal liability	36
3.11 Ethical approval	37
3.13 Iteration	37
3.14 Specific considerations for CCAM FOTs	38
4 From Functions to Hypotheses	39

4.1	Introduction.....	39
4.2	Applying this process to Naturalistic Driving Studies.....	39
4.3	Systems and functions	40
4.4	Vehicle systems.....	41
4.5	Nomadic devices	41
4.6	Cooperative systems.....	42
4.7	Combinations of functions.....	44
4.8	General methodology	46
4.8.1	Step 1: Selection and description of functions	48
4.8.2	Step 2: Definition of use cases and situations.....	50
4.8.3	Step 3: Identification of the research questions	53
4.8.4	Step 4: Creation of hypotheses	55
4.8.5	Step 5: Link hypotheses with indicators for quantitative analyses	60
4.8.6	Iteration	60
4.9	Functions, research questions and baseline in CCAM FOTs.....	61
5	Performance Indicators	63
5.1	Introduction.....	63
5.2	Definition of performance indicators.....	63
5.3	Measures	64
5.3.1	Direct (raw) measures	64
5.3.2	Derived measures.....	64
5.3.3	Self-reported measures	65
5.3.4	Situational variables.....	66
5.4	Events.....	66
5.4.1	Crash Relevant Events	67
5.5	The PI-Measures-Sensors matrix	70
5.6	Performance indicators per impact area	72
5.6.1	Indicators of driving performance and safety	72
5.6.2	Indicators of system performance and influence on driver behaviour.....	73
5.6.3	Performance indicators of environmental aspects	75
5.6.4	Indicators of traffic efficiency	75
5.6.5	Acceptance and trust.....	77
5.6.6	Driver characteristics	78
5.7	Iteration.....	79
5.8	Performance indicators and user acceptance in CCAM FOTs.....	79
6	Experimental Procedures	81

6.1	Study design.....	81
6.1.1	Hypothesis formulation	81
6.1.2	Experimental design	82
6.1.3	Threats to validity: confounds and other interfering effects.....	83
6.2	Participants.....	84
6.2.1	Characteristics	84
6.2.2	Sample size and power analysis	85
6.3	Experimental environment.....	86
6.3.1	Geographical location	87
6.3.2	Road type.....	88
6.3.3	Traffic conditions and interactions with other road users.....	89
6.3.4	Roads to include	90
6.3.5	Weather conditions	90
6.3.6	Time of day and seasonal effects.....	91
6.4	Conducting a pilot study to test the evaluation process.....	92
6.5	Controlled testing.....	94
6.5.1	Operationalisation of tests	96
6.5.2	Operationalisation tool chain	97
6.5.3	Test execution.....	98
6.6	Naturalistic Driving Study	98
6.6.1	Definition	98
6.6.2	Relation to FOTs.....	99
6.6.3	NDS and the FESTA V	100
6.7	Documentation	101
6.8	CCAM FOTs	101
7	Guidelines for Data Acquisition	102
7.1	Measures and sensors	103
7.2	Data acquisition	103
7.2.1	Background data acquisition.....	103
7.2.2	In-vehicle data acquisition	104
7.2.3	Nomadic devices.....	104
7.2.4	Subjective data acquisition	105
7.2.5	Real-time observation.....	105
7.2.6	Additional data sources in cooperative systems	106
7.2.7	Acquisition of infrastructure data and other services	106
7.3	Specific sensors	106

7.3.1	In-vehicle video	106
7.3.2	Internal vehicle bus data.....	107
7.3.3	Automatic in-vehicle driver monitoring.....	108
7.3.4	Non-video environmental sensing	109
7.3.5	GPS.....	109
7.3.6	Audio and driver annotation.....	109
7.3.7	System function/status.....	110
7.3.8	Vehicle metadata	110
7.3.9	Coding/classification/transcription	110
7.3.10	Geographical Information System (GIS).....	111
7.3.11	Communication unit	111
7.3.12	Application Unit	111
7.4	Mechanical requirements	111
7.4.1	Size and weight.....	112
7.4.2	Connectors and interfaces.....	112
7.4.3	DAS mechanical cover and ease of access	112
7.4.4	Crashworthiness and vibration resistance.....	112
7.4.5	DAS environmental requirements.....	112
7.5	Electrical requirements	112
7.5.1	Power management.....	112
7.5.2	Interference with in-vehicle equipment.....	113
7.5.3	Laws and regulations.....	113
7.6	DAS data storage	113
7.6.1	Storage capacity estimation.....	113
7.6.2	Data retrieval/uploading procedure	115
7.7	System configuration	116
7.7.1	DAS inventory management.....	117
7.7.2	Configuration tools and traceability	117
7.7.3	Switching between configurations	117
7.8	Acquisition of data	117
7.8.1	Start-up	117
7.8.2	Acquisition of data.....	118
7.8.3	Shutdown	118
7.9	Synchronisation	118
7.9.1	Timestamping versus real world event.....	118
7.9.2	Integrated sensing synchronisation	118

7.9.3	Synchronisation with nomadic devices.....	119
7.9.4	Synchronisation of infrastructure systems.....	119
7.9.5	Synchronisation of cooperative systems	119
7.9.6	Synchronisation with interviews and other subjective sensors	119
7.10	DAS status and malfunction management.....	120
7.10.1	System status uploads.....	120
7.10.2	Malfunction management	120
7.10.3	Spare system management.....	120
7.11	System installation.....	120
7.11.1	Installation procedures.....	120
7.11.2	Installation verification and calibration.....	121
7.11.3	Dismounting the system	121
7.12	Proprietary data	121
7.13	Personal integrity and privacy issues in data acquisition and analysis	122
7.14	On-line quality management procedures	122
7.14.1	Remote automatic upload.....	123
7.14.2	Automatic and manual quality checks	123
7.14.3	DAS and sensor maintenance	123
7.15	Data acquisition in FOTs of automated driving	123
8	Guidelines for Databases and Analysis Tools	125
8.1	Database design and implementation.....	125
8.1.1	Preferred database models	125
8.1.2	Data categories in the database.....	126
8.1.3	User data spaces and data collaboration	127
8.2	IT infrastructure implementation.....	128
8.2.1	Video data storage.....	128
8.2.2	FOT Hardware	128
8.2.3	Distributed system at many locations.....	128
8.2.4	Physical access	129
8.2.5	Logical access	129
8.2.6	Personal integrity and confidential data.....	129
8.2.7	Backup	129
8.2.8	Risk management.....	130
8.2.9	Database acceptance	130
8.3	Off-line quality management procedures	130
8.3.1	Quality assurance of objective data.....	130

8.3.2	Quality assurance of subjective data.....	130
8.3.3	Measurement naming guidelines.....	131
8.3.4	Automatic pre-processing and performance indicator calculations during data upload.....	131
8.4	Data analysis tools	132
8.4.1	Data classification/transcription	132
8.4.2	Time history and event analysis with video	132
8.5	Database usage	133
8.6	Databases and analysis tools for FOTs of automated driving	133
9	Data Analysis and Modelling.....	135
9.1	Introduction.....	135
9.2	Large dataset handling.....	136
9.3	Consistency of the chain of data treatment.....	137
9.4	Precision in sampling.....	138
9.5	Pooling of data.....	140
9.6	Requirements for integration	140
9.7	Scaling up.....	141
9.7.1	Scaling up using statistics.....	142
9.7.2	Scaling up using (multimodal) traffic models.....	143
9.8	Appropriate techniques at the five links of data analysis	144
9.8.1	Step 1: Data quality analysis	144
9.8.2	Step 2: Data processing	145
9.8.3	Step 3: Performance Indicator calculation.....	146
9.8.4	Step 4: Hypothesis testing.....	146
9.8.5	Additional Step 4: Data mining	147
9.8.6	Step 5: Generalisation of results	147
9.9	Data analysis in FOTs of automated driving.....	149
10	Impact Assessment and Socio-Economic Cost–Benefit Analysis	151
10.1	Introduction.....	151
10.2	Considerations.....	152
10.2.1	Re-use of data in impact assessment	152
10.2.2	Deployment scenarios	153
10.2.3	General issues of the socio-economic impact analysis.....	154
10.2.4	Assessment scope and process implication.....	155
10.2.5	Geographical scope of assessment	156
10.3	Analysis of impacts.....	156
10.3.1	Safety benefits	157

10.3.2	Mobility benefits	162
10.3.3	Efficiency benefits	164
10.3.4	Environmental benefits	165
10.3.5	Integration of results	165
10.4	Socio-economic Cost–Benefit Analysis.....	166
10.4.1	System costs.....	166
10.4.2	Classification of assessment methods	167
10.5	How to carry out the CBA assessment	168
10.5.1	CBA framework.....	169
10.5.2	Inputs for impact assessment (including cost estimation).....	169
10.5.3	Data needs.....	170
10.5.4	Impact valuation.....	176
10.5.5	Results	176
10.6	Socio-economic impact assessment in CCAM FOTs	180
11	Sharing knowledge, lessons learned and data	181
	References.....	184
	Annex A FOT Implementation Plan (FOTIP).....	191
	Annex B Informed Consent Checklist.....	225
	Annex C Checklist for Participant Introduction	227

Acknowledgements

We would like to thank everyone who contributed to this handbook.

The original document and its first update (Version 2) were produced by the FESTA project consortium in 2008. FESTA (Field opERational teSt support Action) was an eight-month European support action that compiled methodology and created a handbook to guide upcoming Field Operational Tests (FOTs).

Since the FESTA project, the handbook was long maintained and continually updated by FOT-Net, a networking operation open to anyone interested in FOTs and run by the following European Support Actions: FOT-Net (2008–2010), FOT-Net 2 (2011–2013) and FOT-Net Data (2014–2016). In 2016, FOT-Net merged with another EU-funded support action, VRA (Vehicle Road Automation), to form a new networking forum and to coordinate automated road transport research projects. This collaboration has so far resulted in two new support actions: CARTRE (2016–2018) and ARCADE (2018–2021). These projects have compiled the latest two updates of this handbook.

Updates to Version 3 of the FESTA Handbook were compiled by a working group as part of FOT-Net 2, and are based on the results of seminars organised during the previous FOT-Net project and feedback from users of Version 2. Version 4 took into account the feedback received following publication of Version 3 (open consultation July–August 2011) and during a workshop in Gothenburg, Sweden (8 September 2011).

Version 5 was the product of five Working Groups participating in FOT-Net 2 in 2012 and 2013 (Ljung-Aust et al., 2014). The Working Groups focused on the following topics: data analysis, events and incident definition, legal and ethical issues, impact assessment and scaling up, and data sharing. They proceeded in parallel, organising dedicated workshops and setting up links to running FOTs. A final expert workshop was organised in Turin, Italy (4–5 November 2013) to validate and agree on amendments to the handbook.

Version 6 contained updates made by the FOT-Net Data consortium in December 2016 with a focus on data sharing. Also, Data Sharing Framework was published to complement FESTA.

Version 7 in December 2018 was published by the CARTRE (Coordination of Automated Road Transport Deployment for Europe) Coordination and Support Action, addressing the way in which FESTA can be used in projects on testing and evaluating automated vehicles.

This latest version 8 by ARCADE project (Aligning Research & Innovation for Connected and Automated Driving in Europe) compiles the recent lessons especially from ongoing automated driving tests and continues the updates initiated by CARTRE.

This handbook is not owned by a single organisation but belongs to the FOT community. Editorial responsibility is considered to be with the most recent testing-related networking forum following FESTA, FOT-Net and ARCADE legacy.

The following organisations (or individuals from organisations) have contributed directly or indirectly to one or more revisions of the FESTA handbook as members of at least one of the consortia of the FESTA, FOT-Net, FOT-Net 2, FOT-Net Data, CARTRE and ARCADE Coordination and Support Actions:

A.D.C. Automotive Distance Control Systems GmbH	Germany
Association of French Motorway Companies (ASFA)	France
BMW	Germany
Chalmers University of Technology (SAFER)	Sweden
CEESAR	France
Daimler	Germany
Delphi France	France
European Road Transport Telematics Implementation Coordination Organisation (ERTICO)	Belgium
Federal Highway Research Institute (BAST)	Germany
FIAT Research Centre (CRF)	Italy
Fraunhofer Society for the advancement of applied research	Germany
French National Institute for Transport and Safety Research (INRETS)	France
Galicia Automotive Technological Centre (CTAG)	Spain
INFOBLU SPA	Italy
International Automobile Federation (FIA)	Belgium
ITS Bretagne	France
Loughborough University	United Kingdom
Ministry of Transport, Public Works and Water Management (RWW)	Netherlands
MUARC	Australia
Netherlands Organisation for Applied Scientific Research (TNO)	Netherlands
Orange France	France
Promotion of Operational Links with Integrated Services, Association Internationale (POLIS)	Belgium
Renault	France
Robert Bosch GmbH	Germany
Swedish National Road and Transport Research Institute (VTI)	Sweden
Technical University of Aachen (IKA)	Germany
UMTRI	USA
University of Cologne	Germany
University of Leeds	United Kingdom
Volvo Car Corporation	Sweden
Volvo Technology Corporation	Sweden
VTT Technical Research Centre of Finland Ltd	Finland
VTTI	USA

Disclaimer:

The FESTA, FOT-Net 2, FOT-Net Data, CARTRE and ARCADE Coordination and Support Actions were funded by the European Commission in the 7th Framework Programme and Horizon2020. The content of this publication is the sole responsibility of the project partners listed herein and does not necessarily represent the view of the European Commission or its services.

List of Abbreviations

ABA	Aggregation Based Analysis
AC	Alternating Current
ACC	Adaptive Cruise Control
ACEA	European Automobile Manufacturers' Association
AD	Automated Driving
ADAS	Advanced Driver Assistance Systems
ART	Automated Road Transport
BCR	Benefit Cost Ratio
BLOB	Binary Large Object
C2C-CC	Car-to-Car Communication Consortium
CALM	Continuous Air interface for Long and Medium range communication
CCAM	Cooperative, Connected and Automated Mobility
CAM	Cooperative Awareness Message
CAN	Controller Area Network
CARTRE	Coordination of Automated Road Transport Deployment for Europe
CAV	Connected and Automated Vehicle
CBA	Cost Benefit Analysis
CEN	Comité Européen de Normalisation (European Committee for Standardization)
CPI	Consumer Price Index
CPU	Central Processing Unit
CRE	Crash Relevant Events
DAQ	Data Acquisition
DAS	Data Acquisition System
DBA	Database Administrator
DC	Direct Current
DENM	Decentralised Environmental Notification Message
DGPS	Differential Global Positioning System
EBA	Event Based Analysis
ECU	Electronic Control Unit
EDR	Event Data Recorder
EMC	Electromagnetic Compatibility
ETSI	European Telecommunications Standards Institute
Euro NCAP	European New Car Assessment Programme
FCW	Forward Collision Warning
FESTA	Field opERational teSt supporT Action
FFM	Five Factor Model
FM	Frequency Modulation
FOT	Field Operational Test
GDP	Gross Domestic Product
GIS	Geographical Information System
GMaps	Google Maps
GMM	Gaussian Mixture Models
GNSS	Global Navigation Satellite System

GPS	Global Positioning System
GSM	Global System for Mobile communications
GUI	Graphical User Interface
HMI	Human Machine Interface
I2V	Infrastructure to Vehicle (see Vehicular communication systems)
ICT	Information Communication Technology
IEEE	Institute of Electrical & Electronics Engineers
I/O	Input/Output
IRR	Internal Rate of Return
ISA	Intelligent Speed Adaptation
ITS	Intelligent Transportation System
LDW	Lane Departure Warning
LED	Light Emitting Diode
LIDAR	Light Detection and Rating
LIN	Local Interconnect Network
MJPEG	Motion JPEG
MOST	Media Oriented Systems Transport
MP3	Moving Picture Experts Group Layer-3 Audio
MPEG	Moving Picture Experts Group
MPEG-4	MPEG standard, version 4
NAS	Network Attached Storage
ND	Nomadic Device (often referred as Mobile Device)
NDD	Naturalistic Driving Data
NDS	Naturalistic Driving Study
NHTSA	National Highway Traffic Safety Administration
NPV	Net Present Value
NTP	Network Time Protocol
OEM	Original Equipment Manufacturer (used in this document as a synonym for vehicle manufacturers)
ODD	Operational Design Domain
OSM	Open Street Map
PI	Performance Indicator
PND	Personal Navigation Devices
RAID	Redundant Arrays of Independent Disks
RR	Range Rate
RSU	Road Side Unit (see Vehicular communication systems)
SAN	Storage Area Network
SMS	Short Message Service
SPSS	Statistical Package for the Social Sciences
SQL	Structured Query Language
SSD	Solid State Drive
SSS	Sensation Seeking Scale
T-LOC	Traffic Locus of Control
TTC	Time To Collision (relevant parameter in preventive safety applications)
USB	Universal Serial Bus
UTC	Universal Time Coordinated
V2I	Vehicle to Infrastructure (see Vehicular communication systems)
V2V	Vehicle to Vehicle (see Vehicular communication systems)

V2X	Vehicle to X (Any Station) (see Vehicular communication systems)
VMS	Variable Message Sign
WBS	Work Breakdown Structure
WLAN	Wireless Local Area Network
WTP	Willingness To Pay
WYLFIWYF	What-You-Look-For-Is-What-You-Find
ZFS	Zettabyte File System

Glossary

Accident	A traffic collision (motor vehicle collision, motor vehicle accident, car accident, or car crash) is when a road vehicle collides with another vehicle, pedestrian, animal, road debris, or other geographical or architectural obstacle. Traffic collisions can result in injury, property damage, and death.
Application	A software running on a system implementing one or more functions.
Baseline	A reference point for comparison purposes.
Baseline period/phase	Baseline period/phase is the part of the data collection during which the function(s) operate in "silent mode", that is, they collect data, but do not give any signals to the driver. From the viewpoint of the driver the function(s) is/are off.
Baseline within comparison situation	Baseline within comparison situation is the scenario with system under evaluation "turned off".
Comparison situation	Comparison situation is the situation that will be used for comparison within one analysis.
Controlled factors	Controlled factors are those factors that are kept constant within one analysis. The data are filtered such that only occurrences in which the controlled factors assume the intended values are selected.
Cooperative Systems	Vehicle systems based on vehicle-to-vehicle (V2V), vehicle-to-infrastructure (V2I) and infrastructure-to-vehicle (I2V) communication technology.
Data acquisition	Data acquisition is the process of sampling or recording data (real world data) for computer processing. It includes acquisition of pure sensor data, as well as acquisition of data from real-time and off-line services, and self-reported data.
Derived measure	Derived measure is a single measure calculated from a direct measure (e. g. by applying mathematical or statistical operations) or a combination of one or more direct or derived measures.
Direct measure	Direct measure are measures logged directly from a sensor, without further manipulations except linear transformations.
Download	Download is the transfer of data from server to a client.
Event	Event is something that happens in a specific period of time for which one or several criteria are fulfilled.
Event data recorder	Event data recorder is a logging device that, when triggered by an event such as a crash, stores the information about the few seconds leading up to the event (and throughout the event).
External measure	External measure is an external measure is provided by data sources outside of the log equipment used in the study.

FOT	FOT is a study undertaken to evaluate a function, or functions, under normal operating conditions in environments typically encountered by the host vehicle(s) using quasi-experimental methods.
Function	Function is the activity or purpose of a system to be tested in the FOT.
Hypothesis	Hypothesis is a specific statement linking a cause to an effect and based on a mechanism linking the two. It is applied to one or more functions and can be tested with statistical means by analysing specific performance indicators in specific scenarios. A hypothesis is expected to predict the direction of the expected change.
Incident	Incident is something unforeseen in the course of action. In driving a vehicle in traffic, something which changes the foreseeable action (speed, direction) of the vehicle.
Latency	A latency period: the time between stimulus and response. In data acquisition generally the time between real world event (or stimulus) and the recording of that event.
Measure	Measure is the magnitude of a quantity such as distance, relative to a unit of measurement, such as a meter.
Near crash	Near crash is defined as a conflict situation requiring a rapid, severe evasive manoeuvre to avoid a crash.
Naturalistic Driving Study	A research method using advanced technology for in-vehicle unobtrusive recording of driver behaviour during ordinary driving in everyday traffic situations.
Performance indicator	Performance indicator is a quantitative or qualitative indicator, derived from one or several measures, agreed on beforehand, expressed as a percentage, index, rate or other value, which is monitored at regular or irregular intervals and can be compared to one or more criteria.
Pilot study	Small-scale version, or trial run, done in preparation for the major study.
Research question	A question to be answered by the FOT, e.g. by compiling and testing related specific hypotheses.
Scenario	A use case in a specific situation.
Self-reported measure	Self-reported measure is a subjective data reported via questionnaires, interviews, focus groups, etc.
Sensor	Sensor is a device that responds to a physical stimulus (as heat, light, sound, pressure, magnetism, or a particular motion) and transmits a resulting impulse which can be interpreted as a measure by an instrument/observer.
Situation	Situation is one specific level or a combination of more specific levels of situational variables.
Situational variable	Situational variable is an aspect of the surroundings made up of distinguishable levels. At any point in time at least one of these levels must be valid.

System	System is a combination of hardware and software enabling one or more functions.
System state	System state is the current status of a system based on its setting.
Test scenario	Test scenario is the test setup in which scenarios are triggered in order to collect data specific to this scenario.
Treatment period/phase	Treatment period/phase is the part of the data collection during which the function(s) are switched on by the experimental leader, such that they are either active all the time, or can be switched on or off by the driver.
Trigger	"Marker" in the data, indicating instances that can be of interest for research.
Trip	Journey from one place to another. (Project-specific definitions are common.)
Upload	Upload is the transfer of data from client to a server.
Use case	A set of operations in which a system or its specific function is expected to interact with a user or another system to achieve a goal
Vehicle bus	Vehicle bus is an in-vehicle internal communications network that connects different components and modules.

List of Figures

Figure 1.1 FESTA V: The steps that typically have to be considered when conducting an FOT	21
Figure 2.1 Iterative development process	27
Figure 4.1 The three-level model of the driving task, based on Michon (1985).....	57
Figure 5.1 FESTA Traffic efficiency estimation based on FOT results	76
Figure 6.1 Total sample size as a function of the statistical power and the effect size (2-sided test, alpha = 0.05, independent variables).....	86
Figure 6.2 Operationalisation of test scenarios	97
Figure 6.3 A partially overlapping relationship between FOTs and NDS along a continuum of experimental control.	99
Figure 7.1 Data structuring	102
Figure 7.2 Illustration of process for securing proprietary vehicle-bus data using a gateway	108
Figure 7.3 Factors influencing data size.....	114
Figure 7.4 Data retrieval/uploading steps.....	115
Figure 9.1 Block diagram for data analysis	135
Figure 9.2 Deployment of the chain with feedbacks and additional models.....	138
Figure 9.3 Block diagram of data quality analysis	145
Figure 9.4 Block diagram for the procedure of data processing	145
Figure 9.5 Block diagram of translating FOT indicators to large-scale effects	148
Figure 10.1 Scope of the impacts within socio-economic impact assessment.....	157
Figure 10.2 An overall schematic picture of the safety impact assessment in eIMPACT	160
Figure 10.3 TeleFOT mobility model (Innamaa et al., 2013)	163
Figure 10.4 Classification of socio-economic assessments.....	168

List of Tables

Table 2.1 A generic guide to scheduling the 22 activities described in the FOTIP in Annex A of the FESTA Handbook	28
Table 4.1 Use cases, situations, scenarios, and their mutual dependence.....	50
Table 5.1 Example on defining a performance indicator (or a derived measure).....	71
Table 5.2 Example on defining a measure (or a signal).	72
Table 6.1 Complementary uses of naturalistic and controlled tests in cooperative system evaluation.....	94
Table 7.1 Pros and cons of different data retrieval/uploading modes.....	115
Table 7.2 Methods and issues in calculating latency for in-vehicle data sources	119
Table 10.1 An example of summary table of the annual social value of impacts of Motorway Automated Driving Functions (ADF) for different penetration rates, and the annual social costs of implementing the system	156
Table 10.2 Social CBA tabulation.....	177
Table 11.1 Template for gathering and documenting lessons learned.....	181

1 Introduction

1.1 FESTA project and handbook

Field Operational Tests (FOTs) as a concept gained popularity in the automotive sector in early 2000s as an evaluation method for driver support systems and functions. The aim at that time was in proving that such systems can deliver real-world benefits. A number of FOTs evaluating vehicle information and communications technologies (ICT) have been since conducted at a regional, national, European and international level. These FOTs have proven to be highly valuable and have been identified as an important means of verifying the real-world impacts of new systems, and as a means to verify that previously conducted Research & Development has the potential to deliver identifiable benefits.

The **FESTA** (Field opErational teSt support Action, 2007–2008) project was to produce comprehensive guidance to facilitate the successful delivery of soon-to-follow first European FOTs. The original version of this handbook was their main deliverable. This, the latest version of the FESTA Handbook, has drawn from experience and knowledge across numerous FOTs in order to revise and update the best practice presented here. This version 8 is the result of a joint effort of several research institutes, OEMs (Original Equipment Manufacturer) and other stakeholders from across the World to prepare a common methodology for FOTs.

For the purposes of this Handbook, a “Field Operational Test” (FOT) is defined as:

A study undertaken to evaluate a function, or functions, under normal operating conditions in road traffic environments typically encountered by the participants using study design so as to identify real-world effects and benefits.

This means that it must be possible to compare the effects that the function has on traffic with a baseline condition during which the function is not operating. In order to achieve this, the participants’ control over or interaction with the function(s) has to be manipulated by the research team. “Normal operating conditions” implies that the participants use the platforms during their daily routines, that data logging works autonomously and that the participants do not receive special instructions about how and where to drive. Except for some specific occasions, there is no experimenter in the vehicle, and typically the study period extends over at least a number of weeks.

More generally, FOTs are large-scale user tests where, for example, a hundred participants are recruited to try out a system (or a function). The period of testing has commonly varied between a few months and two years. During this testing period, questionnaires, measurements and observations are made to identify how the system potentially changes the participants’ driving and travelling behaviour. FOTs also study the effects on other road users and wider impacts on the transport system and the society. Since the FESTA Handbook was first published, its methodology has been applied in a range of different test campaigns, from mobile phone related testing to assessment of vehicle-to-vehicle communication systems.

This handbook also addresses Naturalistic Driving Studies (NDS), where the purpose is not the testing of functions but the observation of driver (or rider) behaviour in everyday traffic situations, using advanced technology for in-vehicle unobtrusive recording.

The main purpose of this Handbook is to provide guidelines for conduct of FOTs and NDS. For simplicity, mostly the term FOT is used and differences between the two kinds of studies are summarised where relevant.

The FESTA Handbook walks the reader through the whole process of planning, preparing, executing, analysing and reporting an FOT and it gives information about aspects that are especially relevant for a study of this magnitude, such as administrative, logistic, legal and ethical issues. Another aspect of the Handbook is to pave the road for standardisation of some aspects of FOTs, which would be helpful for cross-FOT comparisons. It has to be kept in mind, though, that many traffic parameters are different between countries.

1.2 FESTA V

In Figure 1.1, the “FESTA V”, the steps that need to be carried out during an FOT are presented in the form of a V diagram. There is correspondence between the levels on the left-hand and right-hand sides. The steps are explained in detail in the different chapters of the Handbook. For orientation purposes, a copy of the figure is provided at the start of each chapter, highlighting which step of the FOT chain is described in the current chapter. Additionally, the FOT implementation plan ([Annex A](#)) takes up all the steps and integrates them into a detailed table, which can be used as a reference when carrying out an FOT.

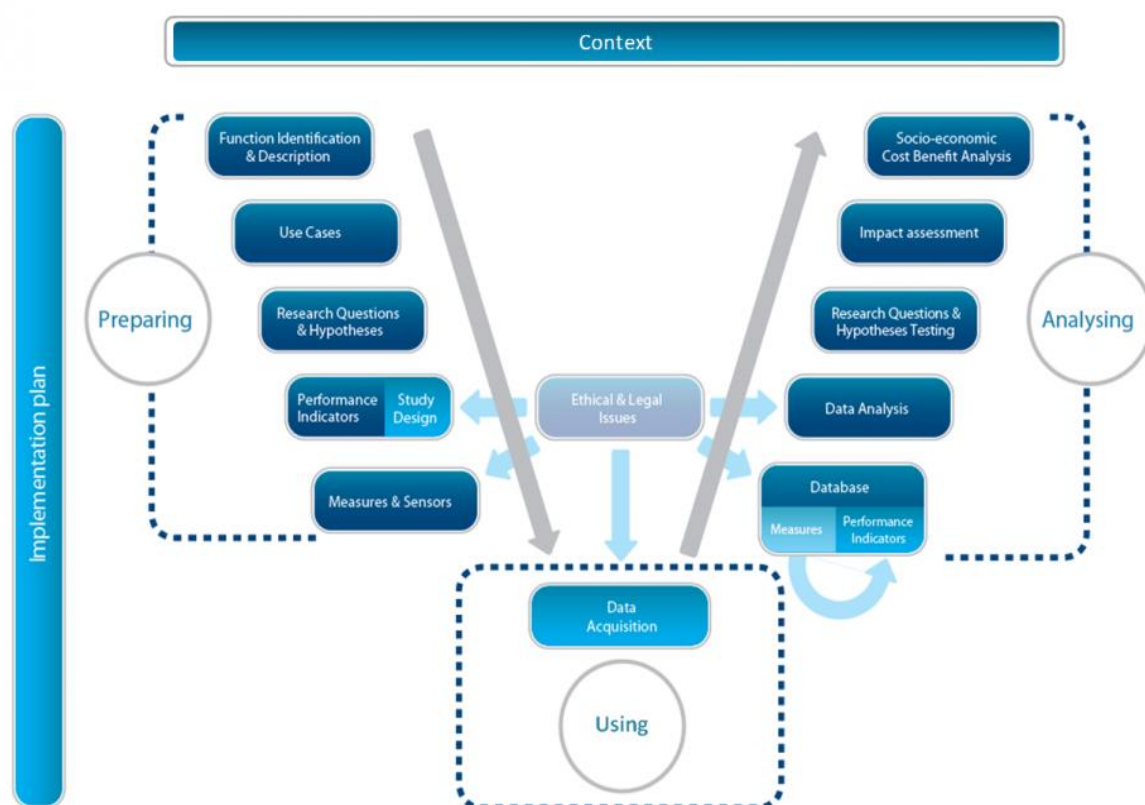


Figure 1.1 FESTA V: The steps that typically have to be considered when conducting an FOT

In order to make the picture more complete, a horizontal bar has been added on top of the diagram that in principle summarises the context in which the FOT is supposed to take place. For instance, the choice of a function to be tested implies that there is either a problem that is to be addressed and that the chosen function is defined to solve the problem, or that a policy objective is stated and that the function tested can be used to reach the objective. An FOT can always be related to a wider perspective than is defined by just a description of the function to be tested.

The top of the V covers setting up a goal for the study and selecting a suitable research team, and also the last steps that include an overall analysis of the systems and functions tested and the socio-economic impact assessment, dealing with the more general aspects of an FOT and with aggregation of the results. The further down on the FOT Chain V-Shape the steps are located, the more they focus on aspects with a high level of detail, such as which performance indicators to choose, or how to store the data in a database. The ethical and legal issues have the strongest impact on those aspects, where the actual contact with participants, data handling and potential data sharing take place.

Representation of the FESTA methodology in the form of a V does not mean that designing and performing an FOT is always a linear process. Decisions made at a certain stage of the FESTA V influence the next steps, and it is likely that there will be a need to sometimes go back and redo some steps. Especially on the left-hand side of the V, iteration may be necessary. For example, one may find that the measures and sensors available do not make it possible to investigate the hypotheses defined earlier, so adjustments to the hypotheses or performance indicators may be needed. Also, the right-hand side of the FESTA V may influence the decisions to be made on the left-hand side. The need to assess the socio-economic impact may influence the definition of functions, use-cases, research questions and other elements of the left-hand side. Consideration of the resources available for data analysis may also lead to revision of the left-hand side. There is, however, the question “when does one stop the iterative process?” From a research perspective, this is a continuous process. However, from a project management perspective, boundaries have to be set to reflect budget constraints and timing aspects.

The first step in the FESTA V is the identification of functions to be tested. Sometimes this may not be the best step to start with. For example, an FOT may not be driven by the technical systems that need to be tested but by a research question or an impact area. When there is a large set of functions available from which a few need to be selected as candidates for testing, definition of the research questions may help the selection process. For example, if safety is to be investigated, different functions may have a higher priority than when traffic efficiency is the main focus.

The FESTA V provides a static picture of the complex design and execution of an FOT but in reality, a more iterative process will be needed, with a starting point suited to the specific aims of a project.

The FESTA Handbook is not meant to be a substitute for consultations with experts, organising a good and capable research team, or carrying out specific investigations into the legal and ethical issues that apply to the current question and situation. It is not an exhaustive action list, and each FOT has its own special issues and concerns that have to

be dealt with on an individual basis. Nor is the advice in it necessarily perfect and representative of the state of the art. On many issues, there will be scope for disagreement with the recommendations or use of alternative sources of advice. But it would certainly be preferable for major departures from the advice to be justified to funding agencies and major stakeholders.

The FESTA project consortium decided early on in the project that the primary focus of the FESTA Handbook would be on the evaluation of Advanced Driver Assistance Systems (ADAS) and in-vehicle information systems for vehicles – both in the form of autonomous systems and of cooperative systems. It was also agreed that the FESTA Handbook should be relevant to the evaluation of Original Equipment Manufacturer (OEM), aftermarket and nomadic systems. The Handbook is therefore designed specifically to guide the evaluation of impacts of such systems, and is less relevant to the evaluation of electronic road infrastructure such as Variable Message Signs (VMS). However, it is seen that many of the activities identified in the Handbook are common to the evaluation of most vehicle- and infrastructure-based ICT technologies.

The FESTA project produced a set of deliverables on the main topics of the FESTA V (FESTA 2008a, b, c, d, e, f, g, h); the handbook brings the information from these deliverables together. In addition to the Handbook itself, there are more relevant documents available at: <https://connectedautomateddriving.eu/methodology/festa/>.

In conclusion, the FESTA Handbook gives an overview and general guidelines concerning the conduct of an FOT. FOTs are designed to contribute to the identification and verification of solutions to a problem, and this handbook is intended to provide a formalised and practical framework, not a cookbook; the methodology described will necessarily have to be adapted to the specific case, in order to increase the efficiency of the approach or to tackle data incompleteness or inconsistency. Furthermore, the results of an FOT may have to be integrated with external sources of information, to achieve a wider perspective and increased relevance for tackling the problem at hand.

1.3 Data Sharing

During the past 20 years, we have seen rapid growth in the number of FOTs performed worldwide. Their data has mostly been used to answer the research questions in the original project. As the number of different datasets has increased and thus the awareness of the substantial effort and funding needed to do these FOTs, the interest in data sharing has come ever more into focus worldwide. The advantages of re-using data from FOTs include the possibility to perform meta-analysis across FOTs and to answer new research questions without having to collect new data. In this handbook, attention is paid across several chapters to ways of ensuring that data can be shared and re-used.

FOT-Net has published a [Data Sharing Framework](#) as a complement to FESTA (FOT-Net Data and CARTRE, 2019). The framework contains a set of guidelines and best practices on data management and sharing.

1.4 Cooperative, Connected and Automated Mobility

Testing of Cooperative, Connected and Automated Mobility (CCAM) in all its different forms is now underway and a large variety of projects have been launched to investigate the effects in real life. Projects may vary between large-scale FOTs and small-scale trials and pilot projects with only a few vehicles. In this handbook, the term “FOT” will be used for reasons of conciseness. Although FESTA is mostly addressing large FOTs, where many vehicles drive for several months, the handbook also contains valuable elements for smaller pilots.

For small projects, an additional document is available with recommendations on how to use the FESTA approach on a small scale: [Micro-FESTA](#) (Barnard et al., 2021).

The term connected and automated vehicles may refer to several levels of automation, ranging from vehicles with systems that may take over some tasks from the driver to completely driverless vehicles. FESTA is strongly centred on the drivers of vehicles, and the changes in their behaviour when driving a vehicle that is instrumented with new systems. This focus is changing with CCAM FOTs towards the behaviour of the vehicle and the wider mobility behaviour of the user of these vehicles.

Two things should be noted here. Firstly, field operational testing is not the only way to evaluate automation; it is a method that is complementary to (or is preceded by) simulator and laboratory studies, expert assessments, technical tests on test-tracks, public acceptance studies, etc. Secondly, FESTA focuses on testing market-ready systems or advanced prototypes in real use, as opposed to covering earlier development steps. A piloting phase (preparatory testing) in FESTA, however, considers iterative steps to be taken to ensure good usability, efficient testing processes and safety of everyone involved. Detailed safety validation of a system and component-level testing precede actual FOT phases. Technical evaluation regarding final performance measurements in operational use, however, is often carried out as part of an FOT. Technical evaluation is to provide complementary insight into the systems as well as detailed input to impact assessment. More generally, FOTs are commonly supported by a set of additional controlled tests, either with a technical or human-oriented focus, that are to examine a specific aspect further, using a constructed setting and additional observation techniques.

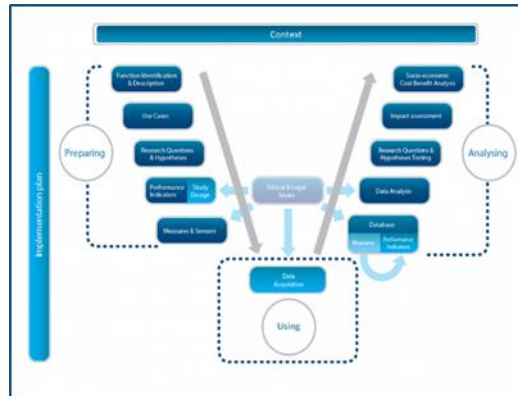
This version of the handbook is not a re-write of earlier versions. In every chapter, a section on automation has been added. Developing a full methodology for CCAM FOTs is a living target and lessons have to be learned from CCAM projects. That is why exchange of knowledge, experience and data will remain of the utmost importance to further our understanding of how FOTs can be conducted as effectively and efficiently as possible. In a new methodology project, starting in 2022, a Common Evaluation Methodology will be developed for CCAM projects.

2 Planning and Running a Field Operational Test

2.1 Introduction

For a Field Operational Test to proceed smoothly, a plan of action must be developed which documents the scientific, technical, administrative, and procedural activities that are needed to successfully complete it. Given that the lifecycle of an FOT typically evolves through many phases, there are many issues to consider. In this chapter, the critical activities and tasks which are necessary to run a successful FOT are documented – in the form of a “FOT Implementation Plan” (FOTIP) – drawing on lessons learned from previous FOTs conducted in different parts of the world.

The FOTIP is described in [Annex A](#) of the FESTA Handbook. In this chapter it is introduced, described, explained and discussed.



2.2 The FOT Implementation Plan

2.2.1 Purpose

The FOTIP is intended to serve primarily as a checklist for planning and running FOTs:

- To highlight the main activities and tasks that would normally be undertaken in successfully completing an FOT
- To ensure that, in running an FOT, researchers and support teams are aware of critical issues that influence the success of the FOT
- By drawing on the experiences of previous FOTs, to highlight the “dos” and “don’ts” of running an FOT
- To provide a consistent framework for planning, running and decommissioning FOTs.

The FOTIP presented in this Handbook is not intended to be prescriptive, but rather to serve as a generic guide in conducting FOTs. By their very nature FOTs are major projects – extensive and expensive. Previous FOTs that have not delivered their anticipated outcomes have not done so primarily because of failures to anticipate problems that compromised their successful execution. The FOTIP attempts to map out all known critical issues that need to be taken into account in planning and undertaking an FOT.

The history of FOTs suggests that no two will be the same, and that there often are many unforeseen tasks and sub-tasks that arise during its lifecycle. The list of tasks contained in the FOTIP in [Annex A](#) of this Handbook is not, therefore, exhaustive. It is based on the collective wisdom of those that have been involved in planning and running previous FOTs. There may be specific requirements for future FOTs conducted in Europe that will need to be decided on a case-by-case basis, especially where FOTs are extended to cover Naturalistic Driving Studies and, potentially, other transport modes such as public transport, cycling, walking or multi-modal studies.

The FOTIP in [Annex A](#) describes what needs to be done, and approximately when, in running a successful FOT. Other relevant chapters in the FESTA handbook describe in detail why these activities are necessary and how they are to be accomplished.

2.2.2 Description of the FOT Implementation Plan

The FOTIP in [Annex A](#) of this Handbook resembles a traditional Work Breakdown Structure (WBS), but without timelines. It is specifically designed in this way so that timelines can be inserted at a later date by those responsible for the overall planning and running of the FOT.

The FOTIP is divided into three columns and two sections below each activity:

- **Column 1 – Activities.** An activity is a high-level task, e.g. “Convene FOT research and support teams” that is usually needed to run an FOT.
- **Column 2 – Tasks and Sub-Tasks.** A task directly supports an activity, e.g. “Appoint FOT project manager”. A sub-task directly supports a task. Essentially, this column contains a series of action statements – “do this” etc. There are very few sub-tasks listed in this column, to contain the size of the document. The document is cross-referenced to other chapters of the FESTA Handbook, which identify the relevant sub-tasks that support these tasks.
- **Column 3 – Person/Organisation Responsible for Activity.** This column identifies the person, team, organisation or combination thereof that would usually be responsible for completion of a task. The FOT project manager is ultimately accountable for successful completion of all tasks, and is therefore included in every task.
- **After Section 1 – Critical Considerations (the “dos” and “don’ts”).** This column contains critical advice for ensuring that an activity or task is successfully completed. For example, “Be sure that the vehicle systems are designed so they do not drain the battery when the vehicle engine is not running,” or, “Do not underestimate the amount of time required to recruit company drivers for the FOT.”
- **After Section 2 – General Advice.** This column provides general advice on how to maximise the likelihood of running a successful FOT, e.g. “The FOT lifecycle is long. Hence, it is advisable to write separate reports on each critical stage of the FOT.” This column also contains explanatory notes, reference to other relevant documents (e.g. FOT reports) and cross-referencing to other chapters in the FESTA Handbook.

The activities and tasks identified in the FOTIP are consistent with those identified in the higher level “FOT Chain” described in [Chapter 1](#) of this Handbook (see Figure 1.1), although the chronological order in which the Activities and Tasks are shown varies slightly between the two. For example, in the FOT Chain, it is assumed that the first step when planning an FOT is the identification of systems and functions to be analysed. In the FOTIP, on the other hand, this task is identified later in the sequence of planning activities (within Activity 2), as there are other planning activities and tasks that necessarily precede the identification of systems and functions to be analysed. The FOTIP identifies the scientific, technical, administrative and procedural activities for planning and running an FOT. The FOT chain summarises the key, high level, scientific and technical steps undertaken when performing an FOT and the sequential links between them.

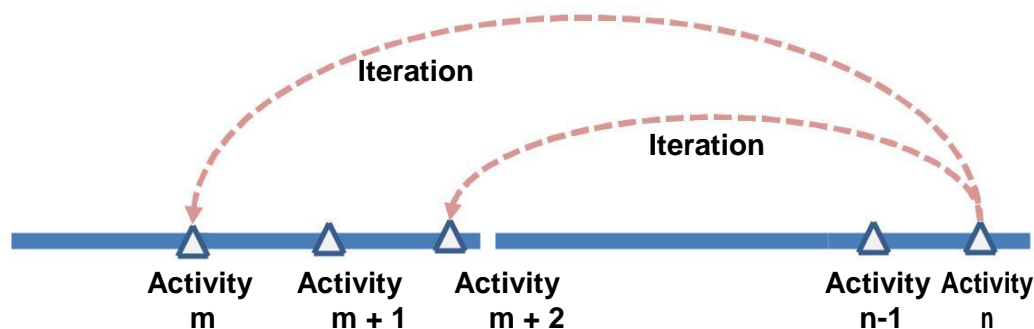


Figure 2.1 Iterative development process

2.2.3 Development of the FOT Implementation Plan

The content of the FOTIP derives originally from several activities undertaken in the FESTA project, based on a comprehensive review of the related literature and several consultations (via workshops, teleconferences, written feedbacks) with members of the consortium and external experts who previously conducted FOTs. It was then revised in rev 4 and rev 5 based on the feedback received in FOT-Net and FOT-Net 2. For version 6, FOT-Net Data added further details related to test data management.

2.2.4 Assumptions underlying the FOT Implementation Plan

There is more than one way to perform a successful FOT. The review of the literature on FOTs revealed that many different approaches have been taken in planning, running, analysing and decommissioning FOTs. The FOTIP in [Annex A](#) of this Handbook draws together procedural activities that are most common to the known FOTs that have been conducted, and the collective wisdom of those who conducted them.

The FOTIP is relevant to FOTs in which the ADAS and in-vehicle information systems to be evaluated are already available on the market, or to studies in which the systems to be evaluated must be chosen by the FOT project team, purchased or developed, and installed (e.g. as in Regan et al., 2006).

The FOTIP provides only a general guide to the sequence in which the required activities should be performed. Some need to happen early in the project and others at the end. Some need to immediately precede others. Other tasks need to proceed concurrently with others. Decisions about the scheduling of activities, tasks and sub-tasks are the responsibility of the FOT Project Manager. Table 2.1 lists the 22 activities identified in the FOTIP, and highlights the main dependencies that exist between them. Within activities, it is up to the FOT Project Manager to further decide which tasks and sub-tasks should proceed sequentially and in parallel.

Some of the major tasks listed in the FOTIP (e.g. “recruit participants” within the activity “Run FOT”) are given only a one-line description and, as such, may appear to be downplayed in the plan. A judgement had to be made about how much detail to include in the FOTIP. Where such one-liners exist, this is because either the task in question is one

that most researchers would normally be familiar with (e.g. recruiting study participants), or because the sub-tasks involved are described in detail in other chapters of the FESTA Handbook. Where appropriate, any known difficulties and concerns associated with major tasks for which only a one-line description is given are emphasised.

Table 2.1 A generic guide to scheduling the 22 activities described in the FOTIP in [Annex A](#) of the FESTA Handbook

Activity	Setup/ design		Preparation			Data Collec- tion		Completion		
Convene teams and people										
Define aims, objectives, research questions & hypotheses										
Develop project management plan										
Implement procedures and protocols for communicating with stakeholders										
Design the study										
Identify and resolve legal and ethical issues										
Select and obtain FOT test platforms (vehicles, mobile devices, roadside units, ...)										
Select and obtain systems and functions to be evaluated										
Select and obtain data collection and transfer systems										
Select and obtain support systems for FOT platforms										
Equip FOT test platforms with all systems										
Implement driver feedback and reporting systems										
Select/implement relational database for storing FOT data										
Test all systems against functional requirements and performance specifications										
Develop recruitment strategy and materials										
Develop driver training and briefing materials										
Pilot test FOT equipment, methods and procedures										
Run the FOT										
Analyse FOT data										
Write minutes and reports										
Disseminate the FOT findings										
Decommission the FOT										

2.2.5 Using the FOT Implementation Plan

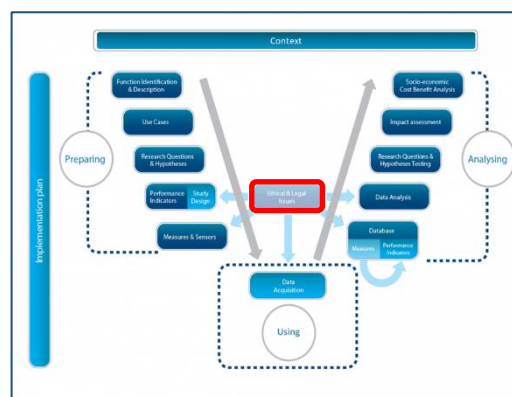
It is suggested that the FOTIP be used as follows:

- Read through the FOTIP before starting to plan an FOT
- Use the FOTIP as a checklist for guiding the planning, design and running of the FOT
- And as a quality control mechanism for ensuring during the study that nothing critical has been forgotten
- Read the FOTIP in conjunction with other chapters in the FESTA Handbook, and refer to other chapters and other FOT reports for detail
- If desired, use the FOTIP as the basis for the development of GANTT charts and other project management tools.

3 Legal and Ethical Issues

3.1 Introduction

Carrying out an FOT usually involves asking participants to share insight into their mobility behaviour or participating in testing a prototype vehicle. Where research involves human participants, there are numerous ethical and legal implications to consider.



3.2 Ethics

An FOT project should follow not only the legal rules of the jurisdiction in which the work is carried out, but the ethical principles which apply to research involving humans. The following provides a guide to the basic considerations which all FOT projects should consider (Vanclay et al, 2013), namely:

- Informed consent
- Specific permission for audio and video recording
- Voluntary participation
- Participants' right to withdraw
- Disclosure of funding sources
- Risk and safety assessment
- The minimisation of harm for all individuals involved
- Personal data collected only where directly relevant to the FOT
- Personal data protection
- Provision of grievance procedures
- Appropriateness of research methodology
- Responsible research innovation (the wider societal impacts of research).

In addition to research participants, the rights of non-participants affected by the FOT should be considered. These may, for example, be passengers in the participating vehicle or other road users captured on data recording devices or who come into contact with the FOT.

FOTs receiving public funding are commonly bound by ethical standards and guidelines named in grant agreements – for example, the European Code of Conduct for Research Integrity. Human rights legislation is relevant: the Universal Declaration of Human Rights, the European Charter of Fundamental Rights and the European Convention on Human Rights as well as the Helsinki declaration, although targeting medical testing, enshrines informed consent before participation in research.

3.3 Legal permissions and cross-border testing

It is not possible to provide a comprehensive guide to all the legal issues that can arise in a particular FOT, as these may be very dependent on the system(s) to be tested and on the study design adopted. It should be noted that the regulations and laws vary from country to country and that even where there are European laws and regulations the interpretation of these may vary between countries. Projects carrying out FOTs in more than one country or that potentially involve cross-border traffic will need to consider the legal implications and relevant permissions required in all relevant countries. In some cases, there may be different regional requirements where there are several jurisdictions within countries.

Consequently when undertaking an FOT, advice should be sought at a local and national level about the measures which should be taken, as well as the timelines and costs involved.

3.4 Participant recruitment

Participation in an FOT should be on a voluntary basis, and based on sufficient information for the individual to make an informed choice about the consequences of participation, and the purpose of the research. Potential risks which the participant may encounter should be set out, as well as what personal data will be collected and how that data should be used and stored. Participants' right to withdraw from the research should be communicated. These matters should be set out in a participant information sheet and provided to the participant prior to the FOT, and their agreement to participate should be confirmed. Please refer to [Annex B](#) for a checklist relating to Informed Consent for participation in an FOT.

It is important to clarify the roles, responsibilities and liabilities to the test subjects. If the testing involves complex responsibilities, an introductory session should be organised for the participants, so that responsibilities in different situations can be clarified. Such introductory sessions should help to avoid, for example, situations where the test subject has not truly read or understood a lengthy participant agreement that potentially includes difficult legal text. Without any effort to clarify the responsibilities, a test subject could later try to question the validity of the agreement.

[Annex C](#) presents a checklist regarding such introductory sessions.

In some countries, it may be a requirement for the participants to undergo a medical examination to prove their capability to take part. In any case, it would probably be sensible to ascertain if they have any medical conditions that might affect their ability to participate. However, in many countries data privacy requirements will not allow enquiry into medical details; this then could be solved by including in participant agreements medical reasons for ending the participation in the FOT.

3.5 Participant agreement

The type of participant agreement required will depend upon the design of the FOT. Where the participant is involved in an FOT in a controlled testing environment (not on a public

road), then the most suitable form of agreement may be implemented by providing an information sheet and a consent form (as detailed in [Annex B](#)).

However, where a participant is provided with a vehicle which is to be used on public roads, a more detailed form of agreement will have to be devised with the assistance of legal experts within the jurisdiction the FOT is to take place.

Issues potentially relevant between the handling organisation and the test participant should be regulated in a contract to provide for legal certainty on both sides (e.g. on obligations, liabilities, insurance issues, information on logging of personal data requiring informed consent, which parties will use the data, data sharing after the project). In this respect, the legal validity of the arrangements must be considered. This is an issue touching the general rule of freedom of contract as a limiting exception. By way of illustration, the handling FOT organisation cannot prompt participants to agree (in a legally effective way) to fully exclude liability. A lawyer can provide advice on this and should definitely be consulted. Further issues of relevance to be considered and covered in this agreement are what shall happen in the event that a participant commits a traffic offence and/or incurs a traffic penalty (speeding ticket, parking ticket, etc.). Another is who is responsible for minor damages to the vehicle and payment of any insurance excess. The issue of who is allowed to drive, e.g. other household members, and under what circumstances, also needs to be considered. Only the participants will have been properly informed about their responsibilities. There is no way to ensure that any third parties are properly briefed.

If personal data is going to be shared across borders, during and/or after the project, it is of utmost importance to set up a process ensuring that the participant agreements used in different countries all include the necessary statements for the data to be used and re-used internationally. A suggested process would be to design a project template that can be adapted to different national legislations where data is being collected and altered. All nationally adapted participant agreements are then gathered centrally, and the statements are checked to ensure that the data can still be shared across borders. Any deviations from the template that imply that data cannot be shared as intended need to be addressed. Solutions might potentially need to be decided by the project management, as deviations might affect the outcome of the project.

Personal data may only be shared from an EU country to a non-EU country under certain conditions. In the example of an EU country sharing data with the UK, the European Commission has determined that the UK offers an adequate level of data protection, and personal data can flow from the EU to the UK.

[Annex B](#) gives a brief checklist for test organisers to consider, when drafting a participant agreement or information sheet.

3.6 Protection of personal data

Personal data should only be collected and processed if it is necessary to achieve a specific objective associated with the FOT. If the aim can be achieved without personal data, then it should be achieved that way. As a general rule, the less personal data which is collected and processed the better.

If it is likely that the type of data being collected will result in a high risk to individuals, then a data protection impact assessment should be carried out. Advice as to how to carry out this assessment should be sought from the relevant Information Commissioner or Data Protection Authority for the jurisdiction in which the FOT is to take place. A high risk is dependent upon the nature, scope, context and purpose of why the data was collected/processed. It may involve the potential for individuals to be discriminated against due to the data collected, where the data belongs to children or otherwise where any harm could be inflicted on the data subject.

Protection of personal data in the EU is stipulated by General Data Protection Regulation (GDPR) from 2018. National laws may state specific additional requirements. Where an FOT is performed in several countries, the legal implications for data protection should be thoroughly investigated.

The general approach to data acquisition in the majority of FOTs will be to ask the participant for his/her informed, voluntary consent. The information provided thereby describes the scope of processing intended over the lifetime of the data, including potential re-use of data after the project. No additional processing (e.g. disclosure of the data that might lead to identification of the persons involved) can normally take place without prior consent.

An issue of major relevance in many FOTs has been processing of video data from cameras directed to the vehicle surroundings. There are two legal privacy issues associated with such cameras: the recognisability of other road users and license plates. Both are critical in respect of data privacy regulations in place in several EU Member States. Video data must be treated with special care, therefore, it is important that personnel handling and analysing the data are given appropriate education on data protection procedures.

With respect to video (and also audio) recording, in the event an FOT takes place in a controlled test environment, where the participants and passengers are known to project management, consent to have video and audio recording should be dealt with in the participation agreement.

However, where the vehicle is provided to a participant for use in their own time, it is possible that non-participants will occupy the vehicle in the situation where it is not possible to obtain explicit consent for their voice and image to be recorded. The options in this case are:

- The vehicle is supplied to the participant on the condition that they retain sole use of the vehicle and that persons who are not part of the FOT are not permitted to use the vehicle
- The participant is permitted to use the vehicle as they wish, however, limitations must be placed on video and audio recordings inside of the vehicle, to account for the presence of a person who has not been able to provide explicit consent for the data to be processed as part of the FOT

- The inside of the vehicle should display a notice that indicates it is a vehicle being used for research purposes, recording devices are in operation, and include contact information.

Technical measures must be employed to protect personal data, such as data server and database arrangements. All data transfers including personal data should be made using encryption. Any personal ID information should be kept separate from the main database and stored with additional protection such as encryption. It has to be recognised that even when data has been pseudonymised, it may be possible to deduce who participated from other sources, e.g. from GIS (Geographical Information System) data in the database.

Where video recording takes place, anonymisation techniques such as facial blurring should be employed to protect personal data as much as possible. For automated anonymisation of video data, new tools are being developed that can be applied to huge datasets. One promising technique is to cover the driver's face with a mask which shows the same emotions as the driver. If datasets can be anonymised effectively, the need for data protection when sharing them would be reduced and more datasets would become available.

Academic and research exceptions to GDPR

There are circumstances where the obligations for GDPR for researchers may be relaxed, where data is processed in the public interest, for scientific or historical purposes and the practical difficulty of complying with all obligations renders the research impossible or nearly impossible to complete.

These derogations are not well understood, and as a result, they are under-utilised. Member states are empowered to make their own law with regards to safeguards and derogations to GDPR, however in general a relaxation of some of the obligations may apply where:

- Complying with the obligation would prevent or seriously impair the likelihood of successfully achieving the research purposes (if the purposes are for public research or research in the public interest).
- The processing is not likely to damage or distress an individual.
- The project continues to employ organisational and technical measures to protect the data.
- The project carefully documents the decision not to comply, specifically explaining the circumstances and reasons.

The data protection authority where the research organisation is based, will be able to provide specific guidance on derogations relevant to that jurisdiction.

3.7 Proprietary data and stakeholder confidentiality

Data ownership and data sharing relate to stakeholder interests. Some stakeholders will regard data as strategic or sensitive. To deal with these stakeholder interests,

agreements on how to address these issues should be proposed as far as possible in advance. There are several ways to address them, such as:

- Project agreements, such as grant agreement, consortium agreement and agreements with external data providers, is recommended to include how to deal with data ownership and re-use as such. A checklist is available in [Data Sharing Framework](#) (FOT-Net Data and CARTRE, 2019).
- Procedures on how to change or introduce new research topics or projects using the collected data, which should assure the stakeholders that the data use and data protection are according to their requirements. This also includes procedures for potential re-use of the data by third parties.
- Anonymisation or de-identification of the data.
- Address ownership of data in the tendering procedures or contracts with the (public) organisation providing the grant.

The confidentiality aspect will affect detailed functional and performance analyses of the system and the publicity of results. It also affects raw log data sharing between consortium partners and required non-disclosure agreements. Generally, there is a confidential detailed level of data, and then there can be levels that are more public, post-processed or aggregated (e.g. indicators related to usage of a system), reaching up to open data.

Data collected from a vehicle bus, an internal communications network, represents a special case. Some of the data may reveal information that is confidential to the manufacturer, who may not want to share the data with third parties. Proper data filtering could be implemented in order to make available to the relevant partners only the data that is necessary to the FOT analysis.

An overview and recommendations addressing the above-mentioned issues, also from a data sharing and re-use perspective, can be found in the [Data Sharing Framework](#) (FOT-Net Data and CARTRE, 2019).

There are some additional points to consider regarding video recording and stakeholder confidentiality. For example, there may be locations encountered where it is illegal or prohibited to video externally, such as border crossings, military locations and private premises. The possibility of this happening needs to be considered; it is likely to be more of a problem in truck FOTs. Data collection could include geofenced areas, where logging is enabled or disabled.

3.8 Risk assessment and safety

A project needs a comprehensive risk assessment plan and will need to be able to demonstrate subsequently that the identified hazards have been properly managed. Organisations will normally have a safety management process for this.

It is incumbent on those conducting an FOT to ensure that the equipment that they have installed in a vehicle and the modifications that have been made to the vehicle systems are safe. Hazards can arise from radio and electrical interference (where electromagnetic compatibility tests should be conducted), from reducing vehicle crashworthiness (installations on the dashboard, interference with airbag deployment, and so on) and from HMI (Human Machine Interface) designs that cause distraction. The potential for failures

to arise from modifications to and interaction with in-vehicle systems needs to be handled by means of a formal system safety assessment. Further, testers need to be made aware of modifications altering the driving experience or requiring their attention, so minimise dangerous surprises.

The rise of advanced driver assistance systems and automated vehicle prototype testing bring further requirements regarding safe design, preliminary safety testing and comprehensive analysis of risks. Such FOT projects have used numerous steps and methods to ensure safety. As a general recommendation, a risk assessment method should be selected to review and mitigate risks from FOT perspective. Unacceptable risks must not remain after mitigation steps.

Detailed safety validation of a system is here considered to be a development step that has already taken place before field trials are to begin. Still, certain systems require tailoring according to the operational area and repetition of key assessment steps.

3.9 Insurance

Insurance requirements extend beyond the insurance of the vehicles and possibly of the participants. There is also a need for indemnity insurance to cover the FOT as a whole. This may be provided by an employing organisation's professional indemnity insurance, but it is vital to confirm that large risks are covered. Insurances and liabilities tend to differ greatly between the EU Member States. It is therefore advisable to seek legal advice in this highly risk-sensitive aspect.

In recruitment, it is essential to ensure that participants have legal entitlement to drive the vehicles in question and are eligible for insurance. It may be wise to have insurance coverage for the fleet as a whole. If the participants are to drive their own vehicles or vehicles that belong to a fleet not under the control of the handling organisation, then insurance coverage needs to be confirmed. Coverage when travelling to other countries may be relevant.

3.10 Responsibilities and legal liability

In the event of an accident, injury or damage, liability and responsibility will be determined by the following factors:

- The legal jurisdiction in which the FOT occurred and any applicable national or regional law or regulation
- Conditions relating to safety, attached to the research funding
- Contractual arrangements between the parties, for example, contractual clauses relating to liability and responsibility between component providers, manufacturers, and software providers
- The nature of the incident and whether it was caused by negligence or by criminal activity (such as dangerous driving)
- The type of damage (i.e. property damage or injury to a person)
- The foreseeability of such an incident and its outcomes, and whether steps could have been taken to prevent it.

Generally, parties are responsible for events or outcomes which had been within their control. For example, if an accident is caused by a faulty component in an automated vehicle, then the manufacturer of that component is likely to be held responsible and liable. However, an automated vehicle FOT may involve multiple parties (manufacturer, software provider, infrastructure provider, driver, project manager) and consequently, there may be multiple parties involved in a single incident.

3.11 Ethical approval

In many countries and in many organisations, there are strict ethical approval and human subject review procedures. These processes may be time-consuming and require thorough documented plans for review. The categories of tests and types of data collection that require detailed ethical considerations in different countries and organisations, are usually available. Cases with clear danger to participants or uncommon handling of personal data will always call for ethical reviews and detailed mitigation of risks. An ethics screening can help to clarify, if further assessment will be necessary.

Different funding instruments generally refer to their preferred ethical guidelines. For EU-funded research projects, European Code of Conduct for Research Integrity serves as the reference document. Human rights legislation is also relevant, as already discussed in [Section 3.1](#).

If the data is going to be re-used after the project, it is important that the research organisation re-using the data investigates whether an ethical approval is needed for the new research project, according to the national legislation in the country where the research project is performed.

In terms of the project timeline, legal and ethical issues need to be considered in parallel during the whole project (and indeed afterwards in terms of data protection, especially if the data is planned to be re-used in new projects, see the [Data Sharing Framework](#) (FOT-Net Data and CARTRE, 2019). It is especially important to take legal and ethical topics into account when drafting the different project agreements between the consortium partners, also bearing in mind data sharing implications. Taking data sharing into account early on clarifies project practices and reduces the need for subsequent bilateral agreements.

As these issues are complex and it is not always easy to find a good solution, it may be useful to consult (next to legal experts) other, comparable FOTs that dealt successfully with them and learn from their experiences.

The FOTIP, discussed in [Chapter 2](#) and presented in table form in the [Annex A](#), provides information about when in the FOT process the various legal and ethical issues need to be considered. The project plan needs to clearly identify the persons responsible for ensuring compliance or involving legal expertise in the respective Member State.

3.13 Iteration

Considering ethical and legal issues may influence the experimental design, data collection and outcomes of the different phases of the FOT chain. It may be necessary to

re-think some phases and to abandon choices made earlier. For example, if it is not possible to collect certain data due to legal or ethical issues, it may no longer be possible to test certain hypotheses or to use certain performance indicators.

3.14 Specific considerations for CCAM FOTs

In this chapter, the need was stressed to start timely preparations for getting legal and ethical approval for FOTs, and to involve legal experts from the very beginning. For projects dealing with automated driving or similar prototypes, this is even more important.

For vehicles driving in automated mode on public roads, it is necessary to obtain permits from the national authorities. Involving the general public in user tests will generally, for safety reasons, require a trained safety or professional driver. This driver may perform a part of the driving tasks in case of lower levels of automation, but may also be present in the vehicle only to take care of emergencies. The safety driver's responsibilities should be clearly documented beforehand, to focus them (e.g. to monitor children at the roadside) and consider that the tasks are reasonable for a human observer.

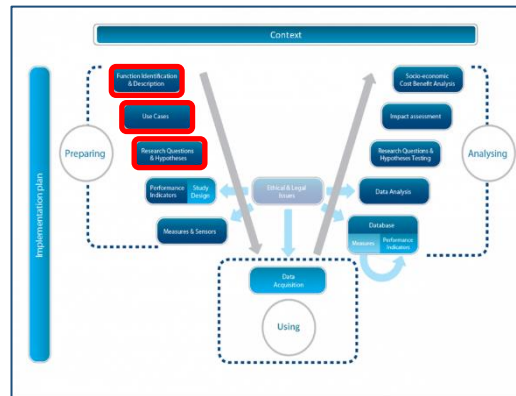
Because it may be difficult to gain authorisation to drive on public roads, automated driving FOTs may choose to use test tracks or closed environments. However, the safety of the participants will remain a key issue in test design and ethical approvals.

When compared to earlier FOTs, a new ethical aspect specifically raised by automated driving tests seems to be the mathematical interpretation of traffic rules and goals related to traffic safety, now that they become programmable. Many companies are developing different safety approaches, considering e.g. safe driving style near pedestrians. However, the actual goals can be difficult to state. For example, very careful driving near persons who have obvious difficulties coping safely in traffic, such as children and people with disabilities, is a common legal requirement. How to translate such legislation about being careful to minimum safety margins to keep in different situations, is far from clear. Further, selecting ultra-cautious driving style reduces the chance of accidents, but slow-moving vehicles may not be well accepted by customers who might look for high efficiency.

4 From Functions to Hypotheses

4.1 Introduction

The final objective of an FOT is to evaluate in-vehicle functions based on Information Communication Technology (ICT) in order to address specific research questions. These research questions are commonly related to safety, environment, mobility, traffic efficiency, usage and acceptance. Specific FOTs can raise additional topics, such as automation, FOTs might study land use. Defining and prioritising research questions, according to resources in use, is a key phase in the beginning of an FOT.



By addressing the research questions, FOTs promise to furnish the major stakeholders (customers, public authorities, OEMs, suppliers, and the scientific community) with valuable information able to improve their policy-making and market strategies. Individuating the most relevant functions and connected hypotheses to successfully address the above-mentioned research questions is one of the major challenges in an FOT. In this Chapter, the process of individuating the functions to be tested in an FOT and the relevant connected hypotheses will be elucidated. Specifically, the reader will be guided in the process of

1. Selecting the functions or features to be tested
2. Defining the connected use cases to test these functions
3. Identifying the research questions related to these use cases
4. Formulating the hypotheses associated with these research questions, and
5. Linking these hypotheses to the corresponding performance indicators.

The steps may be influenced by other elements of the FOT chain. The selection of functions may be driven by the expected socio-economic impacts, or by the research questions. When details are filled in later on in the process, it may be necessary to re-visit earlier steps. Limits in resources or technical capabilities may lead to a decision to limit the amount of functions, use cases and hypotheses. While multi-million FOTs generally cater a large selection of research questions with their data collection, smaller tests must commonly select fewer focus areas.

4.2 Applying this process to Naturalistic Driving Studies

This chapter is written from the perspective of FOTs on various kinds of systems. By contrast, NDS investigate driving itself without the constraints of the experimental conditions that are normally required in FOTs. In FOTs there is a natural progression that starts with a specification of the functions to be evaluated, then moves on to the environments and situations in which the tests will be conducted and the experimental design that will be employed. The specified research questions follow in a natural progression, and, in forming the hypotheses, there is generally an anticipation of the

direction of change, e.g. that a driver assistance system will promote safety or fuel economy. In framing hypotheses for FOTs, we seek to gain a deeper understanding not only of what changes occur but of how they occur, i.e. of how a particular function influences user behaviour.

By contrast, NDS are much more exploratory. They aim for deeper understanding of driving behaviour as it relates to safety and in some cases environmental aspects of driving. They can thus be likened to other observational studies in traffic. Of course, there is a kind of experimental design even in NDS: the vehicles and participants have to be selected, and the choices made here will be determined by the focus of the study and the research topics being addressed.

NDS focus on how drivers manage their driving in changing situations and environments and on how breakdowns in the safe operation of vehicles come about. Hence, they are primarily interested in why problems occur and do not occur. This makes research questions, as opposed to hypotheses, the natural focus of NDS. Those research questions will tend initially to be quite broad – what is the relationship between distraction and safety-critical events or why do young drivers have a problem with loss-of-control on curves at night? The RQs can then be further specified in the form of sub-RQs and so on. From those sub-RQs, appropriate performance indicators and measures can be defined. This then leads to the specification of the data acquisition system and associated subjective and objective data.

One of the challenges in NDS is prioritizing the research questions. It is comparatively easy to generate hundreds of RQs and sub-RQs. The difficulty then is likely to be setting the boundaries of the study by determining which ones are high priority and can be addressed within the resources available. However, if the final database is sufficiently broad and flexible, it can become a valuable resource for the investigation of additional topics and RQs. NDS data will be useful for many other studies on driver behaviour and therefore it is of the utmost importance that facilitating data sharing is taken into account from the very beginning of such a study.

More on NDS can be found in [Section 6.6](#).

4.3 Systems and functions

In the last few years, the number of ICT functions available for use on standard vehicles and more generally while travelling has been rapidly increasing. ICT functions are intrinsically designed to provide the driver or traveller with new, additional information. However, the extent to which this increased amount of information from these ICT functions results has positive effects on safety, environment, mobility, usage, and acceptance in a real traffic situation, is unknown. FOTs warrant to evaluate, for the first time, these ICT functions in a real traffic situation during naturalistic driving. In this Handbook we refer to 1) in-vehicle, 2) cooperative, and 3) nomadic systems intended as a combination of hardware and software enabling one or more ICT functions. Depending on the different systems implementing a specific function, different challenges may have to be faced during the FOT design.

It is important to note that NDS in the future will increasingly include new systems as they increase in market penetration. For example, future NDS will be able to be used for with-and-without analyses of systems that are new today, but commonplace tomorrow (e.g. Forward Collision Warning may become as commonplace as ABS).

4.4 Vehicle systems

Vehicle systems are a combination of hardware and software enabling one or more functions aimed at increasing driver safety and comfort. Vehicle systems promise

1. To increase road safety by increasing driver attention in potentially hazardous scenarios (such as the Forward Collision Warning function),
2. To improve driver comfort by automating some of the operational driving tasks (such as the Adaptive Cruise Control function),
3. To increase driver mobility by furnishing timely traffic information (such as the Dynamic Navigation function), and
4. To increase safety in a critical situation by automating the vehicle response (such as the Collision Mitigation function).

These types of systems are becoming increasingly standard equipment. However, their impact on the driver, the traffic system, the society, and the environment in the short- but especially in the long term is not fully understood. FOTs can help quantify the impact of vehicle systems on driver workload and understand how different functions interact with each other in a real complex traffic situation. Further, FOTs will expose these functions to many improbable scenarios which can be very difficult to consider or test during the functions development and validation phases.

4.5 Nomadic devices

The use of Nomadic Devices for transport and traffic related applications has become increasingly commonplace in the last few years. The first wave of such devices was dedicated satnavs, also known as Personal Navigation Devices (PNDs). The second wave of functions has been a large variety of “apps” for smartphones. In addition, some PND suppliers have formed alliances with vehicle manufacturers so that their products may be fitted as original equipment. Whether fitted as original equipment or functioning on smartphones, the hardware platforms tend to operate autonomously of the vehicle. Indeed, many of these devices can provide traveller functionality – for pedestrian route finding, for public transport information, to locate points of interest – outside the vehicle.

The functionality of PNDs and smartphones has been evolving over time. These days, both feature map updates and cooperative-system-like functionality in the form of live traffic updates and other environmental information.

Specifically tailored aftermarket devices, sometimes offered by the same manufacturers as personal PNDs, are now targeted at the fleet market for management of logistics and other fleet functions. Some of these devices provide links to CAN data. An increasing focus of such systems is fuel economy and feedback to drivers on their efficiency of driving. Another sector for aftermarket devices is the pay-as-you-drive insurance market.

Nomadic devices need to be evaluated from the perspectives of user behaviour and acceptance, safety (particularly in regard to HMI issues), travel and traffic impacts and environmental implications. It should also be recognised that such devices can have broad mobility implications, both in terms of the strategic level of driving (route choice) and in terms of trip generation and mode choice. Any evaluation of usage needs to consider the potential for both in-vehicle and out-of-vehicle usage of these devices.

One of the critical characteristics of PNDs and smartphones is how far the device is integrated within the vehicle. Many devices use specific mounting kits for in-vehicle installation (connection to power supply, GPS and in some cases the vehicle's audio system). Typically, PNDs are mounted with a suction cup directly to the windscreen, while cradles for smartphones may also attach via a suction cup. As a result, they may impede the driver's field of view increasing the risk of accidents. They may break free in case of a collision. There are also often problems with small screen size, visibility and inadequate audio volume. Nevertheless, the popularity of these devices has not been affected. To increase usability and reduce negative side effects, automotive connectivity (e.g. Bluetooth profiles) has been developed so that such devices can use or be used with the vehicle's built-in HMI.

The general ease of use of a device will have a major influence on acceptance and willingness to pay. Here, ease of use refers not just to the usability while driving but to the user experience in all aspects of usage – pre-trip, in-trip and post-trip. Post-trip functionality is very relevant to usage in the fleet market and to support for and feedback on eco-driving.

Due to the fast innovation cycle, FOTs studying nomadic devices may require state-of-the-art planning in order to keep up with the introduction of new features and functions. Updates or changes to the tested system need to be carefully considered and controlled, so that result analysis remains clear. Nomadic device FOTs will also need to consider the surrounding infrastructure since, rather like cooperative systems, many functions rely on information and support from the outside. Weather forecasts, traffic information, updates on road conditions, dynamic speed limit information and speed advice are all dependent on service providers.

4.6 Cooperative systems

Cooperative Systems are vehicle systems based on vehicle-to-vehicle (V2V), vehicle-to-infrastructure (V2I) and infrastructure-to-vehicle (I2V) communication technology. Communication technology has enabled a new class of in-vehicle information and warnings which are precise in terms of time and location. It is already seen that the integration of cooperative systems with automated functions can provide a new level of ADAS. Infrastructure-based information tells the vehicle cruise control, for example, what is the appropriate speed to keep on a specific part of road or warns the driver in case of ice on the road.

Cooperative systems can differ from in-vehicle systems in several areas, which directly influences the FOT planning and operation. The main issue lies in the essence of cooperation: all V2V functions rely on more than one vehicle being in communication range, and V2I relies on roadside stations. Before starting the FOT, it is hard to estimate

the number of vehicles in one area and how often vehicles will pass a Road-Side Unit (RSU). This *penetration rate* is a crucial factor in designing and evaluating an FOT.

Estimates from the first V2X FOTs predicted a minimal penetration rate of 10% for V2X functions to show a noticeable effect on traffic safety and efficiency. Given the size of the FOT area and the average distribution of vehicles, the number of participating vehicles should usually be considerably higher.

The penetration rate also influences the Frequency of Events (FoE). In cooperative systems FOTs, the FoE is roughly related to the number of necessary vehicles for a function. While some functions work with only two vehicles (e.g. a slow vehicle warning), other functions require several more (e.g. traffic jam ahead warning). For certain functions the combined frequency of events might make a naturalistic FOT unfeasible. At least, it becomes necessary to focus test subject recruitments and installations to a small area.

A proper assessment of the penetration rate impact can be derived from dedicated simulations. A simulation environment should consider traffic effects, communication models, applications and their respective influence on each other. Such dedicated simulation environments are able to predict the frequency of events for the developed functions and thus support the FOT design and setup process greatly in cooperative systems.

There are several implications derived from this effect for the test setup and execution of cooperative systems FOTs:

- The number of equipped vehicles is considerably higher or the FOT needs to run for considerably longer to collect enough data to be representative.
- It might be very difficult – even impossible – to run the FOT uncontrolled for some functions (e.g. Emergency Electronic Brake Light). Controlled testing implies the usage of dedicated tools to specify the test cases, the test scenarios and to run the test. They also need to be linked to the vehicles and drivers to control and monitor the running test. See [Section 6.5](#) for further detail.

FOTs can assess the technological and business feasibility of cooperative systems and may be necessary to complete the final validation of such systems. In fact, testing to validate cooperative systems may require a very complicated setup, which may not be possible unless in a real-traffic setup.

One important issue is the installation of infrastructure devices that could be needed. Permission for the testing and team involvement should be planned in advance considering the impact on normal traffic during the installation phase.

Assessment of RSUs and antenna positioning could be rather time consuming. Protection and maintenance of roadside equipment should be ensured.

A specific complexity of cooperative systems is due to the distributed nature of involved components. It is not sufficient only to log the vehicle state and the function state, but in order to evaluate completely the functionality of a cooperative system it is also necessary

to log incoming and outgoing V2X data and the Local Dynamic Map state on the vehicle side.

These data types are provided to or received from other nodes of the V2X network and directly influence the system functionality. Consequently, the scope of logging has to be widened; each node (ITS station) needs to be equipped with a logging device. This device or software component has to be able to log all important internal sources and the network:

- Vehicle Data (CAN Bus)
- Position and Time (if not present on CAN)
- Networking: Incoming and Outgoing Messages (CAM, DENM, etc.)
- Local Dynamic Map Status (lists of neighbour stations and events)
- Function Status (to monitor how applications are working)
- HMI (interaction with the driver)
- Video
- Other Facilities (connection to backend services, etc.)

Roadside stations provide similar measures to those in vehicles (except vehicle data, HMI and maybe video) and it is advised to use a similar architecture for the logging system. In contrast, the measures available to an ITS central station (e.g. test control) can differ widely and are specific to each FOT. Additional roadside sensors may be needed to provide information related to interaction with non-connected vehicles and other situational variables.

For controlled testing scenarios it is advisable to include monitoring features in the logging. Monitoring is about real-time transfer of selected information available to the logger. In a central counterpart, this information is displayed to visualise the current test progress. The monitoring value set is a subset of logging and should at least include vehicle ID, position, heading and speed. It may include information from tested functions to monitor in real time, if the test goals (e.g. to trigger the functions in a specified way) have been met or if a test run needs to be repeated.

When large international FOTs are planned, cooperative systems may require interoperability tests. The costs and organisational and legal aspects (prototypes in foreign countries) should not be underestimated when planning these testing sessions.

4.7 Combinations of functions

There are many FOTs that investigate the impacts of a combination of functions – sometimes because systems and functions come in a bundle. One such common bundle is the combination of Adaptive Cruise Control (ACC) and Forward Collision Warning (FCW). Both functions make use of the same sensors, and indeed second-generation ACCs generally implement a warning function to indicate to the driver when the deceleration demanded of the ACC in order to prevent a collision with the preceding vehicle is beyond the function's designed capability. In other cases, an FOT may be investigating a function that resides on a platform which offers many other functions. This is almost invariably the case when studying functions that reside on a nomadic device such as a smartphone or PND. In some cases, a project will create a new function and provide

it to users on a standard consumer nomadic device. It is not practical or reasonable to demand of users that they do not use the full functionality of the device, and attempts to disable features may well annoy participants.

In planning the evaluation, it is important to consider how functions may interact with each other and how those interactions might affect user behaviour. This needs to be done at the stage of an FOT when research questions and hypotheses are initially formulated. What needs to be considered is:

- Can the effects of the various functions be disentangled? Note that it may not be possible or feasible to do so, particularly if the functions are closely coupled together.
- Does the experimental design need to be modified to enable both the single effects of each function to be investigated as well as the effect of the functions in combination?

Some systems are now so integrated that it is no longer feasible or even safe to disentangle them completely.

Where bundles cannot be disentangled, it is only possible to investigate and report on the impacts of the combined systems. But when, for example, there are separate apps on a nomadic device, then consideration needs to be given to how those apps might interact. The investigation can be carried out in a naturalistic manner, i.e. the participants are free to choose when and when not to use the applications and functions that are not the immediate focus of the FOT. Alternatively, it can be carried out in a more experimental manner by means of instructions to participants, systematically enabling and disabling the functions, or carrying out controlled drives with the functionality of the system(s) carefully manipulated.

In formulating hypotheses, it is useful to think both of the singular effects of a system or function and of the synergistic effects that one function may have on another. Thus, the recommended procedure is to start with the individual functions and then to proceed to combinatory effects. The process should therefore be to:

1. Begin separately with the individual functions, generating a list of research questions and hypotheses
2. Examine commonalities and conflicts between the systems and functions and derive hypotheses from those commonalities and conflicts:
 - a. Can they generate simultaneous messages and/or warnings?
 - b. Do they have a common interface and can they both be activated simultaneously?
 - c. Are the same performance indicators relevant to each system and/or function?
 - d. Are there common factors influencing usage?
3. Distinguish between hypothesis additive effects when the two systems interact with each other and multiplicative effects when the presence of a second system will alter the effects of the first. Additive (or subtractive) effects mean that the size of the effects will change. Multiplicative effects mean that the relationship is different, i.e. that there is an *interaction* in statistical terms. Consider situations in which such multiplicative effects might be important.

The application of this procedure should produce a comprehensive set of hypotheses on how the functions might interact and should affect the subsequent experimental design. As in all such work on the preparation of research questions and hypotheses, the reasonableness of pursuing every possible combination in a structured experimental design needs to be considered. Any additional function can impose a huge cost in terms of the increase in the number of possible combinations. It can be argued that there is a rationale also for multiple baselines, i.e. with all functions off, and with one function at a time off.

Complex experimental designs have large practical costs associated with them, and the benefits of such designs need to be carefully considered. The costs can be in the form of the number of different baselines that may be required and of the time needed for data collection on each combination. A full experimental design in which ordering effects are considered may well be totally impossible for practical reasons. This can all lead to excessive time required for FOT execution. But another side effect can be the sheer difficulty of getting the participants to comply with all the different conditions of the experimental design. These are arguments for using a more naturalistic approach in which the participants are free to use whatever combination of systems and functions they choose. Of course, system and function state will need to be recorded. This naturalistic approach has some disadvantages:

- Not all combinations may occur and not all participants may experience each combination
- It may be hard to take care of seasonal effects
- There may be insufficient sample sizes in some conditions so that experimental power is inadequate

However, the naturalistic approach also has advantages in that:

- Participant compliance will generally be assured
- The frequency with which the various combinations are used can provide useful information for the scaling-up process.

Alternatives which offer greater efficiency than a full experimental design can be proposed. Laboratory experiments on driving simulators can be adopted to examine synergistic effects and *a priori* analysis can be applied at an early stage in an FOT to identify combinations that are of particular interest and which should therefore be the focus of attention. This can lead to a satisfactory but incomplete experimental design.

4.8 General methodology

The main advantage of an FOT is that it has the potential to give insight into system performance in naturalistic driving situations, as free as possible from any artefact resulting from noticeable measurement equipment or observers in the car. Therefore, the first step when planning an FOT is to identify systems and functions where considerable knowledge about their impacts and effects in realistic (driving) situations is of major interest, but is still

lacking (see [Section 4.8.1](#)). Another domain for FOTs is the area of systems and functions which need a certain penetration rate to work at all, like an especially cooperative system.

After identification of the functions and system, which should be tested in an FOT, the goal is to define statistically testable hypotheses and find measurable indicators to test them. To reach this goal, several steps need to be taken, starting from a description of the functions down to an adequate level of detail (see [Section 4.8.1](#)). This means that the main aspects of the functions], the intended benefits and the intrinsic limitations have to be described to fully understand objectives and limitations and to derive reasonable use cases.

Secondly, these use cases need to be defined (see [Section 4.8.2](#)). Use cases are a means to describe the boundary conditions under which a function is intended to be analysed. A general starting point is given by the functional specifications from the function description part. But it might also be of interest how a function performs when certain preconditions are not met and to identify unintended and unforeseen effects.

Starting from the use cases, definitions-specific research questions need to be identified (see [Section 4.8.3](#)). Research questions are general questions to be answered by compiling and testing related specific hypotheses. While research questions are phrased as real questions ending with a question mark, hypotheses are statements which can either be true or false. This will be tested by statistical means (see [Chapter 9](#)). One might already have a very clear idea from the beginning which hypotheses are to be tested in a very specific situation during the FOT. However, this very focused view might result in an extremely limited experimental design where important unintended effects should be considered. The process for defining hypotheses developed in FESTA aims to prevent these potential issues (see [Section 4.8.4](#)).

Finally, hypotheses can only be tested by means of reasonable indicators (see [Section 4.8.5](#)).

In the FESTA methodology, function description is a starting point. However, this may pose some problems. Not all FOTs are testing one pre-defined function; sometimes a set of functions or systems are to be evaluated. Or, in NDS, the focus of the study may not be on functions but on driving behaviour in general. Stakeholders may have different ideas about the functions they want to test, and function descriptions are not always clear.

One issue is whether a function is generic or manufacturer-specific. In other words, how far should the particular manufacturer's specification and functionality be considered? The specifics of the function can clearly have an impact on acceptance and behaviour. An example here might be an LDW (Lane Departure Warning) that gives auditory and visual warnings and a different one that provides haptic feedback through the steering wheel. It is possible that one design will turn out to be more effective than the other.

When it is possible to start with a clear function description, this will allow a detailed planning of the data collection plans, of the experimental design and hence of the costs. But it is of course still necessary to further specify research questions, hypotheses and performance indicators. In the case that functions are not well-defined or a decision has not been made before the start of the project as to which functions to investigate, it may be more

advisable to start with defining the research questions and select functions that seem most useful in answering these questions.

The issue of how many functions can be investigated in an FOT is a matter of the resources to be deployed and of whether the impacts of each function are to be investigated separately or alternatively, and whether the functions are to be treated as a package. Multiple functions can have interaction effects with each other, and combinations of functions can therefore have impacts which are not simply the sum of the individual effects. On the other hand, it may not be feasible to get the functions to work separately – for example, FCW is now generally provided in combination with ACC so that when using ACC it is not possible to switch off FCW.

Similar considerations apply to the number of research questions that can be addressed in an FOT: available resources, stakeholder interests, and practicalities related to collecting sufficient data for evaluation. The early phases of planning a test are also about reaching common agreements on goals and their priority.

The larger the FOT, the more data can probably be collected and different aspects evaluated. Still, clarity on main goals and stakeholders' interests is important to get all parties committed and to give the FOT an efficient start. The final outcome of the process of generating hypotheses and indicators should not be a mere wish-list of different data types to collect, lacking justification on points, but a commonly agreed scope for testing.

4.8.1 Step 1: Selection and description of functions

Usually, it is quite clear from the beginning what functions or at least what type of new functions or features will be the object of an FOT. However, in order to further select and identify details and boundaries of the tested functions – but also in case the functions have not yet been decided – a Stakeholder Analysis is recommended. During this analysis, the needs of the different stakeholders need to be identified and merged into a common requirements description. Stakeholders are those whose interests are affected by the issue or whose activities strongly affect it, those who possess information, resources and expertise needed for strategy formulation and implementation, and those who control relevant implementations or instruments, like customers, public authorities, OEMs, suppliers, and the scientific community. It is of vital importance that all relevant stakeholders are included in the analysis to guarantee that the selection process will not itself bias, from the beginning, the appraisal of the FOT results.

The basis for all following FOT planning steps is a sufficient description of the selected functions. It is suggested to collect the necessary information into a spreadsheet, structured in two parts:

- A first one, the functional classification, where a short high-level description of the main aspects of the function should be given. This information is usually provided through the system specifications given by the system vendor or OEM.
- The second part of the description should comprise limitations, boundary conditions and additional information which is necessary to understand how the function works.

The boundary conditions part should describe where and under what circumstances the system/function will operate according to its specifications, where the FOT should take place, and which type of data needs to be recorded during the FOT to enable a good interpretation of the results. It consists of:

- Infrastructure requirements, cooperative systems and nomadic devices requirements. Here all required actors besides the actual system need to be mentioned, which might have an impact on system performance, service availability or similar. It is intended to trigger the consideration of factors which are external to the system/function under evaluation
- Demographic requirements/driver requirements. Especially the user or driver recruitment process needs to take into account whether a function is particularly designed for a specific group of users or drivers. Drivers differ on a large variety of characteristics, which may all have an influence on how they drive and use different systems and services. These differences may be important to take into account when planning an FOT. Four categories of driver characteristics may be distinguished:
 - Demographic characteristics: gender, age, country, educational level, income, socio-cultural background, life and living situation, etc.
 - Driving experience, and driving situation and motivation: experience in years and in mileage, professional, tourist, with or without passengers and children etc.
 - Personality traits and physical characteristics: sensation seeking, locus of control, cognitive skills, physical impairments or weaknesses etc.
 - Attitudes and intentions: attitudes towards safety, environment, technology etc.
- Geographical requirements/road context. This is necessary for systems which, concerning their functionality, depend strongly on the horizontal or vertical curves of the road layout or on the road type. For example, certain speed limit information systems depend largely on the availability of speed limit information on a digital map, which at present is only commercially available on high-class roads.
- Geographical requirements/environmental restrictions. Certain systems are especially designed for specific environmental conditions or, on the other hand, specifications might indicate that the system under evaluation will not work under certain environmental conditions. In this case the location of the FOT needs to be selected carefully and the relevant data must be recorded during the FOT. For example, most of the functions using a perception system will be affected by adverse weather conditions. If this is the case, it is necessary to log relevant data and take it into account for later data analysis.
- Geographical requirements/traffic context. The performance of certain systems might depend on the traffic context, that is, the traffic density (e.g. given by the Level of Service) or might even be designed to work in specific traffic densities only. Like the other geographical requirements, this needs to be taken into account when an FOT is planned, performed and the data is analysed.

- Other limitations. All other limitations need to be mentioned which might have considerable impact on the performance of functions or systems, since these limitations have a major impact on the experimental design and data analysis.

4.8.2 Step 2: Definition of use cases and situations

FOTs will typically test technically mature ICT systems, as they are the ones that can be tested with large user groups. Therefore, systems and functions to be tested are on the market or close to market introduction. FOTs often involve a couple of variations of the same system, e.g. a navigation system. But the list of functions to analyse grows too long if all possible implementation variations and technologies are considered separately. The use cases put the systems and functions at a suitable level of abstraction in order to group technology-independent functionalities and answer more holistic research questions described later.

Table 4.1 Use cases, situations, scenarios, and their mutual dependence

Subject	Definition	Comment	Example
Use Case	A set of operations in which a system or its specific function is expected to interact with a user or another system to achieve a goal	A use case in driving context is often a combined state of a system and a driver	Car following
Situation	One specific level or a combination of specific levels of situational variables	Thus, a situation is a state of the environment	Rainy weather + darkness + motorway driving
Scenario	A use case in a specific situation	Use case + situation = scenario	Car following on the motorway in rainy weather and darkness

A use case is a textual presentation or a story about the usage of the system told from an end user's perspective. Jacobson et al. (1995) defined use cases as follows: "When a user uses the system, she or he will perform a behaviourally related sequence of transactions in a dialogue with the system. We call such a special sequence a use case." Use cases are technology-independent and the implementation of the system is not described. Use cases provide a tool for people with different backgrounds (e.g. software developers and non-technology oriented people) to communicate with each other. Use cases form the basic test case set for the system testing. There are a number of different ways to define a use case. Use cases in FESTA are very general descriptions, like "car following". This needs to be refined to a reasonable level of detail, which is done by describing "situations" (see [Table 4.1](#)). It is the detailed scenario description that triggers the development of specific hypotheses for later analysis.

Situations are defined as a combination of certain characteristics of a use case. Situations can be derived from use cases by compiling a reasonable permutation of the case

characteristics. Situational descriptors characterise a given situation as the combination of several situational variables (i.e. rainy conditions and speed over 100 km/h and speed regulator active). Situational variables are defined in more detail in [Section 5.3.4](#). Situational descriptors can be distinguished as static or dynamic, where static describes attributes that will not change significantly during one ride of the vehicle, such as age or gender of the driver. Nevertheless, this information needs to be stated, since it is one of the main inputs for filtering huge amounts of data in the later stages of data analysis. The second type of attribute is dynamic, since it can change during one ride of the vehicle, such as the system action status (on or off), traffic conditions, road characteristics or the environmental situation.

The identification of possible situations is covered from three viewpoints:

1. Systems and vehicle specification
2. Environmental conditions specification and
3. Driver characteristics and status specification.

The situational descriptors conform to the following structure:

IDENTIFICATION AND DESCRIPTION

Use case name	A name for identification purposes.
Description	General description of the use case to get a quick overview.
Occurrence	Information about the anticipated quantity of occurrences has implications for the amount of data to be analysed.

SYSTEMS AND VEHICLES

System status	Depending on the hypotheses, the analysis might concentrate on situations where the system is activated or present. <i>Example: ON/OFF (baseline) or IDLE/ON/OFF</i>
System action status	Depending on the hypotheses, the analysis might aim to compare driving performance between different system statuses, e.g. whether the system is in active mode. <i>Example: acting/not acting (meaning e.g. ACC controlling car speed or not)</i>
System/function characteristics	Depending on the hypotheses, an analysis of system performance with respect to special system/function characteristics might be conducted, e.g. examining differences in system performance between nomadic smartphones, PND...) or effects that depend on vehicle type. <i>Example: passenger vehicle/truck/bus</i>

Interaction between systems	The system and especially driver behaviour might change depending on whether the system under evaluation is a support system or whether interactions between two or more systems are foreseen. <i>Example: interaction between Blind Spot Warning and Lane Departure Warning.</i>
-----------------------------	--

ENVIRONMENTAL CONDITIONS

Traffic conditions	Performance of some systems might differ depending on traffic density. Others might only be reasonable with a minimal traffic density. <i>Example: Level of Service A and B</i>
Environmental situation	System performance differs depending on lighting and weather conditions like rain/snowfall/icy roads, etc. <i>Example: normal/adverse weather conditions</i>
Road characteristics	E.g. type of road gradient, super elevation, curvature, curviness, since some systems are dedicated to improving driving performance in curves etc. <i>Example: urban roads/rural roads/highways</i>
Geographical characteristics	Information about geographical characteristics relevant to the systems. <i>Example: mountainous/flat areas, metropolises with high street canyons.</i>

DRIVER CHARACTERISTICS AND STATUS

Driver specification	The characteristics of the users have an impact on driving performance. Even if no specific impacts are expected of certain characteristics, some outcomes may be explained better with more knowledge about the participants. A minimum set of data such as age, gender, income group and educational level is easy to gather from participants. Information about driving experience is also important. Section 5.6.6 gives more detail on driver characteristics and usage of questionnaires for better understanding of driver behaviour and acceptance evaluation.
Driver status	Mindset of the driver <i>Example: attentive/distracted/impaired</i>
Purpose, distance, duration	Describes the different attributes of a trip (time between ignition on and ignition off). All three aspects have an impact on driver behaviour and hence on patterns in the data

A basic set of rules governs the design of situations for an FOT:

1. Complementary: situations should not overlap, for easier analysis.

2. Entirety: the sum of all situations should describe the complete use case.
3. Baseline: The same situation without the use of the systems (system off or non-present) is defined as the baseline. The baseline is the basis for the benefit assessment of the system and the comparison between systems. Therefore, for the same use case, there can be many baselines depending on the number of situations.
4. Comparability: functions compared in an FOT need to have the same use cases and therefore same baseline and situations.
5. Variability of situation parameters: depending on the point of view (user, trip, vehicle, single FOT, multiple FOTs, etc.), attributes describing a situation can vary considerably or not at all.

This list is non-exhaustive and can be extended if necessary.

Finally, from all the possible situations the relevant ones will have to be selected for scenarios of interest in an FOT. Scenarios are defined as a use case in a specific situation, therefore one or more scenarios should be considered for each use case. All other situations should be considered outside the scope of the FOT study. However, if feasible, data should still be collected in all situations for possible future use in an alternative study.

The process of defining the use cases will help with the next steps of the FOT – the definition of research questions and hypotheses and finally the identification of needed indicators. The scenarios as they are defined at this stage of the FOT are not detailed enough for data analysis purposes; therefore, following the definition of indicators the scenarios (and their situations) will need to be further described in terms of events. Only then can the scenarios be classified with a quantitative measurement tool with respect to the defined indicators.

4.8.3 Step 3: Identification of the research questions

In general terms, the goal of any FOT is to investigate the impacts of mature ICT technologies in real use. The core research questions should therefore focus on **impacts**, but there are other questions that 'surround' this core.

The research questions specific to an FOT can only be fixed once the overall goals of the FOT have been established. However, FOTs dealing with transportation generally start from key safety, mobility, efficiency, and environmental impacts and related research questions. The range of questions is considered below. The list should be considered a first step in any FOT and not a comprehensive set of questions.

LEVEL OF SYSTEM USAGE

What factors affect usage of the functions? Examples include:

- Purpose of journeys where the system is used
- Types of road on which the system is used
- Traffic density
- Headway
- Weather conditions

- Ambient lighting

How do driver characteristics affect usage of the functions? Examples include:

- Personal characteristics (e.g. age, vision)
- Socio-economic characteristics (e.g. family, friends, employment status)
- Journey-related characteristics (e.g. other car occupants, shared driving)

IMPACTS OF SYSTEM USAGE

What are the impacts on safety?

- Exposure
- Risk of accident or injury
- Incidents and near accidents
- Accidents

What are the impacts on **personal** mobility?

- Individual driving behaviour
- Travel behaviour
- Comfort

What are the impacts on traffic efficiency?

- Traffic flow (speed, travel time, punctuality)
- Traffic volume
- Accessibility

What are the impacts on the environment?

- CO₂ emissions
- Pollution
- Noise

IMPLICATIONS OF MEASURED IMPACTS

What are the implications for policy?

- Policy decisions
- Laws, directives and enforcement
- Future funding
- Public authority implications
- Emergency service implications

What are the implications for business models?

- Predictions for system uptake
- User expectations
- Pricing models

What are the implications for system design and development?

- HMI design and usability
- Perceived value of service
- Device design
- Communications networks
- Interoperability issues

What are the implications for the public?

- Public information/education
- Changes in legislation
- Inclusive access to systems
- Data protection

When defining the research questions, the trade-off between an in-depth analysis of a few research questions vs. a general analysis of many research questions should be considered, bearing in mind that focus on a few research questions often produces scientifically sound results. However, it is important that even the unaddressed research questions are saved in order to serve as an inspiration for future studies.

4.8.4 Step 4: Creation of hypotheses

Once the key research questions for the FOT have been identified, hypotheses can be derived. The process of formulating hypotheses translates the general research questions into more specific and statistically testable hypotheses.

FESTA distinguishes between more general and open research questions and more specific hypotheses. The definition of a research question in the case of an FOT is “a general question to be answered by compiling and testing related specific hypotheses.” An example would be, “Does having a Forward Collision Warning system improve safety in driving?”

A hypothesis is here defined as: (the definitions used here and below are from the glossary of the EuroFOT project):

“A specific statement linking a cause to an effect and based on a mechanism linking the two. It is applied to one or more functions and can be tested with statistical means by analysing specific performance indicators in specific scenarios. A hypothesis is expected to predict the direction of the expected change.”

The term “function” is used because a particular system may have a number of distinct functions – for example, one system could provide both Adaptive Cruise Control and a

Forward Collision Warning. It is also the case that one function can be provided by different systems. An example of a hypothesis might be, "Forward Collision Warning will have the direct effect of an increase in minimum Time to Collision (TTC)."

There is no process that can guarantee that all the "correct" hypotheses are formulated. To a large extent, creating hypotheses is an intuitive process in which a combination of knowledge and judgement is applied. Nevertheless, a number of recommendations can be made about how this process should be conducted. These recommendations have been tested in FESTA and FOT-Net workshops and modified based on the experience of and feedback from FOTs.

Two complementary ways to develop hypotheses have been used. Both need to be followed, while it is not of importance which step is taken first. One of the steps follows the sequential check of specific areas in which functions can have an impact; the other is fully based on the description of specific scenarios. While the one step results mainly in general hypotheses, the other triggers the development of very specific hypotheses in specific driving situations or scenarios.

Top-down approaches

The six areas approach

The six areas of impact defined by FESTA are based on Draskóczy et al. (1998). Although this approach was originally designed for formulating hypotheses on traffic safety impacts, it is in fact equally applicable for efficiency and environmental impacts.

The six areas are:

1. Direct effects of a system on the user and on driving
2. Indirect (behavioural adaptation) effects of the system on the user
3. Indirect (behavioural adaptation) effects of the system on the non-user (imitating effect)
4. Modification of interaction between users and non-users (including vulnerable road users)
5. Modifying accident consequences (e.g. by improving rescue, etc. – note that this can affect efficiency and environment as well as safety)
6. Effects of combination with other systems.

It is of no particular importance to which of these areas a particular hypothesis is allocated.

The six areas are instead to be used as a checklist to ensure consideration of multiple aspects of system impact.

In applying this procedure, it should be noted that:

- Area 1 includes the human–machine interaction aspects of system use.
- The **driving task** (see Figure 4.1) can be defined, according to Michon (1985), as having three levels: strategic, tactical (manoeuvring) and control. All three need to be considered. All are affected by input from external conditions, which because

they are external to the driver can include the vehicle and devices within the vehicle.

The *Strategic Level* covers longer-term planning and responses to traffic conditions including potential modifications to:

- Mode choice
- Route choice
- Exposure (frequency and/or length of travel).

The *Tactical (manoeuvring) Level* includes potential modifications to speed choice and the effects of such modification on manoeuvring and interaction with other road users. These typically happen in time spans of seconds.

The *Control Level* also includes potential modifications to speed choice and the effects of such modification on vehicle control. Such modifications can occur in milliseconds.

- Consideration should be given to such **mediating factors** as user/driver state, experience, journey purpose, etc.

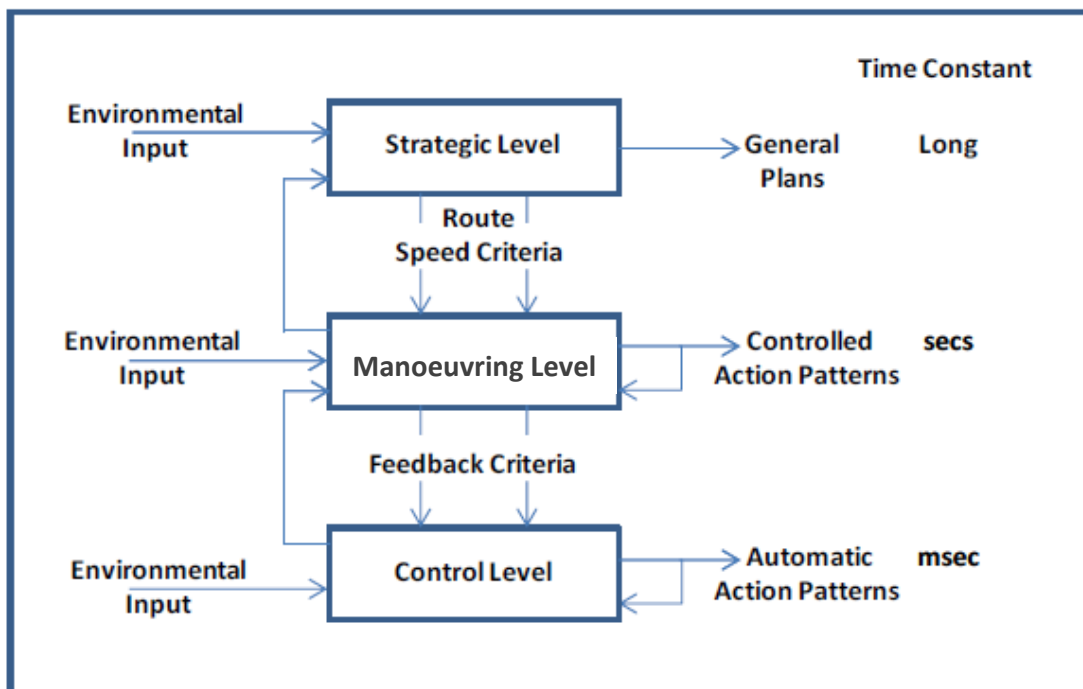


Figure 4.1 The three-level model of the driving task, based on Michon (1985)

It should also be noted that the **effects** of system use may be:

- Short-term or long-term in terms of *duration*
- Intended or unintended in terms of *system design*

Impact area approach

Another useful top-down approach starts from the most relevant impacts areas, which are: Efficiency, Environment, Mobility, Safety and User Uptake.

The basic principle for generating hypotheses using this top-down approach lies in a theoretical understanding of the factors that influence the different impact areas. It should be noted that there are likely to be overlaps of these factors among the impact areas under consideration, and hence the same research questions and resulting hypotheses will be applicable across more than one impact area. The approach will result in *generic* research questions that are independent of any system functionality.

The procedure for generating hypotheses in this top-down approach is as follows:

- The impact area should be considered in its entire context and primary measures affecting that area identified.
- Secondary factors of these measures are then identified that can be used to explain the variations in the primary measures.
- Finally, the variables affecting the secondary measures are identified.
- The variables identified form the basis of the generic research question “*Is there a change in the variable?*” and the hypothesis based upon an anticipated effect of the variable “*The variable will increase/decrease.*”

This procedure should be undertaken for each of the impact areas.

Using the Safety Impact Area as an example:

- The primary measures affecting safety would be the “Number of events (accidents, near misses) that occur” and the “Severity of the event”.
- Secondary factors affecting the first of these measures would, for example, be
 - “Exposure of the vehicle on the road”, “Driving style of the driver”, “Distraction of the driver from the driving task” and “Any interaction with the fitted device”.
 - Considering the factor “Exposure”, this can be measured with the following variables: “Length of journey”, “Number of trips undertaken” and “Road type used”.

These variables lead to the following research questions:

- Does the system affect the length (miles) of journeys?
- Does the system affect the duration (hours) of journeys?
- Does the system affect the number of journeys undertaken?
- Does the system affect the road type used?

This leads to the generic hypotheses that can be tested in a statistical manner. The direction each hypothesis should take (e.g. increase or decrease) is based upon the anticipated effect once the top-down approach is integrated with the bottom-up (system defined) approach.

- Journey lengths will increase/decrease when the system is used compared to when it is not used.
- Journey duration will increase/decrease when the system is used compared to when it is not used.
- The number of journeys will increase/decrease when the system is used compared to when it is not used.
- The use of rural roads/motorways/major roads will increase/decrease when the system is used compared to when it is not used.

Bottom-up: the use case approach

This process leads to the development of hypotheses concerning specific scenarios. These scenarios are derived from the combination of use cases and situations. Scenarios should be covered systematically. It is recommended that a structured approach be used in scenario development and that an Excel spreadsheet is used as a record.

Prioritizing the hypotheses

A complete list of the hypotheses that have been developed should be recorded. If it is considered that some are too trivial or too expensive to address in the subsequent study design and data collection, the reasons for not covering them should be recorded. It should also be noted that there are standardised techniques for observing driving behaviour with manual observers that may be less resource intensive than using dedicated data recording. Observations using such techniques can be carried out at various times during the study, preferably along a fixed route.

A huge number of research questions and associated hypotheses from the top-down and the bottom-up approaches will be developed. A key task is to integrate both sets of hypotheses in the context of each FOT. It is envisaged that the bottom-up approach will form the basis of the hypotheses list for an FOT and that the top-down approach will be used to check that nothing significant for a particular impact area has been omitted.

Once the integration has taken place, the list of hypotheses is still likely to be large. In order to derive a final, manageable set of research questions and hypotheses that can be applied throughout the various test sites, a cost-benefit approach is proposed. Using this approach, an assessment is made regarding the likely “costs” of collecting the data.

Costs can be represented in terms of effort required to derive a performance indicator expressed predominantly in terms of resources. This should be offset against the likely “benefit” that proving/disproving the hypotheses will have. This is measured by way of the likely contribution towards providing a significant answer to the research question and thus the level of contribution to the impact assessment. To some degree, this will depend upon the stakeholder needs and requirements, and therefore a prioritisation of their needs should be considered.

Summary

The basic set of recommendations is:

- A structured approach should be applied linking a **top-down** approach at the global system level with a **bottom-up** approach which looks more at system states

and what can arise from them. FESTA considers it mandatory to combine the two approaches.

- A multidisciplinary team should *jointly* develop the hypotheses. A workshop at which participants can brainstorm and debate is recommended to achieve this. Participants in the process should include design engineers, traffic engineers and behavioural scientists, ideally including both behavioural psychologists and human factors experts.
- The process should iterate between the top-down and bottom-up approaches. It is not particularly important which is performed first, but it is important to cross-check one approach by using the other.
- An important output of the process is the initial selection of the performance indicators to be used in testing the hypotheses.

4.8.5 Step 5: Link hypotheses with indicators for quantitative analyses

Some of the hypotheses will already incorporate an indicator which needs to be measured, e.g. a very concrete hypothesis like “The function will increase time to collision (TTC)”. In this case, it is obvious which indicator to choose, while the method to measure TTC might include complicated procedures and/or costly measurement equipment. [Chapter 5](#) gives an overview of a wide range of reasonable indicators. These should be considered when planning the experimental design, since a detailed description of how to calculate the indicators from measurements is also provided.

Other hypotheses might be rather unspecific, but still reasonable after rephrasing into testable ones. This rephrasing goes hand in hand with the identification of related reasonable indicators. For example, a hypothesis like “The function will increase lane changing performance” is not directly testable, since “lane change performance” is not an indicator itself. Hence, surrogate measures must be identified to evaluate lane change performance. These surrogate measures or indicators can be found e.g. in publications of corresponding research projects. If appropriate information cannot be found or is not accessible, new performance indicators need to be developed. Those indicators and the measurement methodology must be valid, reliable and sensitive; that is, the measurement must actually measure what it is supposed to measure, the measures must be reproducible and they must be sensitive to changes of the variable. A sensitivity analysis should be performed beforehand during a pilot study to make sure that the new performance indicator is suitable. When one or more surrogate measures have been identified, the initial hypothesis can be reformulated into one or more testable hypotheses. In the above example, reasonable indicators associated with “lane change performance” might be: use of turning indicator or number of lane change warnings. The initial hypothesis will then be reformulated into: “The system will increase the use of the turning indicator.” and “During the system use, the number of lane departure warnings will decrease.” The next step is then to evaluate how the indicators “use of turning indicator” and “lane departure warnings” can be measured. In this context, [Chapter 5](#) provides useful information.

4.8.6 Iteration

Iteration is especially important when defining research questions and hypotheses, because usually a selection has to be made from a large number of possible hypotheses, based both on their relation to the main impact areas and research questions and on practical issues. Another important iteration point is the impact areas. The final question

of the impact assessment may drive the design of the FOT in all its aspects. When practical issues, such as which data-loggers to use, make certain choices hard to realise, iteration to earlier stages is necessary. Cost-benefit analyses and feasibility assessment of different options for the FOT may also drive the design. It is important that there is a good communication between the project members who are in charge of defining research questions and hypotheses and those who will be analysing the data, in order to ensure that the questions can indeed be answered.

4.9 Functions, research questions and baseline in CCAM FOTs

The steps in this chapter start with the definition of functions; for automated vehicles this may not concern one single new function but a whole range of interrelated functions – or at least, components. Several well-working components are commonly needed to implement automated driving, for example: HMI, maps, positioning, environmental sensing (including object and lane detection), route and collision avoidance features, and connectivity.

The driving style of a vehicle, used methods and safety considerations may change according to different route segments and different weather conditions. To support analysis of automated driving in different conditions and areas, it has become common to define and utilize Operational Design Domains (ODD). The ODDs basically define where and under which conditions the automated functions are working confidently. Such definitions can also enable grouping of related scenarios. For example, if an ODD is defined to be a highway of certain type, driving strategy and accident probabilities related to animals crossing such a road could be studied as a group. This would reflect the traditional analysis of accidents on different road types. However, an ODD could also be defined as a small segment of a single route, for example a specific bus stop. Splitting a fixed route into segments can enable detailed analysis of operations and risks in these different parts. Currently, ODD-related evaluation concepts are still new and vary between projects.

Automated driving is often closely related to connectivity between vehicles, and between vehicles and the infrastructure. In this handbook, sections are devoted to cooperative systems, as several large FOTs have been conducted on these systems. However, as connectivity becomes a major feature in automation, not only the behaviour of the vehicles needs to be studied but also the functioning of other connected components and their communications.

Experience from FOTs shows that defining and selecting research questions may be one of the hardest parts of setting up an FOT. As CCAM FOTs are dealing with new types of vehicle and mobility, as well as with new impact areas, many new questions will arise.

We may distinguish three types of research questions for CCAM FOTs:

- User-centred tests addressing questions about how the user/driver reacts to automation and uses it; how (s)he understands the capabilities of the vehicle; what (s)he will do when not supervising the surrounding vehicles/traffic; the transition from automated driving back to driver driving during the journey; drivers' situation awareness; the interaction between automated vehicles, their drivers and other road users including vulnerable road users and drivers of conventional vehicles;

user acceptance; changes in mobility patterns; etc. In industrial and workplace context, changes are studied regarding work practices, interaction etc.

- Vehicle-centred tests addressing the question of how the automated vehicle behaves in different traffic conditions and ODDs. Here questions about the situational picture that the automated vehicle creates, and the interaction of the automated vehicle with other automated and conventional vehicles, other road users and with the infrastructure need to be answered. The tests could cover a variety of driving conditions, levels of automation and penetration rates. The traffic flow of mixed traffic of lower levels of automation, and of mainly fully automated vehicles may have different dynamics. New types of traffic accidents and conflicts may appear.
- Context-centred tests addressing questions of how mobility of different user groups and the services they use change, what the impacts are on the traffic flow level or on the transport system level, how the position of other road-users like pedestrians and cyclists is affected, what ethical choices might be involved, and what would be the impacts on the built-up environment, industry and society. These types of questions are extremely important but not easy to investigate with FOTs as these impacts typically take a longer period of time to evolve than the duration of an FOT.

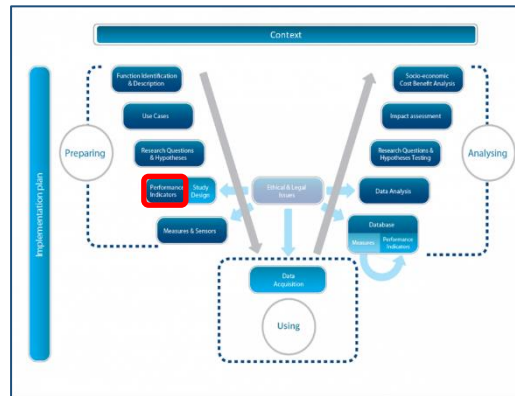
In FESTA, it is recommended that driving with the system is compared with driving without it (the baseline). For fully automated vehicles we no longer have a system that can be viewed as independent from the vehicle itself, the whole vehicle is now the system. Some forms of automation, like autonomous or driverless vehicles, represent a radical change in transport, with no clear baseline available, and studying new emerging patterns may be of more interest.

If we want to compare automated driving with the current situation, Naturalistic Driving Studies might serve as the baseline. In these studies, "normal" everyday driving has been studied in large projects in several countries. Comparisons with lower levels of automation can be made, but may not always be feasible or useful, as these several levels do not provide a continuum in automation.

5 Performance Indicators

5.1 Introduction

During the process of developing hypotheses it is important to choose appropriate performance indicators (PIs) that will allow answering the hypotheses, but that will also be obtainable within the budget and other limitations of the project. Many different kinds of PIs have been used in previous studies



and are related to various aspects of driving. A definition and description of the PI are given below. It is explained how PIs are related to measures, and the different types of measures that have been identified are described. Examples are provided to illustrate the concepts.

Once the PIs and measures have been defined and linked, it is necessary to test how well they work, which in practice means testing the whole data transmission chain from sensors/devices, vehicles and/or roadside equipment to processed and uploaded data in the research database. The moment to run these tests is the “piloting phase”, which is a preparation phase just before starting full-scale user tests, further described in [Chapter 6](#) on experimental procedures.

5.2 Definition of performance indicators

Performance indicators are quantitative or qualitative indicators, derived from one or several measures, agreed on beforehand, expressed as a percentage, index, rate or other value, which is monitored at regular or irregular intervals and can be compared to one or more criteria.

- Hypotheses steer the selection of PIs and the criteria against which they should be compared. Hypotheses are seen as questions that can be answered with the help of measurable PIs.
- Criteria can be baseline, different experimental conditions, absolute values, etc. This depends on the research questions and hypotheses.
- New PIs or combinations can be developed during the course of the study. They will have to be validated in follow-up studies.
- A denominator is necessary for a PI. A denominator makes a measure comparable (per time interval/per distance/in a certain location/...). Therefore “crash” or “near-crash” in themselves should rather be considered as “events”, since they only become comparable when they get a denominator, like “number of crashes per year per 100.000 inhabitants.” For certain PIs, either time or distance can be used in the denominator (e.g. number of overtaking manoeuvres, percentage of time exceeding the posted speed limit).

For PIs measured via rating scales and questionnaires, focus groups, interviews, etc., the “denominator” would be the time and circumstances of administering the measuring instruments, for example before the test, after having experienced the system, and so on.

PIs are very diverse in nature. There are general performance as well as detailed PIs, observed and self-reported (subjective) PIs, PIs calculated from continuous and from discrete data, and so on. An example of a rather general PI based on continuous log data would be the mean speed on motorways, whereas an example of a PI based on discrete, self-reported data would be the level of perceived usability of a function. Some PIs can be based on either self-reported, discrete measures or on logged data, such as the rate of use of a system. The participants can be asked how often they use a function, but the actual function activation and the different settings chosen by the driver can also be logged from the system.

All PIs are based on measures, which are combined and/or aggregated in certain ways, and which are normalised in order to allow comparisons. The measures are described below.

5.3 Measures

Four different types of measures are identified, namely Direct Measures, Indirect Measures, Self-Reported Measures, and Situational Variables, which are described in more detail below. A measure does not have a “denominator”. Therefore, it is not in itself comparable to other instances of the same measure or to external criteria. The measure itself, however, can very well be a fraction (like speed). Several PIs can use the same measures as input, and the same measures can be derived from different types of sensors. An example would be speed that can be read from the CAN bus, logged from a GPS receiver, or calculated by an external sensor registering wheel rotations.

For detailed data analysis, it may be important to note the source of the data in PI generation, e.g. by naming the indicator as speed_can or speed_gps.

5.3.1 *Direct (raw) measures*

A direct measure is logged directly from a sensor, without any processing before saving the data to the log file (note that linear transformations like the conversion from m/s to km/h are not considered to be processing). How the sensor arrives at its output is not relevant for the classification. Longitudinal acceleration, for example, is a direct measure if logged directly from an accelerometer, but not if derived from the speed and time log. In this case, it would be a derived measure, because it is not directly available from a sensor and has to be calculated from other measures, i.e. pre-processed, before logging. Further examples of direct measures are raw eye movement data, the distance to the lead vehicle as measured by radar, and a video film of the forward scene.

5.3.2 *Derived measures*

A derived measure is not directly logged from a sensor, but is either a variable that has been filtered, for example, or which is a combination of two or several direct or other derived measures. An example of a derived measure is time to collision (TTC), which is based on the distance between a vehicle and another vehicle or object, divided by their speed difference. The distance to a vehicle or object on a collision course is a direct measure from radar, for example. The speed difference between an own vehicle and other

vehicle or object is another derived measure, based on own speed as read from the CAN bus, for example, and the calculated speed of the other vehicle or object. Another example of a derived measure based on raw eye movement data and vehicle geometry is pre-defined zones that the driver looks at, such as a mirror, windscreen or radio.

One major issue with derived measures is that there is to date rather little standardisation on how to calculate them. Slightly alternative definitions exist also for the TTC that was used before as an example. That means that derived measures calculated in one study may not match those calculated in another study. It also introduces the potential that the calculation procedure adopted may not conform to the best scientific standards and that, as a result, the findings may lack validity. This problem is currently being addressed by the Safety and Human Factors Standards Steering Committee of SAE International, which has adopted a task (Task J2944) termed “Driving Performance Definitions” (https://www.sae.org/standards/content/j2944_201506/). The task is focused on measures related to longitudinal and lateral vehicle control.

It is important to document the expectations of the measures, and this will form the first version of metadata documentation in the project. As an example, vehicle speed must be recorded in at least 10 Hz. In the analysis phase ([Chapter 9](#)), the documentation must be completed with an in-detail description on the origin of the data and the processing steps that have been performed. A proposed structure for managing data and metadata documentation is given in the [Data Sharing Framework](#) (FOT-Net Data and CARTRE, 2019).

A special case of derived measures is those that are coded by a human observer after data logging is completed. Examples might be gaze direction coding, classifications of scenarios or classifications of secondary task engagements. These measures are considered to be “derived”, because data reduction by a human observer is more than only a linear transformation, and they can be based on more than one direct measure. In case of secondary task classification, one might use both a video of the driver’s hands and a log file of an eye tracker, and for scenario classification both a road database and a video of the forward view might be used.

5.3.3 Self-reported measures

A number of PIs are based on self-reported measures, which are gleaned from either questionnaires, rating scales, interviews, focus groups, or other methods requiring introspection on the part of the participant. These subjective measures are typically not logged continuously, but rather only once or a few times during the course of one study. The measures related to self-reported PIs could be the answers to each single question or the checks on the rating scales, while the sensors would be the questionnaires or rating scales themselves. It is more difficult to make a meaningful distinction between measure and sensor for semi- and unstructured interviews and especially for focus groups.

Subjective data, e.g. on acceptance and trust of a system or function, can provide valuable PIs, and in particular such data can be related to function usage in cases where this is within the control of the operator. Consideration should be given to tracking such acceptance and trust over time, as the levels may change with experience of the function.

5.3.4 *Situational variables*

Situational variables are properties of the traffic system that the vehicles have driven in. They can be logged like direct measures or computed like derived measures. They can also be self-reported and they can correspond to events. Their commonality is that they can be used as a differentiation basis for other PIs, in order to allow for a more detailed analysis. It might, for example, be of interest to compare certain PIs in different weather or lighting conditions, on different road types, or for different friction conditions. These situational variables are not linked to any specific PI. In principle, many types of measures can be used as situational variables, such as when analyses are performed for different speed intervals.

Data on situational variables is essential to collect, since it helps to establish important control factors that are needed when analysing the effects observed in the FOT. Ideally a lot of in-depth data is collected, such as:

- Video data (manual video annotation takes a lot of time so it should be automated as much as possible)
- Questionnaires and travel diaries (at certain points during the test, too frequent interventions will disturb the naturalistic approach)
- Data on surroundings (such as surrounding vehicles, headways and traffic state) and conditions (such as traffic density or weather)
- Description of data and tests for evaluation, such as who drove the vehicle, which functions were studied in a particular test drive, circumstances when driving, other functions in the vehicle, date and time of the test, etc.
- Audio data; e.g. to give test drivers the options to tell what happened and what their experiences are, to make it as easy as possible for them – e.g. using a voice memo
- Logging of function states and messages (system state, e.g. on/off, what information is presented to the driver).

Collecting all these data types is costly and time-consuming. It is important to address the specific purpose of the analysis and focus on the specific need. However, some basic data might not be the specific interest in this project but of great importance for future usage, and it is important to find a balance between these two criteria.

5.4 Events

Events can be seen as singularities based on direct measures and/or derived measures or a combination of these. They can be very short in time, like a crash, or extended over a somewhat longer period, like an overtaking manoeuvre. One or more preconditions must be fulfilled for an event to be classified as such, that is, one or several “trigger” criteria must be exceeded. For the event “overtaking manoeuvre”, for example, the non-technical definition might be: A vehicle in a vehicle-following situation changes lanes, accelerates and passes the vehicle in front, then changes lanes back into the original lane, in front of the vehicle(s) that have been overtaken. Depending on the infrastructure design, the definition might need to be extended to motorways with more than two lanes in each direction, for example.

Several performance indicators can be related to one event type, for example for an overtaking manoeuvre it could be of interest to determine the number of overtakings, the duration of overtaking, the distance/time spent in the opposite lane, and so on. For a more technical definition that sets the trigger criteria of when exactly an overtaking manoeuvre starts and when it ends, either the literature has to be consulted or an own definition has to be developed. This can possibly be based on previous data, or, if nothing else is available, on the data from the current FOT.

Events are very important to NDS/FOT studies, because a core type of analysis performed in almost every NDS/FOT is what can be called *Event Based Analysis* (EBA).

5.4.1 Crash Relevant Events

A basic use of EBA is to identify short driving segments called *Crash Relevant Events* (CREs, typically in the order of 5–10 seconds), during which the crash risk is judged to be higher compared to other driving, and then to analyse why these events occur, and/or whether their frequency or dynamics change when particular safety systems are made available to the driver.

A key element to NDS/FOT success is therefore defining CREs in a proper way. If the selected events are indeed crash relevant, then extrapolation to the general driver population is both possible and credible.

While simple in theory, identifying CREs in NDS/FOT data is more difficult in practice. To begin with, actual crashes are incredibly rare, so the final database will usually contain fewer than needed for statistical analysis.

Surrogate events therefore have to be used. These need to be driving situations where there is no actual crash, but where the event still unfolds in such a way that its presence can be used as an indicator of crash risk.

A key challenge for all NDS/FOT studies is therefore how to couple surrogate (non-crash) events to crash causation mechanisms. Ideally, one would only use CREs that are known with certainty to be predictive of actual crash involvement. Unfortunately, while a lot of effort has gone into developing algorithms, filtering techniques etc. that allow for efficient yet relevant CRE selection, precise definitions with a clear-cut, indisputable connection to crash involvement have yet to be fully established. Selecting suitable CRE definitions is therefore a major decision point for a project, and should be treated as such.

Currently, there are four main approaches to CRE definition in use. Short introductions to these are given below.

Driver response based CRE definitions

The first approach can be called the “driver response based” approach, and it builds on the general idea that CREs can be identified from the way drivers respond to them. The most common version of this is to look for extreme vehicle kinematics. The basic assumption is that drivers prefer to travel in comfort and generally will not expose themselves to drastic events unless necessary. Thus, abrupt velocity and direction

changes in the vehicle may indicate unplanned and urgent responses to unexpected situations.

A less common version is to use what can be called the startle response in the driver to find events. The general idea is that unexpected traffic situations that include a perceivable threat to the driver triggers a response in the form of a general tensioning of the body. This “jerk” may be used as a tell-tale that the driver did not expect the situation and that he/she perceives it as genuinely threatening.

In the driver response based approach, CREs are identified based on how each driver evaluates the situation. For example, while one driver may brake hard at a certain time to collision threshold, another driver might not brake at all. Hence a selection of CREs based on this approach will include the first event but not the second (since the driver did not respond, it is by definition not an event). Driver response based CRE selections will therefore reflect the normal variability in any driver population in terms of driving style, risk perception and capacity to respond. It follows that representative selection of drivers becomes a key issue when using a driver response based CRE selection.

Safety function response based CRE definitions

When the study is an FOT, i.e. designed to assess the impact of one or more active safety functions, then a very natural approach to CRE identification is to use the function itself to detect CREs. After all, that is what the function is designed to do. For example, if an FOT is set up to assess the effects of Forward Collision Warning (FCW) on crash risk, the warnings issued can be used as event identifiers.

The downside of this approach is that any CRE that occurs outside the function’s detection capacity will be missing from the analysis. Without alternative processing or data collection, it is not possible to estimate the frequency of CREs which the function in principle needs to detect but in practice cannot. On the upside, a very realistic assessment of function availability and usage is obtained. Since the function can only do something when it is turned on, true availability and usage rates are automatically represented in the dataset.

Driving context based CRE definitions

A third approach to CRE identification is to base it on driving context. The underlying assumption here is that too small margins equal elevated crash risk. In other words, there exist situations where the safety margins are inherently so small that the slightest mistake or variation could lead to a crash. Prevention of whatever leads to these small margins will thus enhance traffic safety.

The definition of what constitutes too small margins can be static, such as when lane markers are used to indicate boundaries that should not be unintentionally crossed. The definition can also be dynamic. For example, if a vehicle is closer than X to another vehicle and simultaneously closing in faster than Y, it might be considered as being in conflict regardless of whether an action is taken or not.

In a driving context approach, the CRE definition is independent of how drivers resolve the situation. This means all drivers are equally covered, independently of their capacity or

willingness to respond. It also means that drivers with a more aggressive driving style will contribute more events per driver to the analysis than those who are less aggressive, since they end up in small-margin situations more often. If the analysis team believes that small margins are predictive of crash involvement, this is okay, since these drivers then should have a higher crash risk. If the team is dubious about this assumption however, this approach might not be the right one for the study.

Driving history based CRE definitions

The fourth approach is to look for unusual events in a driving history perspective. The underlying assumption is that unusual events in a person's or group's driving history are unusual precisely because drivers try to avoid them. Presumably at least a certain portion of them would be crash risk related.

The advantage of this approach is that it will find the most unusual events that occurred during the study for each person (or group), and it is reasonable to assume that those are events which the drivers would prefer to avoid in the future. The corresponding disadvantage is of course that those events may be special for other reasons than being safety-critical. Even if drivers try to avoid them, they may have little or no connection to traffic safety.

Combined approaches

Sometimes projects make combined use of the approaches above. For example, in the EuroFOT project, after multiple versions had been tried, the CRE definition for lead vehicle conflicts that was finally settled on included the following criteria:

- FCW warning issued (Function based)
- Brakes applied within 5 seconds after FCW warning (Driver response based)
- Max Brake Jerk > 10 (Driver response based)
- Max Brake Pressure > 20 bar (Driver response based)
- Lead vehicle = moving (Driving condition based)
- Direction indicators not in use prior to warning (Driving condition based)

Coupling CREs and crash risk – the CRE causation and applicability question

The discussion above illustrates that performance indicators can be built by counting events, or by considering certain aspects of those events. However, it also illustrates that each approach to the CRE definition outlined above also represents a different view of how the coupling between CREs and crash risk should be made, and it is not obvious which is the best way to go. What is clear, however, is that different CRE definitions will lead to different results. Fitch et al. (2008) describe how another project (Hanowski, Blanco et al. 2008) approached the same dataset that Fitch and his colleagues were analysing with a different CRE selection method. Interestingly, while both projects found hundreds of what they judged to be relevant CREs in that dataset, only seven of the 596 CREs found by Hanowski et al. overlapped with those identified by Fitch et al. Clearly, the CRE defined in a project matters.

Thus, FESTA neither can nor will try to provide specific trigger values for events, nor will the exact measures that have to be included for the definition of a certain event be provided. Projects must make an informed, explicit, conscious and preferably well-

documented decision as to how to use the available approaches to best fulfil their goals, and then set up the CRE detection correspondingly.

Some general recommendations can be made, though. First off, it is important to point out that *not everything can be seen* in the data typically collected in NDS/FOT studies. For example, mental states do not show in video and CAN data, so if the objective is to analyse how intentions, expectations or levels of attentiveness contribute to crashes, probably the study to be conducted should not be a pure NDS/FOT. A mixed approach, where the NDS/FOT is complemented by additional data collection for capturing drivers' intentions and expectations, is probably more suitable.

Second, it is important to note that CREs do not have to look like crashes to be relevant for analysis. This comes back to the underlying assumptions about what mechanisms are predictive of crash involvement. If, for example, it is assumed that high levels of variability in normal driving is an accurate crash predictor (a version of the driving context-based CRE selection approach above), then each event that captures the tails of the driving parameter distributions is crash relevant, even if it does not look spectacularly dangerous on video.

Third, the selection criteria have to match the time scale of the event to be analysed. This applies in particular to using physiological driver parameters for CRE selection. Many such parameters, like respiratory rate, simply change too slowly for precise correlations with traffic situation changes (i.e. capturing hard brakings by looking at respiratory rate is not likely to succeed).

Fourth, it may strengthen the credibility of the results if the CRE analysis is restricted to injury-related CREs. Doing this typically includes finding out at which travel speeds and/or on which road types injuries occur for the crash type to be analysed, and then limiting the CRE analysis to events which fall inside those speeds and road types.

5.5 The PI-Measures-Sensors matrix

Establishing a process to come up with data logging needs is one of the key early phases of an FOT. This process commonly starts from an initial listing of research questions and related performance indicators (PIs). Each PI should include information on which measures its calculation would be based. Required data collection frequencies should be defined and calculation formulas could already be given. The measures lead to consider data sources such as sensors, paper diaries and existing data logging products. Overall data and sensor needs should be outlined for project-wide discussion and further data logging development.

To complement this initial list of data needs, evaluation experts should also estimate, how much data (e.g. test duration or number of events/trips) would be necessary to reach statistical conclusions.

Discussion on data needs touches especially system integrators and software developers, who need to be aware and comment on the data requirements and available measurement options early on. Test site leaders should be made aware e.g. on data logging differences between baseline and treatment periods. Baseline data collection might have to start much

earlier than the actual system tests. The discussions are iterative in nature and reaching a final agreement on data collection can be a process of many months. Research questions and data needs are the main reason why evaluation work needs to start early on in FOT projects.

During detailed data logging development, it is common that PI definitions and measures/signals are further defined regarding e.g. their final naming in data, pre-processing techniques, values for the signal not being available etc. The list of PIs and derived measures commonly grows during data management and evaluation tool development. However, the sensors to install need to be fixed much earlier.

An Excel workbook was developed in the original FESTA project, in 2008, which listed several PIs, measures and sensors to support vehicle data collection planning (FESTA D2.1, 2008b). While there are little changes to logging vehicle speed or GPS, even today, the variety of data that different FOTs collect has since grown to include, for example, communication, situational data and automatically processed video. Instead of attempting to provide a direct comprehensive update, the following tables outline aspects to keep in mind, when defining PIs, measures and sensing.

Table 5.1 gives an example of defining performance indicators and, more generally, derived measures. While (key) performance indicators are to indicate meaningful changes, the term derived measures is simply about next steps of calculations for evaluation and e.g. visualisation purposes. However, both are linked to direct measures coming from data sources.

Table 5.1 Example on defining a performance indicator (or a derived measure).

Performance indicator ID	P01
Name/description	Mean speed of the vehicle
Name in data	mean_speed_ms
Unit	m/s
Required frequency	10 Hz
Required resolution/accuracy	0.1 m/s
Equation	(Calculation formula)
Required measures	Speed_CAN OR Speed_GPS OR Speed_WheelUnitDistance
Subjective/objective	objective
Qualitative/quantitative	quantitative
Related research question IDs or evaluation areas	safety, environment, efficiency, acceptance
Ethical issues	may highlight illegal behaviour (speeding)
References	(Bibliography)

All the direct measures would then be described in another table (example: Table 5.2), providing details on sensor sources and signal processing. Besides the term measures, also the term signals is in common use in this context, as the data is usually lightly processed sensor data.

Table 5.2 Example on defining a measure (or a signal).

Measure name (key)	Speed_CAN
Description	Vehicle speed obtained from the CAN bus
Measure group	Vehicle data
Unit	m/s
Required frequency (Hz)	10
Required resolution/accuracy	0.1 m/s
Lower limit (for validity checks)	-360
Upper limit	1000
Datatype	double
Pre-processing/interpolation	(sampling technique)
Sensor source	Sensor ID
Value for not available	-1000

Sensors and their properties and installation details should finally be outlined in a sensor table. A project may see the need to list e.g. sensor dimensions, power requirements, calibration needs, environmental constraints and crashworthiness, as part of integration work and data logging development.

As part of initial discussions in an FOT, the listed sensors and data sources may need to be linked back to research questions, to clarify where the need for such installations comes from and which analyses would be impossible without.

5.6 Performance indicators per impact area

The PIs are split into different sub-groups, depending on which area of the traffic system they are concerned with.

5.6.1 Indicators of driving performance and safety

Driving performance is discussed and analysed in relation to traffic safety. Given that accidents are usually multi-causal, the desired set of indicators should cover a number of factors. Otherwise, any FOT is likely to miss essential information that is required to produce reliable and valid results.

Traffic safety is regarded as a multiplication of three orthogonal factors, namely exposure, accident risk and injury risk (Nilsson, 2004). The driver's decision-making and behaviour cover all these aspects. Typically, strategic decisions are highly relevant for exposure, tactical decisions for the risk of a collision, and operational decisions for the risk of injuries (Michon, 1985). Consequently, an FOT should cover all these aspects, because it is essential to cover driver tasks and driver behaviour widely, and include decisions like whether to use the vehicle at all, route planning before the trip, timing of the trip, etc. However, as those decisions often lie outside what can be influenced with available countermeasures, the main focus is usually on driving performance while actually driving a vehicle.

The most common approach to traffic safety in NDS studies is to contrast driver behaviour: normal driving and the sequence of events leading up to conflict situations (i.e. near-crashes or crashes).

In summary, an *indicator of driving performance* is a behavioural variable which *indicates* the quality of the behaviour in respect to road safety. The behaviour is measured directly from the driver (e.g. frequency of glances towards given object) or indirectly from the vehicle (e.g. speed).

5.6.2 Indicators of system performance and influence on driver behaviour

In this part, indicators were developed that describe the actual performance of the system to be tested. These indicators are mostly related to both *safety* and *acceptability*. Here the focus is directed at the question of whether the system actually functions the way it is supposed to under realistic conditions. False alarms and misses are obvious indicators in this regard. Relations exist with indicators of acceptance and trust which examine the subjective opinion of the participants on how the system worked.

Furthermore, indicators that describe the influence of the system on the driver and the interaction between system and driver are described. They will enable assessing the driver's willingness to use the system in various situational contexts. They will also contribute to the identification of potential misuses of the system leading to incidents or conflicts. In a longitudinal perspective, they will also contribute to an analysis of the learning and appropriation phases.

Intrinsic performance of the system

The first issue is the intrinsic performance of the system studied. It is related to the *precision and the reliability of the system*. Does the system perform as expected? In this case, we need indicators signalling any deviations, such as false alarms and misses, but also indicators about the context in which these deviations occur. Ideally, the origin of the deviation should also be identified. The identification of false alarms or misses may be based on automated sensors or may require a video recording of the driving scene. For example, in the French LAVIA (ISA) project, loss of the recommended or target speed was automatically recorded, while mismatching between the target speed and the posted speed limit was identified from a video recording of the driving scene.

The intrinsic performance of the system should be distinguished from the operational envelope of the system (i.e. the use cases for which the system was designed to work).

This is important when assessing opinion on the performance of the system: when asking the driver to assess the system performance, the limits of the system operation should be differentiated from system deviations. Two main indicators related to the operational envelope are 1) availability of the system over driving time (percentage of the driving time the system is available, e.g. some systems are only available above a certain speed, for special road characteristics, etc.); and 2) frequency of take-over requests (the system is active but not able to provide assistance due to system limits, e.g. for ACC the maximum brake rate is limited).

Both intrinsic performance and the competence envelope are assumed to play a role in driver opinion of the system.

Modes of drivers' interaction with the system

The second issue is driver interaction with the system. This goes beyond the analysis of overall driving performance when using support systems. In fact, 1) it is examined how drivers use and interact with the system; and 2) it is examined how this interaction may affect driving behaviour and performance.

How drivers use and interact with the system

Some support systems require/enable the driver to activate/deactivate the system, to override the system, to select one system among other systems available, to select or to register some vehicle-following or speed thresholds, and so on. In other words, using a system implies the application of a number of procedures, and these procedures should be registered and analysed. This is the case for systems such as speed limiters, cruise control, adaptive cruise control or navigation systems, for example. These procedures may be classified as the driver's direct or indirect interventions, depending on whether they are applied through vehicle controls (brake or accelerator) or through system controls. As for the indicators of system performance, the situational context should be taken into account. This is important for identifying potential misuses of the system leading to incidents or conflicts as described above. In a longitudinal perspective, these indicators will also contribute to an analysis of the evolution of system usage from the learning and appropriation phases to the integration phase. Furthermore, the frequency with which the system "interferes" with the driver's activity has to be assessed. For example, when driving with a speed limiter, how often is the system "active", that is, effectively limiting the vehicle speed?

How this interaction may affect driving behaviour and performance

For analysing the effect of driver interaction with the system on driving behaviour and performance, various levels of analysis could be employed depending on the desired level of granularity of analysis. Obviously, this granularity depends on the recording means available as well as on the time required for performing such analyses. For example, studying changes in glance behaviour requires video recordings and is time consuming.

For an analysis of behavioural changes at a more general level, synthetic indicators should be conceived. These indicators are assumed to reflect changes at the tactical or strategic level of the driving task. Indicators such as "lane occupancy" and "frequency of lane change" are often used to assess changes at the tactical level. Changes at the strategic level could be reflected by changes in the itinerary chosen or in driving time.

Recommendations:

- Classify the support systems by type and level of interaction implied by their use;
- Classify the performance indicators according to the level of granularity of analysis that they permit;
- Classify the performance indicators according to the means and time required for collecting and analysing them.

5.6.3 Performance indicators of environmental aspects

Exhaust emissions include many different substances like HC, CO, NO_x, PM, CO₂, CH₄, NMHC, Pb, SO₂, N₂O and NH₃. Greenhouse gases – CO₂, CH₄ and N₂O – represent the same society cost anywhere, while costs for other substances depend on the geographical position.

There are two alternatives for quantifying exhaust emissions: measured exhaust emissions or calculated. For measurements there are again two alternatives: on board or in the laboratory. The laboratory alternative demands use of logged driving patterns. Because of the high complexity and costs of such measurements, calculated emissions are in most cases the only reasonable alternative.

Models for exhaust emissions in general include three parts: cold start emissions, hot engine emissions and evaporative emissions. The following formula is a rough description of an exhaust emission model:

$$\Sigma(\text{Traffic activity}) \times (\text{Emission factor}) = \text{Total emissions}$$

Traffic activity data include at least *mileage* and *engine starts*. Hot emission factors for one vehicle are functions of the driving pattern and vehicle parameters. Cold start emission factors are functions of the engine start temperature, trip length and average speed. Evaporative emissions are to a large extent a function of fuel quality and fuel tank temperature variations.

Models on a micro level, including engine simulation, should in principle be able to describe most ICT functions. This is not the case for models on a macro level in general. Micro models are often used for emission factor estimation and macro models for total emission estimations.

The conclusion about what to include as PIs would then be exhaust emissions or measures with high correlation to exhaust emissions.

5.6.4 Indicators of traffic efficiency

The efficiency of a traffic system can be measured as e.g. traffic flow, speed and density in relation to the optimum levels of these properties given the traffic demand and the physical properties of the road network.

A combination of FOTs and traffic modelling is required to allow estimation of traffic efficiency impacts of the tested technologies. A schematic picture of the proposed methodology is shown in Figure 5.1.

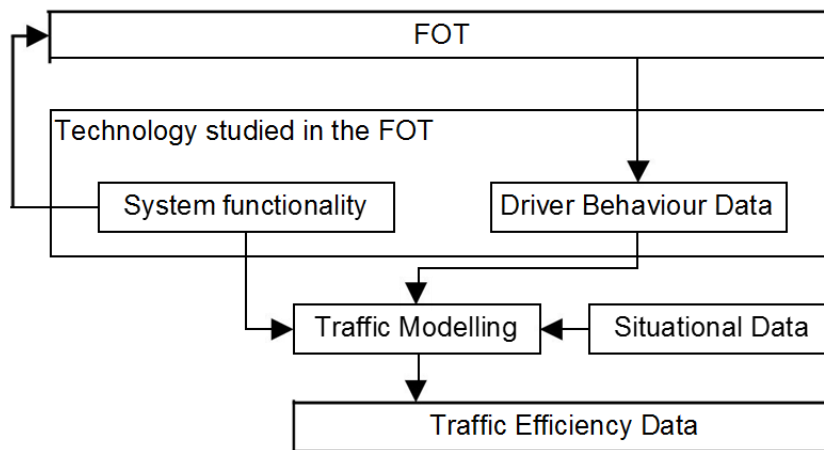


Figure 5.1 FESTA Traffic efficiency estimation based on FOT results

Driver behaviour data is based on data collected in the FOT. This behaviour data will, together with the system functionality of the tested technology, be used as input to traffic modelling in order to aggregate the individual driver/vehicle impact on traffic efficiency effects. This requires that both driver/vehicle data of equipped vehicles and properties of the traffic system that the vehicles have driven in (henceforth referred to as situational variables) are collected in the FOT.

Situational variables are not necessarily directly relevant for PIs or derived measures, but must also be measured or recorded, as they provide key background information that complements the driver behaviour data and is sometimes needed to derive the driver behaviour data. Examples include light conditions and road type.

The appropriate traffic modelling approach will differ depending on which type of driving tasks are supported by the considered technology. Michon's (1985) hierarchical driving model can be applied to select a traffic modelling approach. To model systems that support tactical or operational driving tasks, it is appropriate to apply a traffic microsimulation model. A microsimulation model considers individual vehicles in the traffic stream and models vehicle-vehicle and vehicle-infrastructure interactions. To model systems that support strategic and some types of tactical driving tasks, it is appropriate to apply a traffic simulation model. A mesoscopic model considers individual vehicles but models their movements and interactions with a lower level of detail than microscopic models.

It is advisable to study traffic efficiency for a series of scenarios with varying levels of traffic penetration of the tested systems. The systems should also be studied in representative traffic volumes. This is achieved straightforwardly by running the traffic simulation model with different inputs. The situational data will also contribute to the differences between the scenarios (both measured and modelled).

Outputs from the traffic models will be used to make comparisons of traffic efficiency for the studied scenarios. Example outputs of interest are traditional quality of service and traffic efficiency indicators such as speed, travel time, and queue length.

5.6.5 *Acceptance and trust*

Acceptability indicates the degree of approval of a technology by the users. It depends on whether the technology can satisfy the needs and expectations of its users and potential stakeholders. Within the framework of introducing new technologies, acceptability relates to social and individual aspects as well.

Regarding the dimension of “Acceptance and Trust”, the following subjective PIs should be focused on during FOTs:

Ex-ante usefulness (level of usefulness perceived by the user prior to usage): before using a system, what are the dimensions of usefulness that occur to the future user immediately? What are the benefits he expects from using the system?

Ex-post usefulness (level of usefulness perceived by the user after practice with the system): after first use of a system, what are the user’s impressions regarding the system’s benefits? Ex-post usefulness is to be analysed in relation to the statements of the indicator on “ex-ante usefulness.”

The reactions to both indicators will give useful information for system acceptance. The measurement of these two indicators can be operationalised via self-designed questionnaires, based on established methodological approaches (see Nielsen, 1993; Grudin, 1992). A qualitative approach like a focus group with a formalised protocol and individual in-depth interviews is also appropriate.

The **observed rate of use** of the system or of specific system parts represents an additional indicator for system acceptance and perceived usefulness.

Perceived system consequences (perception of positive or negative consequences of the system's use) is another key indicator for system performance: the user expresses his/her impressions and attitudes regarding the potential consequences when using the system, which can be positive as well as negative. These impressions can best be collected via an interview and be exploited in focus groups, which have the advantage of group dynamics that can provide additional information on the subjective norm. Construction of standardised questionnaires is possible as well (for a methodological background on this indicator, see Featherman and Pavlou, 2003).

Motivation (level of motivation/impetus to use the system) should be connected with the indicator **Behavioural intention** (level of intention to use the system). Both indicators can best be investigated via self-designed questionnaires based on established methodological findings (see Armstrong, 1999; Ajzen and Fishbein, 1980).

The **Response to perceived social control / response to perceived societal expectations** indicates the impact of perceived social control of the user’s behaviour. This indicator is a more sociological one, which should give an indication whether the user feels a social benefit (for example, social recognition) when using the system, or on the contrary, that he/she hesitates to use the system due to fear of social disapproval when using the system (see Castells, 2001).

Usability / level of perceived usability concerns the aspects of the user's general capacity to interact with the system (including installation and maintenance issues, see Grudin, 1992; Shakel & Richardson, 1991). For these indicators, a combination of in-depth interviews, focus groups and self-designed questionnaires based on established methodology is recommended.

5.6.6 Driver characteristics

Even though driver characteristics are not PIs in themselves, they are important as situational variables, which is why they are included in this section. The focus here is on describing the drivers that participate in the study, as compared to selecting drivers based on certain characteristics, which is treated in [Chapter 6](#). Drivers differ on a large variety of characteristics, which may all have an influence on how they drive and use different systems and services. These differences may be important to take into account when planning an FOT. Four categories of driver characteristics may be distinguished:

- Demographic characteristics: gender, age, country, educational level, income, socio-cultural background, life and living situation, etc.
- Driving experience, and driving situation and motivation: experience in years and in mileage, professional, tourist, with or without passengers and children etc.
- Personality traits and physical characteristics: sensation seeking, locus of control, cognitive skills, physical impairments or weaknesses, etc.
- Attitudes and intentions: attitudes towards safety, environment, technology etc.

Studies often focus on characteristics of individual drivers. However, drivers are not alone on the road. There are other road users and there may be passengers in the vehicle, which may influence the driver's behaviour.

There are several reasons for considering driver characteristics:

- To make sure that the sample of drivers is representative of the target population
- To explain the outcomes of the FOT
- To improve systems and services, taking into account differences between drivers

Driver characteristics may play different roles in FOTs:

- Characteristics of drivers possessed before the FOT may play a role in how they behave in traffic during the FOT
- Although some characteristics are stable, others may change when using a system or service in the FOT. Attitudes may change radically before and after using a system for a longer period of time.

In general, it is useful in an FOT to gather as many characteristics of drivers as practically possible. Even if no specific impacts are expected of certain characteristics, some outcomes may be explained better with more knowledge about the participants. A minimum set of data such as age, gender, income group and educational level is easy to gather from participants.

Next, information is needed about driving experience. Usually this is measured by means of self-reports. The amount of practice, i.e. the mileage of an individual driver, can be

collected by asking the subject for an estimation of his/her overall mileage since licensing or the current mileage per year. However, be aware that these self-reports are not very reliable.

For further understanding of driver behaviour, one may consider using questionnaires on attitudes, driving behaviour and personality traits. A well-known questionnaire about (self-reported) driving behaviour is the Driver Behaviour Questionnaire. Some widely used personality tests are the Five Factor Model (FFM) test and the Traffic Locus of Control test (T-LOC). Special attention may be given to the personality trait of sensation seeking, which is correlated with risky driving. The Sensation Seeking Scale (SSS) measures this trait. These questionnaires are available in many different languages, but they are not always standardised, and cultural differences may play a role. Personality traits may be easy to measure, just by administering a short questionnaire. However, the concepts and interrelations of factors are very complex, and results should be treated with caution.

When evaluating the acceptance and use of new systems in the vehicle, drivers' acceptability of technology is important. Both social and practical aspects play a role. Technology acceptance has different dimensions, such as diffusion of technology in the drivers' reference group, the intention of using the technology, and the context of use (both personal and interpersonal). Measuring acceptability can be realised via (existing) standardised questionnaires, in-depth interviews before and after "use" (driving), and focus groups.

5.7 Iteration

When the PIs have been defined, it is recommended to re-check whether these indicators are indeed capable of testing the hypotheses defined earlier. If necessary, the hypotheses or the indicators can be adjusted. Available resources will play a major role in determining which performance indicators to use. It is also necessary to look forward in the FESTA chain, and to consider data storage and analysis. If a large number of PIs have been selected, or if the PIs require a huge amount of data to be collected, considerations about data collection and storage capacity come into play, as well as the question of how to analyse this data. For example, video data requires a large capacity and ample resources to analyse it.

5.8 Performance indicators and user acceptance in CCAM FOTs

The traditional impact areas – safety, mobility, efficiency, environment and costs – are no longer sufficient for determining the full impact of road automation. New societal impact areas and research questions become relevant: land use, public health, socio-economic questions, accessibility, and equity. Industrial automation tests involving e.g. automated trucks raise new efficiency, workplace and stakeholder interests.

The Trilateral Working Group on Automation in Road Transportation (collaboration between US, Japan and EU, but also including other countries) has a subgroup on impact assessment. It developed the [Trilateral Impact Assessment Framework for Automation in Road Transportation](#) (Innamaa et al., 2018).

In addition, the working group compiled Key Performance Indicators (KPIs) that can be used in impact assessment in the following areas:

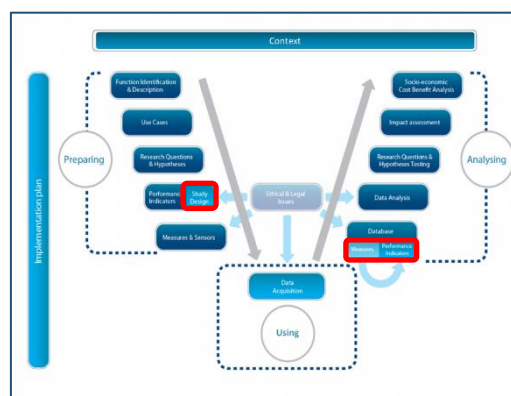
- Vehicle operations / automated vehicles
- Use of automated driving
- Safety
- Energy or environment
- Personal mobility (modal choice, time spent travelling, travelling quality, etc.)
- Travel behaviour (modal share, distribution on routes, etc.)
- Network efficiency
- Asset management (physical and digital infrastructure)
- Costs
- Public health
- Land use
- Economic impacts.

For several of these performance indicators, new types of measures need to be developed, especially for the PIs of more indirect impact areas such as land use and public health.

Special attention needs to be paid to measuring user acceptance. In past FOTs, participants were driving for relatively long periods (sometimes many months) with a system in their car, so they got used to it and were able to form their own judgment on whether the system worked well for them. The users could be surveyed about their trust in the system, how easy it was to use, whether they would buy it themselves, and for how much money, etc. With the first CCAM FOTs and pilots, we usually do not have this same situation. Pilots may have the character of a demonstration where the participants will only be in the vehicle for a short period, and possibly with a safety driver.

6 Experimental Procedures

This section of the handbook provides guidance on the overall experimental design of FOTs in order to ensure experimental rigour and scientific quality. The first section, [Study design](#), provides guidance of the formulation of hypotheses, experimental design and possible confounds. The second section, [Participants](#), provides advice on participant selection, including demographics, driving experience, personality and attitudes, along with consideration of sample size. The third section, [Experimental environment](#), suggests how the road environment (road type, weather conditions etc.) plays a part in the design of an FOT and the subsequent data analysis. In the fourth section, [piloting](#) is described and in the last section the methods of [controlled and semi-controlled testing](#) are explained.



6.1 Study design

6.1.1 Hypothesis formulation

Hypothesis formulation is described in [Section 4.8.4](#).

As a general rule, research practice proceeds in the following way:

1. Formulation of the hypothesis
2. Testing the hypothesis
3. Acceptance or rejection of the hypothesis
4. Replication of the results or (in the case of rejection) refinement of the hypothesis

A hypothesis is a specific statement which can be tested with statistical means by analysing measures and performance indicators (PIs). It is a tentative explanation for certain behaviours, phenomena, or events that will occur. It is essential for an FOT to be designed with clear hypotheses in mind in order to aid the interpretation of the results.

In formulating a hypothesis, consideration should be given to the variables under scrutiny. It is vital that the variables collected in an FOT allow the researcher to accept or reject their hypotheses. To do this, both the independent and dependent variables should be well defined at the start of the FOT. The independent variable is one which can be manipulated by the researcher. As the researcher changes the independent variable, he or she records what happens using dependent variable(s). The resulting value of the dependent variable is caused by and depends on the value of the independent variable. Other variables, known as controlled or constant variables, are those which a researcher wants to remain constant and thus should observe them as carefully as the dependent variables. Most studies have more than one controlled variable.

6.1.2 Experimental design

The two basic types of experimental designs are *within-subjects design* (sometimes referred to as crossed design) and *between-subjects design* (sometimes referred to as nested design). FOTs also need to contain a control condition, in which subjects do not get any treatment. This condition is meant to serve as the baseline: This is how drivers behave in case there is no treatment or no experimental manipulation at all.

Within-subjects design

In a within-subjects design, each subject encounters every level of treatment or experiences all experimental manipulations. For example, in an FOT evaluating navigation systems, every subject drives for some time with (experimental condition) and for some time without (control condition) the system. In this specific case, one half of the subjects would start with the control condition and then switch to the navigation (experimental) condition and half of the subjects would do this vice versa.

This type of design has two advantages: 1) fewer subjects are needed compared to a between-subjects design, and 2) it is more likely to find a significant effect, given the effects are real. The power of a within-subjects design is higher than in a between-subjects design. This is related to the reduction in error variance, since there are no individual differences connected to differences in treatment measures. A disadvantage is the risk for carry-over effects, which means that if a subject experiences one condition, this may affect driving in the other condition.

Between-subjects design

In a between-subjects design, each subject participates in one experimental (or control) condition. The major distinguishing feature is that each subject has a single score (with or without the system). Note that the single score can still consist of driving on various types of roads, during long periods of time or with different types of driving behaviour, workload and comfort.

The advantage here is that carry-over effects are not a problem, as individuals are measured only once in every condition. The total number of subjects needed to discover effects is greater than with within-subjects designs. The more treatments in a between-subjects design, the more subjects are needed altogether. In order to limit the confounding effects due to individual differences in a between-subjects design, one should either use random assignment, in which the assignment of what subject is exposed to what treatment is done randomly, or use matching groups (also called matched pairs), in which one also has to make sure that different groups are comparable with respect to pre-selected characteristics, such as gender and age. In order to do this, one needs to identify the variables that one wants to match across the groups, and measure the matching variable for each participant, and one needs to assign the participants to groups by means of a restricted random assignment to ensure a balance between groups. Also, one needs to keep the variable constant or restrict its range. This will reduce differences within each group and therefore reduce within-treatment variability.

The main drawback with the matched pairs design is in the sampling process. As the number of characteristics that require matching increases, so a correspondingly large sample pool will be required to allow adequate matching. A further problem is that this

design assumes that the researcher actually knows what extraneous factors need to be controlled for, i.e. matched – and in some circumstances this may not always be the case.

Longitudinal and Cross-Sectional Designs

One question an FOT may have to answer is whether an effect of a treatment (e.g. driving with a system) changes over time. To investigate this, longitudinal or cross-sectional designs can be employed. While longitudinal surveys of this type can be very useful, they do not provide an answer to the questions concerning why the changes may or may not have occurred. If things like that are measured in FOTs, one should already have a clear idea why a positive effect may disappear after a while. This could, for instance, be such factors as risk compensation (because the systems warn you, you can drive until you are warned).

One of the difficulties with longitudinal studies is that it is hard to keep subjects motivated throughout the entire study period, or people may move, or become ill. Because of these difficulties, other methods for investigating changes over time have been developed and the cross-sectional design offers an alternative.

The cross-sectional design looks at changes over time by taking a number of cross-sections of the population at the same instant in time. This is obviously quicker and less costly than a longitudinal study, and there is a lower chance of actually 'losing' participants during the run of the experiment. On the other hand, a main drawback of the cross-sectional study is related to the previous experiences of the participants and how this might have an impact on the findings.

Baseline and treatment period

The baseline period is often squeezed into the project and is quite short, especially in relation to the treatment period. Ideally, the two would be equal lengths so that there is the opportunity in the baseline period for the same variations to occur that may occur in the treatment phase (such as seasonal effects, see [Section 6.3.6](#)). The more data available, the more robust the results are.

6.1.3 Threats to validity: confounds and other interfering effects

As a general rule, the results of an empirical study should allow a clear decision as to whether the hypothesised relationships between variables exist or not, i.e. whether the hypotheses can be accepted or have to be rejected. In the best case, the researcher is able to attribute the changes he/she observed at the dependent variable without any doubts as to the manipulation of the independent variable. The internal validity of an experimental or quasi-experimental study describes the extent to which this inference is unequivocally possible because the study has been designed in a way that alternative explanations for the effects are implausible or can be excluded. The internal validity of a study increases to the extent to which such alternative explanations can be ruled out. In the literature, these factors are also described as confounded variables which need to be controlled by appropriate measures right from the beginning of a study.

In the literature, several interfering effects have been described which interfere with the effect of an independent variable on a dependent variable and contribute to a decrease of

internal validity if they are not controlled by measures implemented in the experimental design. The following effects constitute threats to the internal validity of FOTs:

- **History:** Unplanned events unrelated to the study might have an effect on the correlation between independent and dependent variables. For example, during the performance of an FOT an important paragraph of the road code might be changed (e.g. new speed limits for certain road categories), which is accompanied by increased police surveillance activities.
- **Maturation:** Mainly effects due to experience and learning which affect the dependent variable and are (in long-term studies) erroneously attributed to the independent variable.
- **Testing:** If the behaviour of interest is sampled at different times, there might be a biasing effect from the number of times, e.g. from becoming more familiar with the test situation. For FOTs this might become relevant if subjects are tested at different times over the course of the study, but not if their behaviour is sampled continuously and more or less unobtrusively.
- **Selection:** In general, participation in an FOT is voluntary, which means that the strategy of recruiting subjects can have a biasing effect. For example, to offer a certain amount of money (e.g. €500) as compensation for the effort caused by completely finalising the study might be an incentive for participants with a low income, whereas it might insult people with a very high income.
- **Drop-out:** During the run of an FOT one has to take into account that not all subjects will finalise their participation as planned. However, this drop-out can have a biasing effect on the results of an FOT if the subjects who quit early differ systematically from those who finalise as planned with regard to relevant characteristics (e. g. socio-economic status, age, gender etc.).
- **Experimenter bias:** Effects on the dependent variable which result from social interaction between the experimenter and the subjects, which might occur, for example, if at the beginning of an FOT the experimenter explains the system functions very carefully to some subjects due to sympathy, whereas he/she is careless with this in regard to some others.

6.2 Participants

6.2.1 Characteristics

Depending upon the research questions, there is often a need to select a particular group of participants for inclusion in the FOT and ensure that this group is in some way representative of those drivers who will ultimately interact with the system.

The types of variables that should be taken into account include:

- Demographics variables, such as age, gender, social economic variables, and permanent or temporary driver impairments
- Driving experience, in general but also experience with various systems, accident history and the usual time of driving and roads used
- Personality and attitudes.

In [Section 5.6.6](#), more information is given on driver characteristics and measurements.

Recruiting on a personality/attitude basis will ensure that a system is tested on a broad range of drivers who may interact with the system very differently. Recruiting on a personality/attitude basis may be appropriate, since these are likely to influence behaviour directly. Variations in beliefs are likely to explain differences in driver behaviour and system use. Before beginning recruitment for any FOT, researchers must consider the relationship between individual differences and the behaviour which the system is seeking to influence.

In addition to selecting drivers, personality and attitudes can also be used as covariates in analysis in order to identify several differences in driver behaviour and system use between groups. It is not imperative that FOTs base their recruitment on such measures. However, their inclusion within the experimental design provides useful insight into the manner in which individual characteristics influence behavioural adaptation to new systems.

6.2.2 Sample size and power analysis

FOT studies should be able to assess the functionality of the ICT systems and their impact on driver behaviour, traffic safety, environment, etc. When the chosen sample size is too small, it is difficult to statistically prove effects of the system that are actually there. With very large sample sizes the chance of finding an effect increases. However, there are two major drawbacks to using very large sample sizes:

- Every driver/participant needs a car equipped with the system and with a data logging system, which is expensive.
- Small effects which are statistically significant might be found, but they might not be relevant when looking at power effect.

The appropriate sample size for an FOT depends on a number of choices that have to be made in the final setup. These choices are, for instance, the number of ICT systems that are going to be tested and the choice of a between-subjects (two separate groups of drivers with and without an ICT system, but always with data logger) or within-subjects design (each participant drives a certain amount of time with and without the ICT system).

In order to ensure that the chosen sample size is representative for the behaviour of a group of drivers and that it is possible to statistically prove effects that are there, power analysis is needed to calculate the desirable sample size. This power analysis is based on a number of assumptions:

- Suppose an FOT is based on a between-subjects design, such that different groups of drivers each drive with a different system – or at least one group with an ICT system and one group without an ICT system

- The power is 80%, indicating the chance of statistically proving a difference between the groups when it is there (i.e. a chance of 20% of failing to prove it)
- The alpha level is 5% (i.e. the chance of falsely finding a significant effect)
- Two-tailed testing, because there is no reason to assume that either one of the groups performs better/worse than the other.

The effect size is 0.2, which is typical for a small effect that can be expected in an FOT with a lot of disturbing factors compared to more experimental test set-ups. An effect size of 0.5 is typical for a medium size effect. EuroFOT analysis has indicated that it is more effective to increase the number of drivers than to extend the time period of data collection (Jamson et al. 2009).

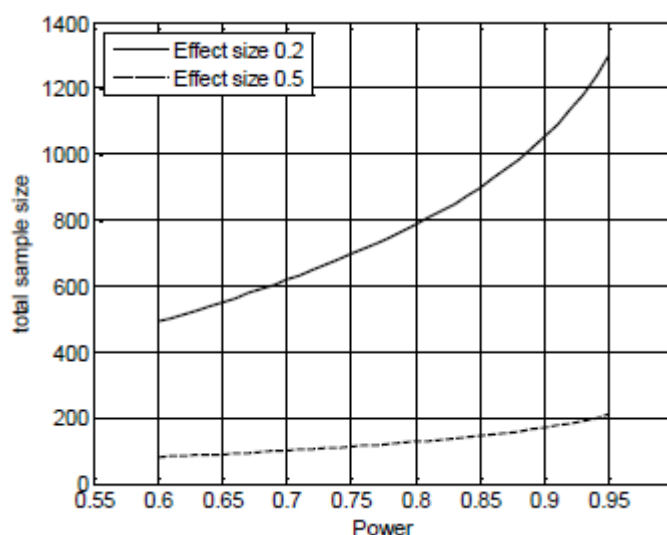


Figure 6.1 Total sample size as a function of the statistical power and the effect size (2-sided test, alpha = 0.05, independent variables)

Figure 6.1 shows that a total sample size of 800 (i.e. two groups of 400) drivers would be needed to be able to statistically prove small size effects between the two groups. The groups are relatively large to compensate for the relatively high number of disturbing factors when trying to find effects in real traffic. If we expect medium size effects, groups of only 75 drivers would be sufficient. If a within-subjects design is chosen, one group of 400 drivers would be sufficient to test both the without- and with-system conditions.

In practice, recruiting the specified target sample may turn out to be very difficult. In order to compare results between countries, it would be ideal to have the same equipment in all countries, as well as the same groups of drivers. This can, however, prove to be difficult because of the penetration rate of different car makes in the vehicle fleet of different countries, and because drivers differ in their 'national driving styles'.

6.3 Experimental environment

The experimental environment is a critical element within an FOT, since it will determine the data that is collected and the ability to fulfil the objectives of the FOT. In general, environmental factors can be treated in several different ways, including:

- Explicitly included in an FOT because there is a particular interest in data connected to that environmental factor (e.g. motorway routes for lane departure warnings)
- Explicitly included in an FOT because these environmental factors are part of the range occurring within a normal driving scenario (e.g. night time driving)
- Measured scientifically so that the data relating to that environmental factor can be included within post-trial data analysis (e.g. vehicle headways)
- Recorded (in varying levels of detail), so that portions of data can be excluded from analysis (e.g. heavy rain, where all or some of the data from a particular day may be discarded; or overtaking manoeuvres where short periods of data within a larger set are discounted during a study of steady following behaviour).

6.3.1 Geographical location

In line with the above, the geographical location can be chosen because it is representative of the intended area of use of a vehicle/system (e.g. predominantly motorway environments). Alternatively, the geographical area can be chosen because it displays the characteristics needed to collect the specific data the study is interested in during the FOT (e.g. the choice of mountainous and/or northern European environments in order to collect data on the use of systems in cold environments).

The population within a particular geographical location may affect the running of the FOT. For example, certain cultural issues, population characteristics, car ownership, use of new technologies, and language issues may be apparent. In addition, the characteristics pertaining to the road and prevailing traffic may be of importance, including:

- Road type and localities present
- Traffic patterns, such as types of journeys (e.g. commuter or tourist travel), traffic flow, traffic density, vehicle types, and frequency and sophistication of journeys
- Other transport options, availability and costs, and inducement or penalties to encourage particular transport mode choices
- Legal regulatory and enforcement environment, such as speed limits, levels of enforcement of traffic regulations (e.g. speed cameras), penalties for traffic or other violations, standardisation (e.g. compliance of road signs with international standards).

The geographical location may also have implications with regards to technical and other study issues, including infrastructure and data communication issues such as:

- Network/beacon infrastructure for vehicle–infrastructure communication
- Network coverage/reliability of telecommunications, especially if automatic over-the-air data transmission is used instead of manual data download
- Localised GPS coverage issues (e.g. urban canyons, foliage cover)
- Logistical issues; both in the validation and experimentation phases, safe and secure access to infrastructure equipment should be ensured for validation of the functions (especially in the case of cooperative systems), for data download (if remote access is not available) and for maintenance. Also, target vehicles should

be accessed for data download (if data is not being transmitted over the air) and for maintenance.

- Availability and quality (resolution, scope and depth of content) of electronic maps that can integrate vehicle location for situation evaluation. Moreover, in case of complex functions and especially for cooperative systems, high accuracy maps may be required in order to implement these functions.
- Availability of other data, e.g. from the police, highway authorities, fleet operators, maintenance personnel.

The most important point in relation to the geographical area is that it must be chosen based specifically on the *objectives* of the particular FOT, and in particular, in relation to the validity of the data being collected. There are two overall considerations:

- Is it needed to consider a particular geographical aspect because it is relevant to the types of vehicles and or systems being studied?
- Does a geographical aspect need to be considered to ensure that the results obtained can be generalised to the wider 'population' of interest (i.e. external validity)?

The starting point is to consider the overall objectives of the FOT, including the types of cars and systems that will be incorporated into the trial. The second major consideration is that of generalisation of the results. In particular, it is necessary to ensure that geographical aspects are included to ensure that the data collected during a specific FOT can be generalised to the wider population of interest. The third factor to consider is whether the geographical factor is of particular interest in terms of data analysis. If it is desirable to analyse results according to the presence or absence of a particular factor, then the geographical environment(s) must include that factor (and possibly a variation thereof).

Finally, it is important to note that the decision to collect data in a specific country might, due to legal requirements in the country, have an impact on how especially personal data could later be collected, handled and shared.

6.3.2 Road type

The road type is the environmental factor that perhaps has the greatest dynamic influence on individual and collective driver behaviour, and hence impacts on safety, mobility, traffic efficiency and the environment within an FOT. It is highly dependent on the geographical area, as discussed above.

The road type will encompass a number of variables which will influence driver use of systems, driver attitudes, driver behaviour, and driver outcomes. The FOT may want to include roads with specific characteristics, including:

- Surfaced or unsurfaced roads
- Minimum, average and maximum speeds of traffic
- Number of lanes and presence of lane marking
- Visibility (of the environment and other traffic)

- Types of manoeuvres that a driver will need to undertake (e.g. stopping at traffic lights or overtaking manoeuvres)
- Typical vehicular headways
- Presence of safety features such as rumble strips or speed cameras.

Three main categories of road should be differentiated:

- Urban
- Rural
- Motorway

Note that road classifications differ in different countries and there is no standard European classification. Ideally, a map and a database of the region of deployment of the FOT should be established in order to reduce the time needed afterwards for collecting this type of data (on the basis of the video recording of the road scene). An electronic map containing at least the type of roads and the speed limits in force (and location of speed cameras) would greatly facilitate the task.

6.3.3 Traffic conditions and interactions with other road users

Traffic conditions and interactions with other road users are important considerations. A distinction needs to be made between:

1. Traffic conditions in a general sense, which characterise a general level of constraints and which, in the same manner as the infrastructure zones, define the driving environment
2. Other road users and their behaviour, which characterise an individual level of interaction between the driver and one or more other road users in the driver's immediate proximity.

The traffic, as a general and contextual entity, can be characterised using several dimensions, for example:

- Density: expressed in terms of the number of vehicles travelling in a given space
- Stability: this can be within a traffic stream (in which case it is expressed in terms of the frequency of speed variations on a traffic lane in a given unit of time) or between different traffic streams (in which case it is expressed in terms of the frequency of lane changes in a given unit of time)
- Speed: the average speed of traffic
- Composition: types of vehicle (light vehicle, heavy vehicle, van, motorcycle) and their relative proportions in a given traffic stream.

The interactions at individual level between the driver and one or more other road users in the immediate vicinity can also be characterised using several dimensions:

- The category to which they belong (light vehicle, heavy vehicle, van, motorcycle, pedestrians)
- Their speed and acceleration (direction and rate)

- Their manoeuvres and behaviour (merging into the subject's lane or pulling out into a lane, merging from an entry slip road, braking, etc.).

Other characteristics to be taken into account are:

- Route choice
- Temporary road/traffic variables
- Traffic encountered
- Impact of road measures on driver behaviour
- Static and dynamic variables associated with the road.

6.3.4 Roads to include

When setting up and running an FOT, it is necessary to consider the extent to which specific road types need to be incorporated into the trial and hence which participants need to be selected. The basic questions to consider are:

- Are specific road types needed to answer the research questions for that sample?
- Would any system of interest be used on a range of different road types?
- Is driver behaviour (in terms of safety, mobility, traffic efficiency and environmental impact) expected to differ according to the road type they are travelling along?
- Is it needed to be able to compare results according to different road types?
- Is it needed to include specific road types in order to generalise the results to a wider population?
- Are interactions with other road users to be included in the analysis? If so, video equipment needs to be installed.

By considering the above questions, one can determine whether a range of different road types are needed, or whether the FOT can concentrate on collecting data based on specific road types. In an FOT, the objective is usually to study the normal driver behaviour. This means that drivers should not be encouraged to change their normal routes.

6.3.5 Weather conditions

Weather conditions are hard to predict, control for, or measure accurately in an FOT. However, weather conditions and associated factors such as ambient lighting are relevant aspects for all FOTs, irrespective of the overall purpose of the study. A well designed FOT must consider a range of weather-related issues, with a view to including, targeting or excluding particular weather conditions. In order to include weather as an experimental variable within analysis, or to specifically include or exclude data for analysis, it is necessary to use a consistent taxonomy and definition of weather conditions.

Related to how weather factors are measured is the level of accuracy that is employed in the measurement of weather factors, including location and time attributes. A further complication with weather factors is that it is often combinations of weather and other dynamic and static factors that have a practical impact on an individual driver or general traffic conditions within an FOT. Extreme weather conditions present a risk to FOTs because they often can't be predicted, and can make journeys impossible, prevent access to vehicles, or in the worst case destroy equipment.

Data may be confounded due to abnormal weather, for example snowfall increasing driver headways and reducing traffic speed or bright sunshine causing glare on screens in vehicles, or momentary distraction to drivers.

There are several ways of potentially measuring weather conditions:

- In real time using direct measurement of the factor, e.g. vehicle sensor to measure ambient temperature (which could then be used to link the use of features to outside temperature)
- Indirect real-time measurement using a surrogate sensor, e.g. recording the use of the windscreen wipers to indicate when it is raining
- Subjective rating scales (completed by the driver or other), e.g. driver assessment of the degree of rainfall
- Post-hoc data mapping – the use of weather records to estimate weather conditions
- Post-hoc analysis of video data by a trained data coder.

At a general level, there are four main considerations with regard to weather:

- Which weather conditions are relevant?
- Should they be 'designed in' or 'designed out' of the study?
- Do weather conditions of interest have a macro (e.g. a rainy day) or micro (e.g. reflected glare) level impact?
- What level of data is needed, and how is this obtained?

6.3.6 Time of day and seasonal effects

Temporal factors, such as time of day, and seasonal effects have a considerable impact on the planning of FOTs and the analysis of data. They can really cause problems for explaining the effects that are found (e.g. whether they are caused by the system under test or by seasonal circumstances). In contrast to the weather effects outlined above, the temporal factors can usually be predicted, and so it is usually easier to deal with the issues successfully. The main issues that have to do with the time of day, week, and seasonal variations are:

- Influence on driver state (e.g. sleepiness)
- Disruption caused by external events, e.g. school opening times
- Influence on traffic levels
- Other temporal influences on traffic
- Impact on vehicle occupants
- Glare
- Ambient light levels
- Seasonal confounding of data collection
- Influence on route choice
- Pragmatics to do with drivers' work and life schedules
- Using time of day as a surrogate; for example, time of day can be used to specify or control for traffic levels or ambient light levels.

Time of day and seasonal effects are different to weather issues in several ways, including:

- Time of day and seasonal effects are much more predictable than weather conditions
- They are often proxies, i.e. not important in themselves, but important because they result in variation of a factor of interest (e.g. traffic levels, or level of the sun above the horizon).

These two factors mean that greater emphasis should be placed on planning around relatively predictable time of day and seasonal effects, and considering their impact on the FOT. There are different ways to (partly) deal with seasonal effects: have a control group, or adjust the length of the test. The latter means that either there is a short time period for the FOT, so that baseline and treatment phases take place in the same season, or that the FOT is very long (more than a year), so that baseline and treatment phases include the same seasons.

6.4 Conducting a pilot study to test the evaluation process

A pilot study can be defined as a “small scale version, or trial run, done in preparation for the major study” (Polit et al., 2001); it goes before large-scale quantitative research and is very useful to test the research instruments, identify any performance problems and ensure a reasonable durability of the technology instruments adopted. Conducting a pilot study is a fundamental phase to get warning in advance about practical problems or difficulties that may affect the study, and it is also necessary to prepare the deployment of the FOT and to support the design of the relevant tools for the evaluation process. This task should be performed early in the evaluation process. It represents an important step for mobilisation and dialogue between the various teams involved in the FOT and for promoting a common framework and consensus for the evaluation process.

The relevance of conducting a pilot study and the time required are often underestimated. To better understand the importance of this step, below is a list of general reasons for conducting a pilot study (for a wider overview, see Polit et al., 2001):

- Developing and testing the adequacy of research instruments
- Assessing the feasibility of the full-scale study
- Testing the research protocol
- Testing whether the sampling frame and technique are effective
- Verifying the likely success of the proposed recruitment approach
- Identifying logistical problems which might occur using the proposed methods
- Testing variability in outcome to help determining sample size
- Collecting preliminary data
- Verifying what resources (finance, staff) are needed for a planned study
- Verifying the proposed data analysis techniques to uncover potential problems
- Testing the research questions and research plan
- Training the researchers, both in data analysis and in personal integrity issues.

Going more into detail, in FOTs these preliminary field tests have to deal with three main levels of analysis with specific objectives:

1. The first preliminary field tests have to check *the technical functioning of the data collection systems in real driving situations*. They should enable to identify potential problems of sensor calibration or drift and thus to establish the periodicity of maintenance procedures during the FOT. They should also permit to validate the data collection procedure from data acquisition and data transmission to data storage.

The technical teams involved in the FOT should be in charge of these field tests.

2. The second level of a preliminary field test deals mainly with the issue of *assessing the usability and usage of the systems under study and of identifying the main critical issues associated with their use in real driving situations*. This is particularly relevant for:

- Structuring the familiarisation phase of the drivers before their participation in the FOT
- Contributing to the design of the questionnaires for the subjective assessment of the systems
- Testing and/or improving the various tools developed for data processing, such as automatic identification of critical use cases and scenarios and video-based identification of triggering events or categorisation of road and traffic contexts
- Identifying a number of critical scenarios when using the systems – scenarios that could be investigated more extensively when the data gathered from the FOT is processed and analysed.

This test requires the participation of a sufficient number of drivers (depending on the target population in the FOT) and should be performed in real driving situations. An experimental journey on the road could be designed for that purpose (depending on the hypotheses formulated). This level of analysis provides useful data for designing the relevant tools for the evaluation process as mentioned above, for estimating the time required for data processing and data analysis and thus calibrating these phases in the FOT. It may be seen also as an opportunity for training the team(s) in charge of data processing. Finally, it represents an important step for testing some of the hypotheses formulated in the FOT and/or for refining them. In this phase, it's important to underline that the drivers used in the pilot study will not be part of the final sample and therefore most of them do not need to be naïve.

Psychologists, ergonomists, and human factors experts should perform these tests in close cooperation with the team in charge of statistical analysis as well as the team in charge of developing data processing tools.

To test whether what is asked from participants is realistic, it is a good idea to pilot yourself before letting 'real' participants undergo the testing. Let someone (or several persons) from the project team drive in an FOT vehicle, answer the

questionnaires, fill in the travel diaries, etc. This is especially relevant for people working on subjective data collection and analysis.

3. The third level consists of *testing the feasibility of the overall evaluation process* from the selection of the participants through to data collection and evaluation. It is a kind of final rehearsal before the deployment of the FOT. It enables in particular a check of the communication process between the various teams involved in the practical deployment of the FOT, of the robustness of the technical tools designed for data collection and transmission, and of the robustness of the evaluation tools used in the assessment.

The result of the pilot can be a no-go if too many problems are still present. In this case, it could be reasonable to delay the start of the data collection phase and to repeat some earlier steps. This means that there are feedback loops in the piloting process.

6.5 Controlled testing

As described in [Section 6.2.2](#), a power analysis is required to determine the necessary sample size for conducting an FOT. The estimated or simulated frequency of events and the penetration rate are a key element in this calculation. It might prove that a naturalistic FOT is not feasible, due to a low frequency of events resulting in a very high number of needed vehicles or a very long experimental period.

In such cases, one possible option is to allow controlled or semi-controlled testing. This means that all or a certain group of drivers are instructed before or during the test execution to behave in a certain manner. For instance, a professional driver might be instructed to simulate a car breakdown in order to trigger the car breakdown warning function in passing (uncontrolled) vehicles. In the controlled approach, the test drivers are called into the test and asked to drive the test route with some arrangements. Preferably, the tests will be conducted in real traffic. Some tests, however, must probably be organised on a closed test track. One test may include several runs of the route. Several situational variables can be fixed in advance. The tests can be designed so that some variables are systematically controlled during data collection. Based on the practical constraints, different levels of control can be chosen from totally naturalistic to totally controlled, taking into account that controlled testing breaks with the principle of un-interfered experiments and should be chosen only if the FOT boundary conditions and/or power analysis do not allow a naturalistic test of the function under test. Controlled testing can also be used as a supplement to naturalistic FOTs.

Table 6.1 provides an overview of differences between controlled tests and naturalistic driving studies (based on the classification from the [DRIVE C2X](#) project)

Table 6.1 Complementary uses of naturalistic and controlled tests in cooperative system evaluation

Comparison criteria	Naturalistic studies	Controlled tests
General versus experimental design	Normal day-to-day driving. Data collected continuously. Usually the same drivers for a long time (or full study). Do not bias drivers but get their natural response and acceptance. In a cooperative setting, really evaluate if vehicle interaction is enough for statistical significance.	Controlling the exact studied scenarios and interaction between vehicles. Easier to have different user groups (old, young etc.) but for shorter times. Necessary for function and technical evaluation. Controlled number of tests gives easier statistical control.
Acceptance, long term	Yes, but care should be taken not to influence by administering many questionnaires throughout the study.	Not possible. Short study time.
Acceptance, short term	Possible, but takes calendar time since interactions cannot be forced.	Yes, but care should be taken to limit too many repeated interactions over a short time. May give unpredictable and unreliable results.
Impact on environment	Yes, as long as enough system interactions happen. Compliance with system possible to study.	Yes, but assumes driver compliance.
Impact on safety	Yes. Will likely result in a wide variety of situations. Possible to compare treatment/baseline for e.g. crash-relevant events. Naturalistic distraction and TTC distributions (including compliance).	Yes, but difficult to cover crash-relevant events, distraction, and compliance.
Impact on efficiency	Yes, a variety of situations and compliance can be included. Need a large baseline to compare with.	Specific situations may give a good statistical base.
Impact on mobility	Yes, but long term analysis/visibility is needed, plus enough interactions all along. Questionnaires relevant.	Difficult since compliance can be very high if drivers are told to perform tasks. Questionnaires relevant.
Driver behaviour: route choice	Learning effect can be studied. Difficult to know if route choice is an option. Indirect effects.	Difficult since drivers will do what they are told.
HMI	Short and long term HMI usage and acceptance	Short term HMI usage and acceptance. Easier to

Comparison criteria	Naturalistic studies	Controlled tests
	possible. Evaluation if reaction time changes over time. Time on task possible if video is available.	evaluate in a controlled environment but difficult in the long term. Necessary before naturalistic deployment.
Usage (function)	Are users turning it off over time? Compliance with system information possible to study.	Difficult
Technical evaluation of wireless communication	Study of robustness possible, but in-depth analysis difficult due to uncontrolled scenarios.	Necessary to study in depth communication aspects.
Function validation	Long-term functionality evaluation possible but impractical for optimisation and technical debugging. It may give optimisation parameters in the long term and give information about function validity for real traffic scenarios.	Necessary to validate and optimise functions to technical boundary conditions. It can give fast and reliable results that can be quickly followed up.

6.5.1 Operationalisation of tests

Controlled testing requires a strict operationalisation process from high-level hypotheses down to individual tests to be performed. A three-step process is advised:

In controlled tests, all drivers are instructed to follow a defined *test scenario*. This scenario is created from the hypothesis defined and tries to provoke a system behaviour, which causes the activation of the function to gather data needed to prove or disprove the hypothesis. The scenario should therefore contain:

- Functions addressed
- Hypothesis addressed
- Description of desired situation
- List of desired participant types
- List of desired vehicle types
- List of vehicle groups (e.g. one group as broken down vehicle, one group for passing vehicles).

It is sufficient if scenarios are described in non-formal text. However, it might be advisable to use a pre-defined scheme to describe them. To follow up in the operationalisation, the scenarios have to be further refined into test scripts. A *test script* builds upon one scenario and maps it to a given area and a given project setup. To generate this test script, each group has to be mapped onto the road network of the test site. A route is created which

defines for each group where exactly the vehicles will drive and what timing is desired/expected.

A baseline can be created by assigning a separate control group to the test script with systems switched off.

This test script therefore contains:

- One route for each participant group with timing information (including individual vehicle timing offset)
- A desired state for the functions to test
- A desired state for the logging and monitoring systems.

In a final step, the test script is turned into a *test case* before actually starting the test. This is where the actual drivers and vehicles are assigned to the groups. Also a date and time for the test case are fixed. One test script might be scheduled several times as a test case to gather enough qualified data to filter out outliers in the execution. Drivers and vehicles may change for different test cases of the same test script. Figure 6.2 is an example of this process.

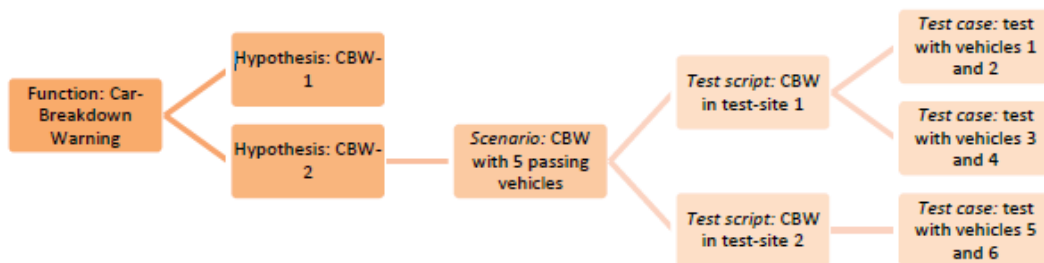


Figure 6.2 Operationalisation of test scenarios

6.5.2 Operationalisation tool chain

In larger FOTs a dedicated set of tools is highly advised for the operationalisation process.

In a scenario editor tool, all scenarios can be entered in pre-defined fields. These map to the textual information needed to describe the scenario, but also define formal aspects, such as desired number of test iterations or whether a pre-validation in simulation is necessary. It should also list the necessary PIs.

The script editor tool is map based. It loads scenarios and maps the implicit information on what should happen to explicit routes in one specific location. For each of the driving groups one route needs to be created. Intelligent mapping tools allow to use the underlying road network data on the map (e.g. OSM, GMaps) to automatically follow the street. To get a first idea of how the script will perform, a real-time minimal simulation can be used to see virtual vehicles move along the defined routes. Thus, synchronisation between groups can be reached to successfully create the desired situations.

The script editor tool also creates log profiles to be taken during the test, based on the PIs contained in the scenarios. For this process the measures needed for all PIs are merged. Sophisticated scripts can also contain time-bound or location-bound markers, which are executed in the vehicles once they pass the given point. These markers are used by the test system, for instance, to trigger:

- A change of log profile (e.g. extended logging, when entering the test area)
- A driver instruction
- Activation or deactivation of functions (e.g. for a control group)
- A synthetic function behaviour (e.g. turn on the broken-down vehicle warning).

In a final step of operationalisation, the test script has to be mapped to the current test site situation. A test case is generated from the test script by allocating available vehicles and drivers to the groups shortly before starting the test. This should not be done in advance, since fluctuations in vehicle pool and drivers are to be expected for larger fleets. The tool chain can support this with a dedicated control connection to the vehicles.

6.5.3 Test execution

In theory, a controlled test can run unsupervised. In practice, controlled tests need live supervision to have an acceptable success rate. (Note that a test is determined to be successful if the desired scenario was created, not necessarily if the function was triggered).

The supervision of a controlled test is preferably managed with a *test control* tool. This tool displays in real time the status of all participating vehicles (*monitoring*) and the selected test case. Thus, the operator can monitor test progress and determine deviations from the original script.

A way to directly interact with test drivers is desired. Using the same connection as the monitoring data, the test control tool can send messages back to the vehicles. These messages can contain:

- Textual instructions to drivers (to be displayed on HMI)
- Voice instructions to drivers
- Scenario script and test case information, e.g. test name, schedule, route information
- Trigger information for the test system itself (log profile changes)
- Trigger information for the system under test.

It has to be decided whether driver instructions are necessary for the FOT. If so, they can be either displayed on the system HMI or on a dedicated device.

6.6 Naturalistic Driving Study

6.6.1 Definition

The opposite to controlled testing is the Naturalistic Driving Study ([NDS](#)) or observation, a research method using advanced technology for in-vehicle unobtrusive recording of driver

(or rider) behaviour during ordinary driving in everyday traffic situations. This method yields unprecedented knowledge primarily related to road safety, but also to environmentally friendly driving/riding and to traffic management. A central focus of NDS is to understand explanatory factors associated with crashes and predict involvement in crashes. The naturalistic method “refers to a method of observation that captures driver behaviour in a way that does not interfere with the various influences that govern those behaviours” (Boyle et al., 2009). NDS are defined as “those undertaken using unobtrusive observation or with observation taking place in a natural setting” (Dingus et al., 2006). The following characteristics could define the naturalistic driving approach (Sagberg et al., 2011):

- Unobtrusive recording of driver and vehicle parameters. Normal driving, i.e. driving purpose and driving destinations as defined by the driver, and driving taking place on roads open to ordinary traffic, and with the vehicle that the driver normally uses (owned, leased, or company vehicle)
- No observer present in the vehicle.

6.6.2 Relation to FOTs

Naturalistic Driving Studies tend to focus on crash-explanatory factors, and FOTs generally focus on evaluation of systems or functions. However, the collected data in both types of study can be used for many alternative purposes such as analysis of environment, efficiency and mobility impacts. NDS and FOTs are preferably seen as different methods because (a) the study design is different (participant selection, experimental conditions, vehicle sample, etc.), and (b) the research questions and hypotheses are different. In particular, the main difference relates to the degree of experimental control found in the study, as illustrated in Figure 6.3. It is recognised that NDS and FOTs have some common methodological aspects and that there are gradual transitions between them.

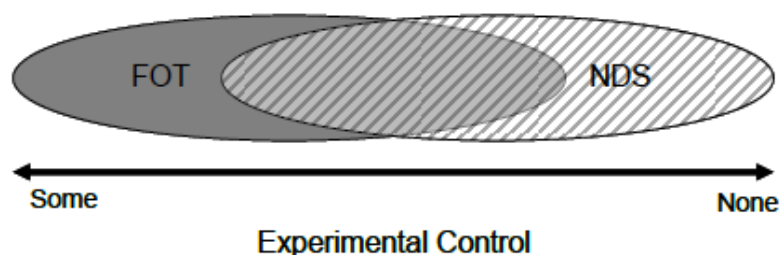


Figure 6.3 A partially overlapping relationship between FOTs and NDS along a continuum of experimental control.

A more generic methodological approach is emerging whereby naturalistic data collected by either NDS or FOTs can be used for similar purposes. For example, an NDS may be used to evaluate the impact of on-market intelligent safety functions (Antin et al., 2011) and an FOT may be used to study crash causation (Olson et al., 2009). Even though there may be differences in purpose between NDS and FOT studies, the technology for driver behaviour observation may be the same, and consequently experiences from FOTs have been important inputs to the planning of a large-scale European NDS (Sagberg et al. 2011). Between the FOT and the NDS lies the Naturalistic FOT. The Naturalistic Field Operational Test (N-FOT) is defined as a study undertaken using unobtrusive observation in a natural setting, typically to evaluate the relationship between (permanent or temporary)

driver-, vehicle-, or environmental factors with crash risk, driving behaviour, and countermeasure effectiveness (Victor et al. 2010). This definition accommodates for both accident-oriented research of explanatory factors associated with crashes (common in NDS), as well as for the evaluation and development-oriented research on new technology and solutions (common in FOTs). N-FOT studies can, for example, assess the relationship of an in-vehicle system (dynamic vehicle factor) or age (static driver factor) or distraction (dynamic driver factor) or speed cameras (static environment factor) with crash-risk, driving behaviour and/or countermeasure effectiveness. Environment sensing and video are believed to be essential for identifying near collisions and other incidents, and for validating that intelligent vehicle systems (e.g. collision warning, lane departure warning and intelligent speed adaptation) perform as expected.

6.6.3 NDS and the FESTA V

Although FESTA V was originally developed as an implementation planning tool for FOT studies, it is a highly relevant and useful tool also for NDS. However, some modifications are required. These modifications are primarily related to the shift in focus away from function assessment in FOTs towards *behaviour* assessment in NDS. Broadly speaking, the bottom or tip of FESTA V is most relevant for NDS and the top of FESTA V is less directly relevant.

The horizontal context bar is also useful for NDS; it is recommended that activities that deal with the more general aspects of an NDS and with high level aggregation of the results take place within this horizontal context bar. According to Sagberg et al. (2011) this type of activity includes:

- User/stakeholder identification
- Topics selection
- Dissemination
- Identification of constraints, such as available technologies and budgets.

FESTA V has two “scaling-up” steps: impact assessment and socioeconomic cost–benefit analysis. NDS also have scaling-up activities, such as aggregation of research question results, and analysis of their implications. For example:

- Provide an integrated overview of the NDS findings
- Use findings to identify new and more efficient safety and sustainability measures related to vehicles, drivers and road infrastructure (such as incentives to encourage adoption of new technologies, education of drivers, regulation enforcement)
- Identify ways for measures and tools to improve safety and sustainability of road transport in Europe based on NDS results
- Demonstrate how naturalistic driving can be used in industrial development of safety and sustainability functions and services.

For further detailed examples of how FESTA V can be interpreted for NDS purposes, see Sagberg et al. (2011) and Victor et al. (2010).

6.7 Documentation

As exposed above, study design is a complex process taking into account a large number of parameters to optimise data collection for a study's specific goals: by their choices, study designers can have different levels of control over experimental design, drivers' sample selection, data collection period and environment. All those preliminary choices may have a large impact on what can be deduced or concluded from a dataset. It is therefore paramount, especially in the optic of data reuse, for any analyst to consider motivations behind a specific data collection, and be informed of possible limitations and biases. Additionally, actual conditions of data collection may largely differ from the initial planning: choices of a specific vehicle brand for technical or organisational reasons may bias the driver sample towards a specific population, delays in planning may lead to different weather conditions, day-to-day operation may lead to different organisation of baseline and treatment periods. As a result, not only should initial plans and their motivations be thoroughly documented, but also *real* conditions of data collection, real driver sample characteristics etc. have to be written, both for initial analysis and for future data reuse.

6.8 CCAM FOTs

The experimental procedures for evaluating CCAM will often be restricted by the technical status and maturity of the automated systems and by the regulatory permissions. As long as it is not feasible to conduct large-scale naturalistic FOTs with automated vehicles, study design may take the form of controlled testing, as described in [Section 6.5](#) on controlled testing.

What is specific for user acceptance testing of CCAM is the large gap between expectations and promises of the vehicle automation, and the actual stage of development. Furthermore, vehicle automation impacts many areas of the future mobility, as well as the liveability of cities and towns. Users are asked to project themselves into the future, they are not only asked to think whether this is a useful and convenient product for them, but to imagine how their daily life and their environment could change.

Although real user tests can be performed, such as with automated shuttles providing normal services, testing automated systems may encounter issues with having artificial scenarios, safety drivers, low speeds, user interfaces that are not fully developed or not easy to understand, and a short time for users in the vehicle. Providing a good user experience by managing expectations and using additional techniques such as videos, storyboards, and focus groups discussing the experiences, may help to elicit valuable information from users.

Specific attention could be paid to the role of safety drivers, if they are used in the study. Their role and their expected behaviour should be specified in detail, in order to avoid over-influencing participants. Safety drivers may also be a good source of information, as they are the ones who may be exposed to driving with the system for long periods. As they usually have a lot of experience in (conventional) driving, and additional training, they may also have a good insight in the differences between automated and non-automated driving. Their opinions and experiences could provide insights that can shape new study designs.

7 Guidelines for Data Acquisition

This section aims to provide guidelines and recommendations for how to handle data in an FOT study. Data acquisition, data storage, and data analysis tools will be covered here.

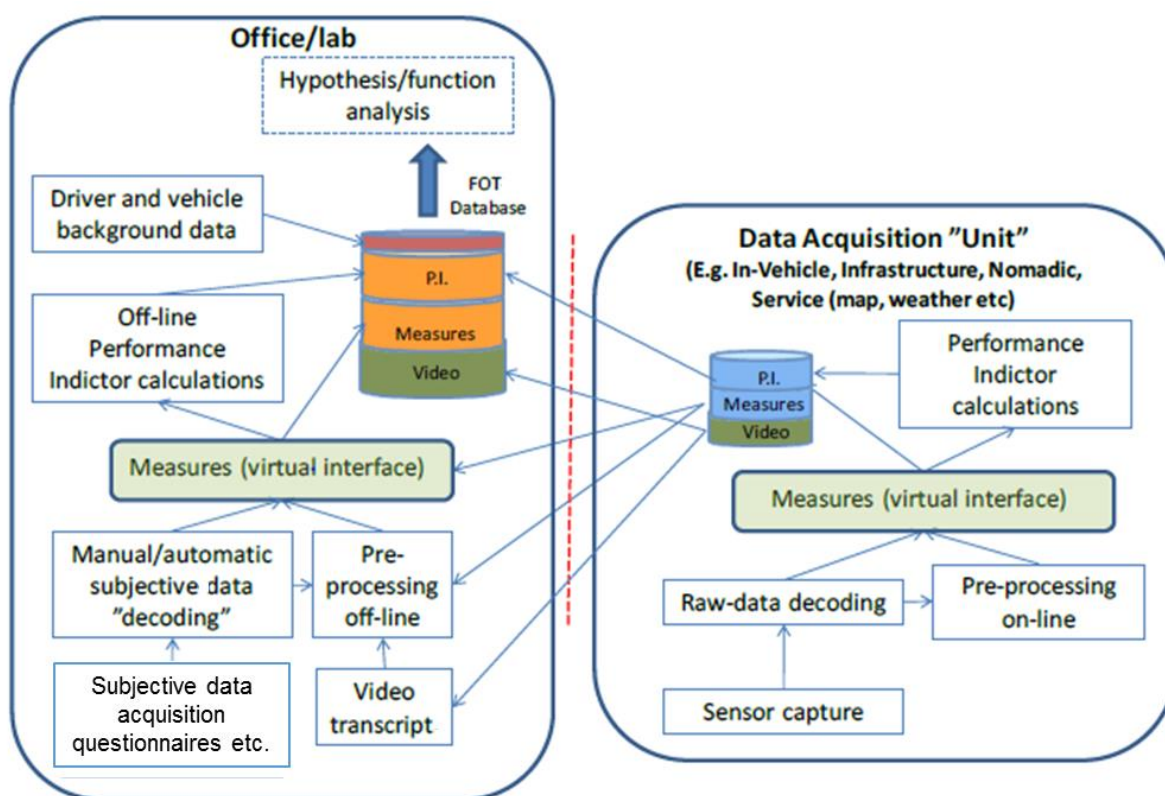
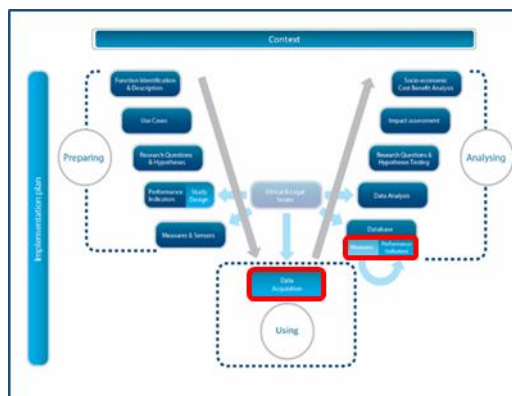


Figure 7.1 Data structuring

Please refer to Figure 7.1 for an overview of a data handling structure for an FOT, and for the naming conventions used in this document. The example data structure above includes data from an electronic data acquisition system, as well as collected subjective data. The Data Acquisition Unit (on the right) comprises sensor systems requiring raw data decoding. The raw data may then be pre-processed, in this case by low-level data processing such as simple filtering or calculation of directly derived results. Both raw data and pre-processed data (derived from raw data) are stored in the same format locally and can be kept locally, until moved from the Data Acquisition System (DAS) to the main storage, and used for analysis.

Acquisition of subjective data may also be performed. *Subjective data is also considered as acquired from a sensor.* This data can similarly be subject to manual or automatic

decoding, database storage (pre-processed or not), and subsequent use in performance indicator calculations.

It is important to create documentation of developed tools and processes in order to enable their usage not only after the project's completion but also within the project. Large-scale projects with long durations tend to be very dynamic, some people may be replaced, others might be assigned different tasks, and therefore proper documentation saves a lot of time for everyone involved. It is also advised to provide detailed test protocols with all relevant test scenarios and potential risks.

7.1 Measures and sensors

FESTA recommends use of a spreadsheet or similar documentation for defining data needs, performance indicators, measures and sensors. Properly handled and thoroughly implemented, the tables are valuable tools for data structuring and for data requirements specifications, and for identification of connections between sensors, measures and PIs.

The data needs are first listed for project-wide discussions of scope and installations. At a later stage, this information will be used in defining data processing steps and, for example, a relational database.

7.2 Data acquisition

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

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

Before the collection of any type of data, considerations on privacy issues are required (see [Section 7.13](#) and the [Data Sharing Framework](#) (FOT-Net Data and CARTRE, 2019)). This has to be done at all sites collecting data if the sites are located in different countries and hence subject to different legislation.

7.2.1 Background data acquisition

Background data about the driver is crucial and needs to be collected integrated in the driver interaction procedure. Due to privacy issues, different parts of the background data may or may not be suitable for storage in a database, or to be used in statistical and other forms of analysis.

Data could be gathered by interviews and/or questionnaires, by different tests, or by specific instruments. The driver background information should be considered as acquired

from a sensor, and preferably be added to the database and to the sensor matrix, though subject to privacy issues.

7.2.2 In-vehicle data acquisition

An in-vehicle DAS is needed in FOTs where the focus is to study in-vehicle systems by collecting data from the systems in the vehicle and also from sensors mounted on the vehicle such as video cameras. A suitable DAS can differ from study to study and a specific solution cannot be recommended for all types of FOTs.

Choosing a widely-used DAS product or using existing vehicle data loggers can save considerable amounts of project time, when comparing against starting data logger development from prototype components. However, as FOTs often test the very latest vehicle features, including elements such as new communication or environmental sensing, data collection has required tailoring and additions more often than not. Where projects lasting many years may afford a data logging development phase running together with finalisation of tested systems, shorter projects need to build on existing DAS products. As a few past FOTs have experienced considerable delays finalising their data collection equipment and processes, FESTA recommends to start data logging selection process as early as the initial data needs have been collected from evaluation leaders. Further, the initial data needs should be gathered early on in an FOT, to start scoping. Final data needs are often a result of long negotiations: how much effort collecting certain type of data would require and what are the evaluation alternatives and priorities.

If the FOT chooses to use various loggers, it has two options regarding post-processing of collected data, to support efficient calculation of indicators, detection of events and other evaluation-related calculation phases:

1. Either the data from different loggers needs to be converted to a defined format, that may have logger-specific extensions
2. The software used for evaluation needs to be adapted to read in data from different loggers and produce harmonized indicators, as far as possible.

Both options have benefits, so the selection depends on FOT type: how different are the tests and loggers, how many partners have resources for indicators calculation, confidentiality of raw data vs post-processed data etc.

7.2.3 Nomadic devices

A nomadic device (ND) or an aftermarket device could be either part of the function/system under test, or it could be part of the data acquisition system, acquiring specific FOT data.

NDs can also be used as data storage tools, as they are easy to install and use on different kind of vehicles. If the vehicle has a dedicated gateway for ND, this option can be used for capture of further vehicle related data.

Using local wireless connections, the storage capacity of an ND could be extended with large capacity hard disks. A possible drawback of an ND, when used as a DAS in itself, is

that test subjects must remember to bring the ND to the vehicle every time they use the vehicle.

7.2.4 Subjective data acquisition

As explained before, subjective data is also considered “sensor” data within the scope of the FOT methodology. All subjective data should therefore be stored and handled logically as if it were collected from a “real” sensor. Subjective data may include data acquired from the test participants in different ways; results from interviews, questionnaires and focus groups. In evaluations of cooperative systems and vehicle fleets, consideration should be given to collecting subjective data also from actors such as control room operators.

The result from the subjective data acquisition should preferably be stored in an electronic format. Electronic compilation of the questionnaire may be considered to reduce the overall manual work and cost, maybe by using web-based tools.

Human-entered survey data usually requires checking quality and completeness before using the inputs to calculate user group averages. Some of the quality-checking can be avoided by careful definition of electric questionnaires, defining acceptable and plausible value ranges already in the survey tool.

For subjective data to be stored and linked with other test data, the following reference information is usually required:

- Subject ID code
- Date and time (hh:mm) of test start
- Date and time (hh:mm) of test end
- If present, reference to objective data (file name, location).

It is important to think about processes and tools for gathering questionnaires at the beginning of the project, e.g. provide smartphone apps to fill out questionnaires.

7.2.5 Real-time observation

In this context, real time observation data is data collected by an observer who directly or indirectly (in real time or afterwards, for example on video) is observing the drivers and systems to be evaluated. The data acquisition process is usually relatively manual, but the results should be transferred to digital format and uploaded to the FOT database for further analysis. Even when using a professional observer, but especially when using several observers, a quality check for the final dataset is important.

Real-time observation data helps provide a more detailed picture of a driver’s behaviour, as well as verifying the information gathered by other instruments. As the overall purpose of an FOT is to collect information on as natural driving as possible, the real-time observation is almost exclusively done through video cameras hidden in the vehicle compartment, as an observer physically in the car will impose the risk of the driver not acting the way he or she would otherwise. Direct real-time observations must therefore be carried out with great care and as unobtrusively as possible, or avoided completely.

7.2.6 Additional data sources in cooperative systems

The cooperative systems architecture implies the possibility of additional data sources. Specifically, RSUs connected to proper sensors may provide traffic information and environment data. RSUs connected to a traffic light controller are able to provide traffic light phases and intelligent traffic control centre dispatch traffic information and alternative routes. It is important that all data records contain a timestamp synchronised to a GNSS clock.

7.2.7 Acquisition of infrastructure data and other services

General aspects

The infrastructure can be equipped with sensors to detect e.g. traffic or weather conditions. Data from such systems can be collected in raw format or in an aggregated form. If data is collected both on the vehicles and on the infrastructure separately, it is necessary to synchronise the two sets of data. It is recommended that GPS time is used as the synchronisation source.

Infrastructure

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

Services

When using additional sources or external services, it is recommended for traceability (during and after the project) to record information about e.g. version of service, update rates and resolution/precision of the information they have during the duration of the study. It is also recommended to invite the service providers for discussions and possibly partnership in the FOT.

7.3 Specific sensors

7.3.1 In-vehicle video

Most state-of-the-art FOT experts consider video as a very important data source for the identification of driver behaviour and reactions, as well as for the process of analysis to understand the underlying context with regards to the surroundings. When a certain situation or event has been identified for evaluation of a particular system, the video provides the analyst with information about the context of both driver behavioural aspects and the interaction with the environment (if external video is used). Even when events are detected automatically using data from various sensors, video can offer confirmation of the event for an analyst, or it can provide important material for dissemination of results.

Video can be used in mainly two ways:

- Driver monitoring: Firstly, driver eye/head movements in relation to the vehicle/environment/context, and secondly driver interaction with the vehicle and

- other driver actions (pedal, gear shift, steering wheel handling, mobile phone, eating, etc.)
- Environment/contextual monitoring: Helps understand the driving contexts by collecting information about the surrounding traffic.

For a specific FOT, the number of cameras needed, the needed resolution, the views captured by the cameras, etc. should be *defined by what is needed to address the hypotheses*. A thorough investigation is needed to find the correct requirements: predict storage needs, evaluate image quality, and set requirements accordingly. Experts from both the study design/analysis planning phase and from the analysis team should be included.

Generally, the following parameters affect video quality and need to be considered: picture resolution, frame rate, colour settings, “regions of interest”, bit-rate strategy, bit-rate limitations, “quality” strategy settings, and other video compression method-related settings.

Recommendations

Make a thorough analysis of video acquisition requirements. Choose a well-tested hardware video frame grabber/compression solution, and select compression suitable for the FOT.

Video data analysis is normally complex and time consuming when done manually. For this reason, it is recommended to try as far as possible to use automated processes.

7.3.2 Internal vehicle bus data

Most vehicle manufacturers feature one or several internal vehicle networks such as CAN, LIN, MOST, FlexRay or Ethernet. An internal network may carry large amounts of useful information for the FOT. However, the detailed message format may be confidential, whereas evaluation usually requires collecting only the basic information carried over the network, such as timestamped vehicle speed.

OEM cooperation

Accessing information from an internal vehicle bus can be highly complex and even void warranty if it is done without the OEM's permission and supervision.

A description of the entire vehicle network will often contain proprietary information, and may reveal detailed information on specific functions and the vehicle system architecture. Thus, non-disclosure agreements are typically required and can be hard to attain.

An option to using the entire vehicle bus description is that the OEM only provides the description of a selection of signals which enable access to the most important data. Still, however limited the access to the vehicle bus is, there may still be proprietary issues.

By using a bus gateway, the OEM will be able to extract data from the bus without providing any information about how the bus actually works. The gateway is programmed by the

OEM to read certain information, decode it and then pass it on to the FOT logger equipment. An illustration of the process is shown in Figure 7.2.

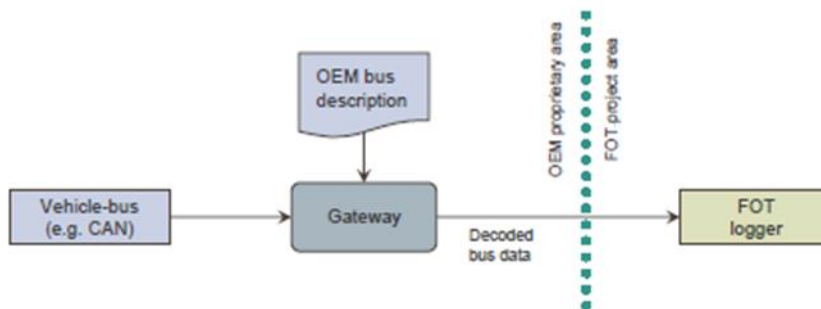


Figure 7.2 Illustration of process for securing proprietary vehicle-bus data using a gateway

When physically connecting to the vehicle bus, it is important to follow applicable standards of the bus in order to prevent interference that may reduce network functionality (and thus warranty). It should also be stressed that every bus implemented by an OEM might not necessarily fulfil the standards in every detail.

7.3.3 Automatic in-vehicle driver monitoring

Driver monitoring systems use both direct and indirect methods. Indirect methods collect inputs from the vehicle and driving style, and derive the driver's state based on these measurements. Direct methods include e.g. monitoring the driver by using a head/eye tracker. Some examples of direct measures are given below.

Head/eye tracker

Two of the main issues for many systems are visual distraction and the effects the system has on the driver attending to traffic versus to the system. By using eye or head trackers in the vehicles, continuous data of some driver state/attention measures can be obtained. The problem with head and eye trackers is mainly the risk of significant data dropouts due to limitations in driver head and gaze tracking.

State-of-the-art head/eye-tracker technology today is relatively expensive, but the benefit of using such a system should not be underestimated. It is recommended that head or eye tracking systems are employed in FOTs where driver state is an issue.

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

FOTs using eye-tracking devices have continued to find this a challenge.

Other

It is strongly recommended to avoid sensors that the driver has to put on or wear. This will help to ensure drivers behaving as naturally as possible.

The latest driver monitoring systems offer features beyond head and eye tracking, such as classification of expressions, detecting hand and joint position, presence of e.g. a mobile phone, and anonymisation of data by replacing the driver's face with an avatar.

7.3.4 Non-video environmental sensing

Sensors already integrated (by OEM/Road Administration)

If a required environment sensor is integrated into the vehicle or the infrastructure, significant effort should be spent on trying to add the sensor data to the used data acquisition system. In this integration several issues have to be considered.

- OEM allowing/disallowing access to data
- System interfaces: Some low-level sensing information (e.g. object tracks from radar) may require special interfaces to be implemented both in hardware and software
- System comparability: If the studied vehicles have different OEM-integrated systems, they will provide varying quality of data as well as different resolution, range, field of view etc.

Add-on environment sensing

If additional sensing needs arise, consider the recommended process for adding sensing:

- Use the sensor matrix for sensing needs and identification
- Requirements for extra sensing need must derive from hypotheses and performance indicator requirements
- Additional considerations: Does extra sensing require additional interfaces? Does it require information from vehicle buses? How can the sensor be integrated without significant effort? Does it require repeated calibration?

General guidelines for specifications are difficult to define. However, e.g. radar and other object tracking sensing systems, required field of view, radial and angular resolution/precision are important to define based on the hypothesis.

7.3.5 GPS

A GPS device can provide a GPS Time reference timestamp and a difference between the GPS time and UTC. It is highly recommended that this is used for synchronisation within a data acquisition unit as well as between systems (in-vehicle, Nomadic Device, infrastructure, and services). Information about the present local time zone can be useful for subsequent analyses and synchronisation with non-UTC devices.

Multipath propagation of the GPS signal is a dominant source of error in GPS positioning, since it depends on local reflection geometry near each receiver antenna. In most cases this can largely be corrected with DGPS solutions. Errors depend on time of day, satellite positioning (in zenith or low orbits) and other atmospheric disturbances.

7.3.6 Audio and driver annotation

Continuous audio recording is potentially a significant privacy issue and is not recommended. In some FOTs, drivers have had access to a comment button on the dashboard, which, when pressed, would start recording any verbal comments during a pre-set number of seconds (usually around 20–60 seconds). Use of the button could be encouraged if the driver feels that it is warranted or at agreed (critical on non-critical) events. Be sure to inform the drivers consistently about the annotation possibility and provide some simple guidelines on how to offer the comment.

7.3.7 System function/status

A system under evaluation, such as an LDW, ACC or FCW, needs to be continuously monitored to ensure that it is operating properly. The system status signal will thus form a measure to be recorded in the data acquisition system. The status signal will depend on the specific system and needs to be provided by the system manufacturer or vehicle manufacturer, thus requiring strong collaboration with the actual provider.

The availability of the tested function, e.g. a driver support system or automated driving, could also be logged to monitor how often the function has been available for use versus how often it was used.

7.3.8 Vehicle metadata

Information about the studied vehicle is important for analysis and study design. The recommendation is that for each type of study, systems, functions and specifications that may act as confounding parameters in a specific analysis should be stored. During analyses, these parameters may have confounding effects if an inhomogeneous test fleet is used. Examples:

- FOTs with focus on safety: Information should be stored about integrated systems that may contribute to driver distraction
- FOTs with focus on environmental issues: A more powerful engine, automatic gear shift or four-wheel-drive are most likely to be confounding parameters in an environmental analysis

Wider metadata – from weather conditions to prevailing economic conditions – is also essential for interpreting data post-study.

7.3.9 Coding/classification/transcription

As part of the data reduction and analysis process, sections of time will often be labelled with classifications according to a coding scheme or syntax. Depending on the study, sections of time can be assigned categories such as “crash”, “near-crash”, “incident”, “curve speed warning”, “lane change”, “crash avoidance by steering”, etc. The categories can be different types of events, scenarios, operational phases and so forth.

When classifications are made, they are often saved and thus become a new data source which is added to the database. For example, an index designating all instances of lane changes in the dataset can be created and saved. Regardless of whether the classification is performed by a human analyst transcribing video or by an algorithm applying kinematic

trigger values to the data, this process of classification should be seen as a type of sensor providing a new data source. Thus, it is comparable to other types of off-line performance indicator calculations (see Figure 7.1). It is recommended to plan for the creation of these new data sources or measures during the data reduction phase, after the data has been uploaded to the FOT database.

7.3.10 Geographical Information System (GIS)

Geographical information, such as speed limit, road curvature, roadside embankments and other on-and off-road information are a valuable source of data. Contextual indicators of events and situations provide added insight into both behavioural aspects and how the infrastructure influences system performance.

GIS-derived information about the current road (road type, speed limit, rural/urban, banking, curve radii, etc.) could be used directly in the vehicles as separate measures. By doing this on-line, the necessary post-processing is reduced, but requires additional software and possibly hardware. One advantage of performing this on-line can be that the absolute position (e.g. GPS) does not have to be stored, thus lessening some privacy concerns. An advantage of using commercial navigation software as part of the on-line map-matching and information extraction can be that there is no need for potential in-project development of map-matching algorithms. This technique has been used in some state-of-the-art FOTs. It is important to validate that map-matching and other GIS data extraction is done in a proper way, and that e.g. speed limit information is up-to-date.

7.3.11 Communication unit

A communication unit is a device that provides vehicle-to-vehicle and vehicle-to-infrastructure connections in cooperative system FOTs. It should also be considered as a measurement source for all communication-based aspects of the FOT. Typical measurements include received and sent messages, network traffic congestion, number of connected nodes, etc.

7.3.12 Application Unit

The application unit is a device intended to run developed functions, and is typically an x86-based computer in rugged configuration. If an application unit is deployed in an FOT it also provides a variety of measurements. All developed functions should be providing measurements of their state and key variables used. The underlying facility layer (e.g. vehicle access, position access, local dynamic map) can also be a valuable source of data.

It must be decided whether it is feasible to also run the data acquisition system directly on the application unit, which simplifies access to all key variables. The downside is that the DAS might consume too much processing power or drive capacity, thus interfering with the functions.

7.4 Mechanical requirements

The following mechanical guidelines and requirements are primarily applicable to FOTs with in-vehicle data acquisition systems.

7.4.1 Size and weight

The system should preferably have a negligible effect on the driver's use of the vehicle, including limitations in storage space.

7.4.2 Connectors and interfaces

Most DAS hardware problems result from physical connector issues (both the DAS and its peripherals). Thus it is recommended to use connectors that have some locking between connector genders. Cable pull-relief should be used when possible.

7.4.3 DAS mechanical cover and ease of access

It is recommended that any indicator LEDs on the DAS be fully accessible, without the need for tools, to non-technicians. Also, having the possibility to connect interface devices without having to remove covers is preferable.

7.4.4 Crashworthiness and vibration resistance

For all FOTs, the minimum requirement for ruggedness of the application unit is that the entire system should operate under normal driving conditions throughout the FOT, and also during harsher incidents. Data acquisition systems must not interfere with available safety systems, and should be integrated safely so as not to cause harm to traffic participants (drivers, pedestrians, other), even in the event of a crash.

7.4.5 DAS environmental requirements

Environmental requirements for the DAS concern mainly temperature. Too high and too low temperatures (both static and transient) will affect the DAS. Important considerations include sufficient cooling of the system itself and of the ambient temperature. The amount of cooling needed will depend on the processor used and the workload. If the FOT is using video compression software, the need for processing power may be drastically higher than if only simple signals are recorded. A high processor load generates significant levels of heat. The need for forced ventilation also depends on the DAS mounting position.

If the DAS is placed in a shielded location the need for water resistance may be negligible, although the internal parts of the DAS should be able to withstand reasonable levels of condensation. If applicable, it is recommended that a simple dust/particle filter is placed by the main air intake of the DAS.

Components with moving parts need special attention. Traditional mechanical hard disk drives (HDD) have been among the most sensitive components. Automotive-grade hard drives are available, but FOTs are increasingly moving to use solid state drives (SSD) in their data logging. Besides enduring shocks better than HDDs, SSDs also offer higher data transfer speeds.

7.5 Electrical requirements

7.5.1 Power management

The main requirement for the DAS setup is that the installation should never affect or impair normal vehicle function, regardless of the environmental conditions. It is one of the most

critical issues for drivers pondering whether to accept the installation of equipment in their vehicles. If FOT installations are rumoured to impede the trustworthiness of vehicle operation, both the study at hand and subsequent trials may be compromised by a shortage of volunteers.

The entire DAS installation should not draw power to the extent that the vehicle's battery charge falls below its ability to start the vehicle. Care should be taken that the system does not draw power (or then only minimum power) if the ignition has been turned on but the engine didn't start, or if the engine stops without the ignition key being removed.

7.5.2 Interference with in-vehicle equipment

In all in-vehicle installations, the aim should be to minimise or even manage without the use of AC-powered devices. DC/AC and DC/DC converters are sources of electromagnetic noise and may affect both standard in-vehicle equipment (such as FM radio) and the FOT installation itself.

Attaching any equipment to the in-vehicle bus systems has to be done very carefully. Transmitting data on vehicle buses should in most cases *not be needed or not done at all* in an FOT implementation. Failure to adhere to this might be dangerous and result in vehicle operational malfunction that may result in significant cost, injury or death, or produce other very unwanted results. Even adding only listening/eavesdropping devices should be approved by the vehicle manufacturer before being implemented.

7.5.3 Laws and regulations

Depending on the FOT study at hand, different regulations will apply. In some cases, CE certification of the FOT equipment may be necessary, and each project must verify what is applicable to the specific study. If the vehicles are to be driven in non-EU countries the specific regulations for each region should be verified.

For wireless communication and for some sensing systems there are regulatory restrictions on transmitting electromagnetic radiation. A few examples of sensing systems that have direct regulatory restrictions are lidar and radar. The restrictions may be based on electronic interference or harm. This has to be taken into account for each individual FOT sensor setup. Still, for each instrument and jurisdiction, care should be taken to investigate the applicable regulations.

Several countries have regulations for equipment employing radar or laser technology on public roads, as these can affect the effectiveness of e.g. authority speed surveillance instruments.

7.6 DAS data storage

7.6.1 Storage capacity estimation

The main aim of the *storage capacity estimation* is to guarantee the availability of free space for recording the vehicle data. Storage capacity depends on the following factors: number of recorded signals, sample rates and sample size of the recorded signals, data

collection method, driving hours, data size reduction (filtering/compression), and data deletion procedure.

Ideally, the sample rate for each signal should be the lowest possible that can guarantee that no information relevant to the FOT hypotheses is lost during the sampling process. Using a dynamic sample rate and a high sample-rate buffer would make it possible to adapt the resolution of the recorded data, depending on the event. The drawbacks of using dynamic sample rate are that the recording system complexity (and probability of faults) increases, and more post-processing of the data will be necessary to handle the different sample rates. Database design and search become more complicated.

The data collection method can be continuous or limited to specific events of interest, such as time intervals during which lateral acceleration exceeds a certain threshold or the vehicle enters a curve at a speed above a given value. In FOTs where only triggered storage is used and where the driver subjectively triggers it in some way, it is important that past data is also recorded by using e.g. a ring buffer (buffering one to five minutes in the past). The level of pre-trig time will differ between projects, and should be defined based on the hypotheses for each study.

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

Compression algorithms can help to reduce the data size. A lossless compression (such as zip) can be used for CAN and sensor data, whereas a lossy compression (such as MPEG-4 and MP3) is normally acceptable for voice and video, respectively. The drawback of compression is that the complexity of the system increases and new possible sources of error and malfunctioning are introduced.

A safe data deletion procedure implies that no data is deleted in the vehicle until a copy of the data has been backed up, verified, and stored in a safe place.

In Figure 7.3 **Figure 7.3** the factors influencing the data size are summarised.

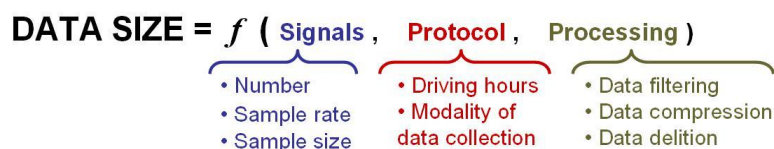


Figure 7.3 Factors influencing data size

Storage capacity may be depleted and the intervals for retrieval and uploading may present some variability. These factors should be taken into account by guaranteeing enough tolerance in the final storage size. Since running out of space to record data would result in data loss, a 20% to 50% margin of storage size tolerance is recommended.

In some studies where levels of allowed data loss are very small, it is recommended to use direct in-vehicle data backups. This can be done, for example, by using several storage media with a data-mirroring solution (e.g. RAID).

When triggered data acquisition has been chosen as the data collection method, great care must be taken to define the trigger definitions. Even if triggered logging is used for the evaluation of effects, most studies will require baseline data. It should be possible to configure and acquire baseline data using the same DAS also for these cases.

FOT activities that use any type of triggered data acquisition and have high-data-rate data sources will need significant amounts of main memory to handle the necessary ring-buffer for pre-triggering storage. Moreover, estimations of needed storage space and data retrieval/upload frequency will be affected.

7.6.2 Data retrieval/uploading procedure

Data retrieval/uploading procedures are needed to make sure 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 prevent data storage waste (caused by double storage). In Figure 7.4 the data retrieval and uploading steps are given.

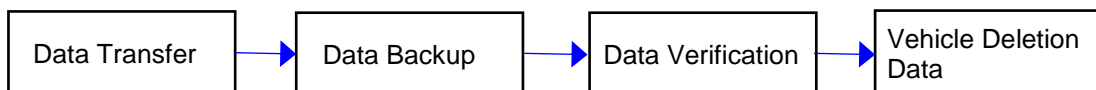


Figure 7.4 Data retrieval/uploading steps

Data transfer

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

Depending on the support used to record the data in the vehicle and the data size, different data transfer modes can be implemented.

Generally, data transfer poses two main problems: It may be time consuming, and data can be lost during transfer. Different transfer modes are compared in Table 7.1.

Table 7.1 Pros and cons of different data retrieval/uploading modes

Data transfer mode	Time efficiency	Cost efficiency	Technical complexity	Data security	Driver comfort
Data “picked up”.	+ Very fast from the vehicle point of view.	- May require paying someone to pick up the data. - May require an extra vehicle chasing the FOT vehicles.	+ Very reliable because simple.	- Data may be misused between “pick-up” and final storage.	- May require the driver to go somewhere or move the data.

Data transfer mode	Time efficiency	Cost efficiency	Technical complexity	Data security	Driver comfort
Data transmitted via wired connection.	- May be slow for a big amount of data.	- May require paying someone to do the downloading.	- The equipment needs to be safely accessed.	+ Can be very secure if the driver is not able to access the data.	- May require the driver to go somewhere or to move the data.
Data transmitted via cellular connection.	- May be slow even for a relatively small amount of data.	+ Download can be automatic. - May be expensive depending on the network provider.	- Automatic wireless download is very complex.	+ Can be very secure.	+ The driver does not need to do anything.
Data transmitted via WiFi connection	+ Relatively fast - Requires vehicle to be in Wi-Fi AP range	+ Download can be automatic. - Requires Access Point installation.	- Automatic wireless download is very complex.	+ Can be very secure.	+ The driver does not need to do anything.

Data backup

Data should be backed up and stored in a safe place as soon as it becomes available. Ideally, the main copy and backups should be in separate, secure locations. In some studies where levels of allowed data loss are very small, it is recommended to use direct in-vehicle backups. This can, for example, be implemented by using several storage media with a data-mirroring solution (e.g. RAID).

Data verification and deletion

Due to the potentially huge amounts of data being handled, data verification is important since the probability of errors during the copying process is high.

Data from the vehicle should be deleted *only* once the data has been backed up and verified.

Data loss

Experience from previous FOTs shows that data loss at the retrieval/upload stage is common, despite the fact that it could be avoided almost entirely with a robust and well-tested procedure. To prevent this loss from happening, it is important to verify that the data is consistent before deleting it from the vehicle. Should it not be consistent, the vehicle data logger should be checked as soon as possible and monitored to identify any issues as soon as they arise.

7.7 System configuration

Although DAS configuration needs can differ, one basic requirement that applies to all FOTs is that it should be possible to find and configure a specific DAS after the study has been completed.

7.7.1 *DAS inventory management*

A system for basic inventory management is recommended for FOTs with more than a few vehicles in use. For such a system to be efficient, sensors, DAS units, vehicles and all other equipment need to be included and relevant supporting procedures developed. For any one point in time, it should be possible to deduce the exact hardware and software configuration of a particular installation.

7.7.2 *Configuration tools and traceability*

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

Similarly to above, it is important that traceability of software configuration changes is maintained. In this way, the exact software configuration of a particular installation at any particular time can be reproduced and reviewed.

7.7.3 *Switching between configurations*

For some FOTs, switching between configurations in each vehicle may be necessary. In addition to having to change driver ID when a new subject is introduced to a vehicle, there are a few other situations where it may be important to know who was driving or what was done with the vehicle. If this is not separated from the subjects, these trips may be classified erroneously. Examples of situations to consider are: the vehicle driven back for overhaul/maintenance, validation testing just prior to vehicle delivery, and other engineering testing.

If possible, the use of remote desktop tools over wireless (e.g. WLAN or 3G) or wired networking (Ethernet) is recommended for remote administration.

7.8 Acquisition of data

When controlling the power supply to the DAS, the start-up and shutdown speeds must be optimised to reduce loss of data. Loss of data can occur both during hardware initiation when no software is started and during hardware termination when no software is able to trigger on a vehicle restart.

7.8.1 *Start-up*

Normally, data acquisition will start as the vehicle ignition is turned on. In order to minimise the data lost during the start-up procedure, the hardware and software must load and initiate as quickly as possible. The start-up time (or the duration when data is lost) should be well monitored and documented, preferably as a property associated with each recorded trip (since it might differ with temperature etc.). Start-up of the DAS hardware should not be done if the voltage is too low.

7.8.2 Acquisition of data

The DAS hardware should be kept powered on and running during the entire trip. To ensure that the host power system is not overtaxed, the power management unit must continuously monitor the power supply and initiate shutdown if a persistent voltage fall is detected. The system must not shut down on temporary variations such as during engine crank. For such circumstances, an energy reserve such as a battery may be required.

7.8.3 Shutdown

Once the data recording has stopped, the DAS should be kept running for a short time (typically a few minutes) in case the vehicle is started again. Otherwise, the DAS hardware should shut down itself as fast as possible. The power management unit should then, after a short interval, cut the power supply to the DAS, regardless of whether it is properly shut down.

7.9 Synchronisation

FOTs are considered to be used for logging of data and not as part of a hardware-in-the-loop component (for example prototype system development). This means that data from the different data sources do not necessarily have to be available for storage close to the real-world event, as long as they are individually timestamped for off-line recreation of the timeline.

7.9.1 Timestamping versus real world event

Latency is the time between when the actual real-world event takes place and when the data from each respective sensor is timestamped in the logger. In most FOTs there is a need to specify the requirements for allowed levels of latency, based on the hypothesis. It is recommended to explicitly evaluate what the latency is for each data source (and for some sources individual measures) and compare this with the defined requirements (based on hypothesis; needs to be done as part of the hypothesis and performance indicator generation). If latency has been measured properly and there is limited jitter, the timestamp can be corrected by offsetting with the latency.

Note that for data sources that are not controlled by the DAS implementers, such as vehicle CAN, it may be even more difficult to obtain the necessary information for the latency from real-world event until timestamping.

7.9.2 Integrated sensing synchronisation

Depending on the methods of analysis and the implementation in the database, the needed level of synchronisation as well as the importance of measuring latency between different integrated data sources in a vehicle will differ. Table 7.2 outlines the issues and methods for calculating the latency for some in-vehicle data sources.

Table 7.2 Methods and issues in calculating latency for in-vehicle data sources

Data source	Methods and issues in calculating latency
Video	May produce significant jitter/fluctuation on itself and other data sources. Approximate latency can be ascertained using a synchronised LED light (measured by digital I/O). Preferably hardware synchronised cameras should be used.
GPS	The latency can be calculated very precisely since the GPS time is the actual acquisition time.
CAN	Difficult to establish latencies for internal systems, but with reference sensors it is possible to get the latency for some measures.
Acceleration/yaw rate	Latency can easily be measured using a reference sensor.
Radar and lidar	A lab setup with reference sensors of tracked object motion can be used. Data can also be compared against other sources such as video and aerial imaging.

7.9.3 Synchronisation with nomadic devices

The recommended method to realise synchronisation between the different sources of data is to use GPS UTC time. Thus, all acquired data should be associated with a timestamp that is represented as absolute GPS derived time. It is recommended to use nomadic devices that have an easy interface with GPS, so that the absolute time information is available.

7.9.4 Synchronisation of infrastructure systems

Also in the case of infrastructure, the UTC time derived from GPS is useful. Another possibility is to use an on-line time synchronisation service like NTP/SNTP. In the case of a traffic data and safety related FOT, data may have to be stored as raw data, accurately synchronised in time, to allow the reconstruction of the scenario in the following data analysis phases.

7.9.5 Synchronisation of cooperative systems

Synchronisation of systems with communication between vehicles can also be realised without a central infrastructure. Again, one option is to use the common reference time provided by GPS.

For ITS stations involved in cooperative systems FOTs and additional external sources, it is advised to use a network time-synchronisation protocol, such as NTP. This method can also be useful when synchronizing the clock of several vehicle computers.

7.9.6 Synchronisation with interviews and other subjective sensors

In many cases, it is enough that the interviewers write the date and time (hours and minutes) of the interview or questionnaire. If the subject is requested to indicate and/or

comment on events during driving (for example by pushing a button), this should be timestamped when logged if possible, to enable synchronisation of the event/comment with other data. The accuracy needed is in most cases less than 5 seconds. For post-hoc structured comments or questionnaires on video or events, it is important to define a process of linking the events to absolute time.

7.10 DAS status and malfunction management

To simplify laymen feedback on system status to the responsible technicians at times of system problems, LEDs or similarly externally viewable information about the system status can help and are recommended. This is especially useful when testing complex prototype systems requiring correct operation of several components, such as automated driving systems.

7.10.1 System status uploads

In order for the people responsible for DAS and sensing in the project to be able to continuously monitor the status of the test vehicles while on the road, a remote (wireless) transfer of the system and sensing status is preferred. (See [Section 7.14.1](#) for details.)

7.10.2 Malfunction management

A process for identification of problems in the vehicle, contacting the driver and exchanging sub-systems or the entire vehicle should be developed.

7.10.3 Spare system management

For quick and efficient problem solving it is recommended to keep spare parts for all components pertaining to the DAS. It has been suggested that a 10% extra number of spare parts should be kept within the FOT, and that this number be added specifically when estimating the total project cost. All spare parts are to be managed by the inventory management system. Testing and calibration of spare parts should preferably be planned and performed within the process. Supporting management procedures need to be developed as well. At least one fully equipped DAS system should be kept “on the shelf”, prepared and calibrated for immediate use.

7.11 System installation

7.11.1 Installation procedures

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

- Mounting: positioning, means of attachment, accessibility, safety and security
- Cabling: dimensions, shielding, drawing, mounting, tolerance and labelling
- Connectors: soldering/pressing, robustness, impedance and labelling to avoid mix-up
- Power supply: consumption, fuse, voltage, source and switching

- Environmental endurance: effects on electromagnetic disturbances (EMC), temperature, humidity, vibration, shock, electric safety and dirt.

It is of great importance that the FOT system installation is adapted to the requirements set by all other systems in the vehicle. If this is not done properly, the installed system could generate disturbances that might void warranty of the original systems – or even an entire vehicle. All systems that may possibly be in conflict with the installed system need to be identified. An adaptation plan must then be developed for each system to ensure that they will be able to operate properly after installation.

The actual installation work needs to be done by operators that are authorised to work on the actual host system, and during the installation work, all changes to the host system (if any) must be documented in detail. Depending on the study and region, the authorising body could for example be the OEMs, insurance companies (voiding warranty etc.), the project and/or a legal entity.

If several vehicles are to be installed, it is recommended to select one vehicle as prototype. The prototype installation will then revise the installation specification continuously during the work.

7.11.2 Installation verification and calibration

When the system is installed it needs to be verified and calibrated before the data acquisition starts. The verification will refer to the installation specification and verify that all requirements are met. Monitoring of all potential sources of interference so that no conflicts are caused is important.

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.

7.11.3 Dismounting the system

When the data acquisition is finished and the system is to be uninstalled, the installation documentation shall be used to ascertain that the host system is restored to its former condition. Finally, all proprietary data needs to be removed from the FOT system before it is disposed of or reused in future projects. Remember to include dismounting costs in the FOT budget.

7.12 Proprietary data

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

- When the OEM is strongly involved in the data acquisition process during the FOT execution, the confidentiality of the CAN/LIN/MOST specification is not an issue within the project.
- When the OEM does not handle the data collection by itself, the usage of CAN gateways is proposed. The CAN gateway has to be programmed by the OEM to provide the data from the CAN/LIN/MOST bus according to the agreed logger specification.

The second issue – reverse engineering of functions and systems – is also an issue within the FOT project. Each project will have to handle this and define what is needed. In some cases, it may be necessary for the OEMs to handle detailed low-level data and aggregate it on a certain level before it is provided to the project partners responsible for data analysis. A general recommendation for future FOT projects is to define in advance what level of system data is needed to answer a specific research question and whether the involved OEMs are able to provide this data to the project.

In some FOTs, OEMs might be interested in the acquisition of additional data which is not directly related to the project and proprietary to the OEM. This should be allowed. The OEM could separate the additional data from the project data before the data is provided to the further project for analysis.

7.13 Personal integrity and privacy issues in data acquisition and analysis

Recommendations for the definition of necessary legal arrangements depending on the specific FOT are not covered here. See [Chapter 3](#) for further information.

Different levels of data security should be implemented in order to cover personal and privacy issues properly. The data access right of a project partner should depend on their specific role in the project.

Several different levels of data security should be implemented. Driver data allowing direct conclusions about the identity of the driver is considered to have the highest requirements regarding data security. Video data of the driver's face and driving diary-level GPS data are common examples. For this kind of data, performing anonymisation steps would enable granting wider access rights for evaluation, whereas the raw data needs to be strictly controlled.

It might also be necessary to implement a location-based control, which deactivates the video recording where required (e.g. when driving in countries or areas with specific legal requirements).

7.14 On-line quality management procedures

In all FOTs, assuring data quality in the data collection and data management process is very important. The procedures for data quality assurance before, during and after the data collection should be well defined. Specifications and plans should be written for each individual FOT.

It is recommended that a quality management team is appointed for each individual FOT with roles such as: periodic quality overview, OEM contact person, subject contact person, DAS and sensor maintenance person, and vehicle maintenance person.

7.14.1 Remote automatic upload

In most state-of-the-art FOTs wireless transfer of vehicle and data status has been used, in order to assess the status of vehicle, DAS and data without having to physically access the vehicle when the vehicles in the study are on the road. Different transfer techniques such as simple text messages (SMS) or GSM/3G have been used for status uploads. A maximum delay from the time of actually collecting the data until it has been analysed for quality and status should be defined. Otherwise the project risks that the vehicles on the field are potentially not collecting the required data. The maximum delay should be set based on the accepted levels of data loss and the length of the study.

When the data has been uploaded and put into a database, trip statistics can be calculated per vehicle or driver. It is recommended to use this to identify extreme/abnormal driver/usage behaviour early in the study so that if necessary, drivers can be exchanged. This driver monitoring early in the study is highly recommended so that the study schema of the specific FOT is kept.

If an FOT is to be executed across country borders and include roaming for wireless services, investigation of the cost/benefit of using quality data upload systems outside of the “home country” should be made.

7.14.2 Automatic and manual quality checks

It may be tempting to do the quality assessment fully automatically, but state-of-the-art FOTs have indicated that by doing this there is a risk of contacting the driver in cases where the error or anomaly is not significant for the study. Because of this, it is recommended that the quality assessment should be set up in different steps and that before (if) employing a fully automatic system, the algorithms for the assessment should be thoroughly validated. An automatic threshold-based warning system can be applied for some hard and very important measures. A tool for maintaining the warning system should preferably be checked by a responsible person each day.

7.14.3 DAS and sensor maintenance

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

7.15 Data acquisition in FOTs of automated driving

Data acquisition in Autonomous Vehicles (AV) poses additional challenges compared to FOTs of ADAS and other systems. The principles of data recording are the same but with an increased focus on time synchronization and the use of environmental sensors. This might be related to the fact that Automated Driving (AD) functions are not yet mature and

the focus is on development rather than evaluation. This should change over time when the functions are introduced on the market and large-scale automated driving FOTs are needed.

Recording of point cloud data stresses the Data Acquisition System (DAS) because of large amounts of data. If the point clouds are not stored, adequate processing (in sensor, AD computing unit or in the DAS) of the data is of great importance to represent the detection of objects, their position, movement and distance. The use of reference cameras is important in validating the AD function, the response of other traffic participants and the driver and passengers inside the vehicle. Camera recordings are important also when validating automatically detected time segments of interest.

The DAS should use a common reference time and the timestamping of data packages should be done as close as possible to the sensor. The principle is the same as for previous FOTs but the increased number of sensors and computers in AVs adds to the problem: common time reference and synchronization among sensor and data recording is challenging.

From the perspective of development (and replay of driving situations), evaluation and liability, the decisions made by AD functions may have to be stored. On what level data and processes need to be recorded depends on the specific use, and legal experts should be consulted regarding liability.

Configuration management is very important in the rapid development of AVs. For each test, metadata about the current setup should be documented. This includes both the hardware (sensor setup) and software (version or configuration) of the current AD functions. It is also important to understand what other vehicle safety systems are active in the vehicle, and in which situations the different systems override the others.

Constant changes to the tested system can make evaluation difficult and also lead to discarding data. Therefore, changes are important to manage especially from evaluation perspective.

Due to the development nature of first automation tests, test data amounts have been large and datasets have been often collected manually: by collecting SSD drives, plugging in a 10 GB Ethernet cable or starting a WiFi upload. Data collection progress, vehicle location and system status, however, have been monitored over mobile connections to ensure test success. Uptake of 5G connectivity and high-speed WiFi could refresh the discussion on when and how to transfer data from vehicles to back office (or cloud) storage. As the technology becomes more mature and automated event detection as well as data anonymisation improves, FOT data needs could also get reduced.

8 Guidelines for Databases and Analysis Tools

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

The following sections focus on databases for large FOTs (where thousands of hours of raw data are collected, even using different server locations). The size and logging needs of the FOT determine the complexity of the necessary database.

The requirement analysis for the database shall consider the usage of the database during the project duration, but also include requirements for possible re-use of the data after the project duration. Recommendations for data handling, especially in connection to data sharing, can be found in the [Data Sharing Framework](#) (FOT-Net Data and CARTRE, 2019).

When designing data analysis, it is important to consider the following general recommendations:

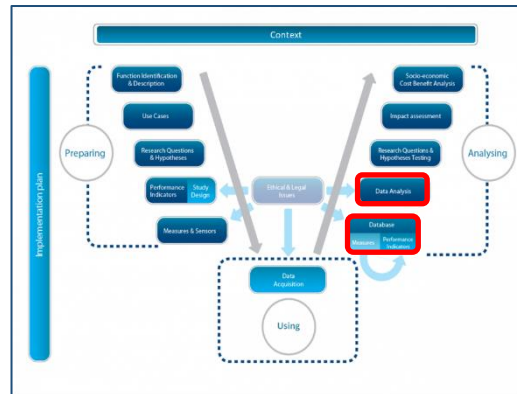
- Create documentation of data structures, tools and processes in order to enable analysis also by persons who do not take part in data collection
- Provide a detailed test protocol with all relevant test scenarios
- Make sure that a complete data analysis process (data processing and analysis tools) is developed before the piloting phase, in order to test and assess all data analysis steps during piloting.

8.1 Database design and implementation

8.1.1 Preferred database models

The main challenge of an FOT database is to make thousands of driving hours manageable from a storage perspective and available for ad-hoc analysis. FOTs often have very specific demands, making it difficult to recommend one generic database model. Relational databases have been used widely, but big data can sometimes be easier to manage and process as plain files, or by using non-relational databases (NoSQL). Combinations of such storage options are popular as well, e.g. database entries containing links to video files. Database design is affected by analysis needs, e.g. if there is a need to easily view certain sensor data or merely study indicators processed from it.

It is mainly the continuous time history data (e.g. vehicle sensor signals logged at 10 Hz) and video data that put high demands on software and hardware. Additional data such as information on vehicle type, video annotations or lists of detected events enrich the time history data, but are of smaller size.



Tip for relational database design: A measure equals a column in a table in the database. In order to avoid costly joint operations when performing analysis, it should also be considered to keep the database as de-normalized as storage allows: Some data in a trips table, such as driver ID or vehicle ID, could for example be repeated in time history tables. From a storage or relational database perspective this is not commonly preferred, but when doing analysis, it will reduce the complexity of queries and save computing time.

8.1.2 Data categories in the database

Besides the sensor and video data, many data types need to be transcribed in the database in order to prepare it for analysis. Often the data is segmented using time or location identifiers that are being used as references within the database. Most commonly “Trip ID” has been implemented in previous databases, but also “Journey ID”, “Event ID” or “Location ID”, depending on why and how the data was collected. In many databases combinations of these segments are being used.

Time history data transcription

Data logger ID and *time* are cornerstone indices in the database designs. This means that sensor data must include a timestamp when inserted in the database. Relational FOT databases commonly use a pre-defined (fixed) sample rate within one table. This way of storing information can support straightforward analyses and visualisation as well as potentially save storage space. Varying sample rates in tables place higher needs on implementation of analysis software and indicator processing. When different sample rates are needed within the same FOT database, the different data sources can be saved into separate tables. Furthermore, data with frequency differing from the default (e.g. 3 Hz vs 1 Hz) should be marked as potentially “incompatible” with the main data.

It is important to have a strategy in case of data loss or error codes. If a sensor gives no data, a NULL value can be inserted. Some FOTs suggest that using the last known sensor value makes analysis easier. The problem with data that is actually not valid has to be dealt with. Error codes must be documented and later taken into account in the analysis phase.

External data transcription

Data from external sources can be collected, especially if other actors such as infrastructure have incidence in the hypotheses and in performance indicators defined (see Chapters 4 and 5). A typical example of this data category is weather conditions, traffic status or event information in the case of cooperative systems, and is usually provided by external entities (e.g. traffic management centres). When collecting external data, it is highly important to be able to synchronise this data with the data collected during the trip (e.g. using GPS time).

Transitional data transcription

Transitional data can be stored separately into tables that only contain data when transitions occur. This is suitable for state change variables (i.e. turn indicators or windshield wipers) to reduce storage overhead and allow high precision timing.

Despite the potential storage overhead, it may also be possible to handle transition data in the same frequencies as measure data, to simplify analysis.

Background data transcription

There are at least two types of *background data* that should be stored in the database model: 1) the driver's and 2) the vehicle background data. Driver data such as age and gender should be stored in a *driver* table but any data to identify the driver should be stored securely and separately (see [Chapter 3](#)).

Subjective data transcription

Subjective data, e.g. questionnaire data, should be saved together with participant IDs and time. To reduce errors, automatic transcription of subjective data is always preferable. Transcription of audio/voice messages is recommended.

Events/classifications transcription

An FOT database can consist entirely of events, if a triggered data collection approach is adopted (as opposed to continuous data collection; see [Chapter 6](#)). In other FOTs, where data collection is continuous, the ability to find and classify events of interest is of central importance. Classification and use of "events" (classified time periods) is an important aspect of FOT analyses (see [Section 7.3.10](#)). Some events are straightforward and simple to identify, for example hard braking defined as peak deceleration $> 0.7 g$, and may not need to be saved as a discrete or transition variable. However, many events involve a considerable amount of effort to find and validate, and are worth saving into a discrete variable database or an index/summary to facilitate data query and analysis.

Event pointers should be saved to speed up analyses and can be used in combination to describe more complex situations with multiple events.

The data can be described by manual or automatic video coding or annotation. This data is often associated to events, but could also be any other time or location segments depending on the analysis (see [Section 8.4.1](#)).

8.1.3 User data spaces and data collaboration

As the analysis work begins, there will be the need for the analyst to store new data (coming from combination or processing of the raw data) in a private user space. Private data should be kept in a private "user space" (database or schema), in order not to risk inadvertent confusion with original project data.

If this new data is also relevant for other users of the database, a solution to share this data in a project internal space should be implemented. The collaboration process should enforce the data to be described and metadata description is needed (e.g. including data origin and function/method/algorithm applied to the data).

Some of the data could be of interest for external researchers, in some cases on an aggregated level and therefore exported or accessible via web interface. Other data, such as video, could be made available to other researchers using more rigorous procedures.

Sharing this data must be approved by all stakeholders in the FOT, preferably early on in the project.

8.2 IT infrastructure implementation

8.2.1 Video data storage

A common way of storing video data in the FOT context is to store the video files on a file server and store the links to the video files in the database (to link the videos e.g. to time or location segment identifiers).

For very large FOTs, large amounts of video data can exceed the limits of file systems or storage appliances. This can cause extra complexity, for example to add logic (scripts) to enable a single mount point that is preferable from a data management point of view. Different file systems and appliances should be evaluated.

It is worth examining different video codecs; using an optimal codec or a change in resolution can reduce the storage need significantly. These decisions are needed early on, as the cost (mainly CPU time) for re-processing can be high and a large FOT might have need for a high-performance cluster to complete in acceptable time frames.

8.2.2 FOT Hardware

When extensive video data has been collected from an FOT, it is recommended to separate the database and video file server in order to configure the hardware individually. Very fast and reliable disks are available these days, even with a limited budget. In many cases, storage at some kind of a disk cabinet, NAS (Network Attached Storage) or SAN (Storage Area Network) is most appropriate. A storage setup with some kind of redundancy (RAID) configuration should be considered, in order to be better prepared if a disk crashes or some data blocks are corrupted. The analysis databases should use faster disks than plain storage servers.

Outsourcing system operations is possible, commonly offering low cost and flexible solutions. However, the costs for network bandwidth, backup, and administration must be thoroughly evaluated with respect to the expected use of the data. Since the data is not in the control of the project organisation, it is important to make sure that the supplier ensures legal compatibility (e.g. regarding the use of PII data) and to agree on precise requirements on data security and protection.

8.2.3 Distributed system at many locations

For the database, a single common database or cloud-based distributed database could be used depending on the requirements of the FOT.

Due to the location and size of the FOT, it might be necessary to establish a distributed solution for data storage. This is especially true when deploying an FOT in different countries with different local data servers. In this case, it is recommended to also establish a central data server in charge of gathering all data (relevant for analyses) from local data storages.

Connection to this central database should be guaranteed so that information can be easily transferred from and to it. A broadband IP connection is then recommended, with simultaneous access. The use of remote desktop technologies should be evaluated to provide the possibility to work collaboratively within a common environment.

8.2.4 Physical access

Physical access, as well as the approval process for access to the hardware, must be documented.

8.2.5 Logical access

Logical access, as well as approval for access to the database, must be documented. A role-based access is advised when any user in a certain role of the database obtains certain access. This also applies to the supporting operating system. Any FOT must define the roles and permissions of the database. These roles can be:

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

It is recommended to organise the data with re-use in mind. This means that policies and data access routines are set already in the data collection project. More practically, this could mean that one uses membership-based data access patterns (to get access to data, a user account is a member of an analysis group, which is a member of a data access group). Since the data access privileges are associated with the data access group, there is no need to reset those privileges after the initial data collection project has ended. The only need for future projects is to tailor the analysis group to the data access groups. Most likely the data access groups follow data access statements or restrictions in the consortium agreement or other contracts (e.g. with external data suppliers or OEMs).

8.2.6 Personal integrity and confidential data

Driver data must be stored according to the access restrictions defined by the steering committee. In a collaborative study, some data may be classified as confidential by one partner or even by a supplier of measurement equipment.

8.2.7 Backup

An FOT database backup strategy should be based on “acceptable downtime”. Off-site backups are mandatory for managing a disaster scenario. The majority of the data is never edited (video and raw data in the database) and data mirroring should be sufficient. The backup policy must be based on the time it takes to recover data and the acceptable loss

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

8.2.8 Risk management

An FOT study can generate huge amounts of data; especially when video is used, and the management must decide on the needs for backup and acceptable downtime for recovery of the FOT data. It is up to the steering committee of the study to have a documented backup policy and crash recovery strategy. Further, the backup strategy might need to vary during the lifecycle of the study (collecting phase, analysis phase). If so, each phase and strategy should be documented. Disaster recovery (when local database and backup hardware are destroyed) strategy must be taken care of and there should be an offsite backup of the data.

It is vital to keep track of where data is created and manipulated. A copy of the collected raw data should be kept as original data in a read-only space to prevent accidental data loss. In addition, it might only be the person responsible for data uploading who can insert new data and the FOT database owner who can delete data.

8.2.9 Database acceptance

Before an FOT is launched, the FOT database architecture should be reviewed by a system evaluator to ensure that all requirements are fulfilled and to verify policy documents.

8.3 Off-line quality management procedures

8.3.1 Quality assurance of objective data

Before data is uploaded to database

Before uploading objective data from a vehicle, a well-defined algorithm should be applied to all the data in order to verify data consistency and validity.

Quality assurance of video data

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

Driver ID verification/input

Again, it may be necessary to have a process that allows the analysts to view, for example, one image per trip and match this with the IDs of the drivers allowed to drive a specific vehicle. If a driver is unknown, then the data for a particular trip may have to be neglected. A software tool for doing this manual identification of drivers is preferable. Be advised that some eye trackers (if available) provide driver ID functionality.

8.3.2 Quality assurance of subjective data

In order to address the validity of the data, the formulation of the questions (and possible answers) is a key issue, especially when designing a questionnaire to be distributed to respondents. Questions must evidently be formulated in a clear and unambiguous way. In addition, questions must also, for example, be specific, not too complicated, and be formulated in simple terms that can be understood by the interviewee. Hypothetical questions are the most difficult and should be avoided.

Regardless of the data source, missing data is a threat to data quality (see [Section 9.8.1](#)). In the case of a missing questionnaire, efforts must be made to ensure that data collection is as complete as possible and reminders must be administered. Furthermore, the number of questions should be thought through in order to limit it. The number of open questions should be as small as possible, to reduce the effort required of respondents. The interviewer plays an important role in collecting data in an interview situation. Interviewer bias, that is the influence of the interviewer on the respondents' response, can be avoided by administering a questionnaire. However, the interviewer may also increase the quality of the data collected, e.g. by answering questions and using probing questions.

8.3.3 Measurement naming guidelines

It is recommended that the FOT project decides on and adheres to a set of naming conventions for measurements. The strategy used should be well documented and thoroughly enforced. Motivations for a clear naming convention include: 1) project-wide consistency, 2) clarity for direct understanding of used measures in analysis, 3) differentiation of non-comparable measurements, and 4) avoidance of confusion.

When specific measurements are named, references to the following measure attributes are recommended: *indicative name*, *associated source*, *sample rate*, and any other FOT *specific descriptor*. The compounds should be joined consistently to create a single word. Possible strategies are: camel case (SomeSignal), underlines (some_signal), or hyphens (some-signal). Depending on the context and FOT-specific requirements, all or only a subset of the compounds can be used.

Examples: [GroundSpeed_GPS_1Hz], [GroundSpeedGps1Hz]

The aim is to clearly understand what a measurement “is”, where it comes from, and how it relates to other measurements. To avoid the risk of making faulty comparisons, measurements that are *non-comparable* should be named *differently*.

8.3.4 Automatic pre-processing and performance indicator calculations during data upload

It is recommended to define procedures and implementation schemes on *how* to add calculation of pre-processing and in performance indicators in the upload process (see [Chapter 9](#)). These calculations should preferably be read-only for the users. The actual algorithms for the pre-processing and performance indicator calculations in this step have to be well defined and tested (on e.g. pilot test data), or based on previous experience. Since the estimation of some specific performance indicator may set specific requirements on the raw data (see [Chapter 5](#)) these constraints have to be taken into account when implementing the automatic pre-processing.

8.4 Data analysis tools

The focus of this section is to describe analysis tools, not analysis procedures or methods.

8.4.1 Data classification/transcription

The exact coding scheme/syntax for events (time segments) will vary widely across FOTs. However, the following features have been identified as important software functionalities:

- Organising or categorising subjects into groups and subgroups
- Defining a set and structure of codes
- Editing or updating the coding scheme
- Defining events as either a *state* event (e.g. glance left, glance right) or a *point* event (e.g. stop light)
- Defining whether state events are mutually exclusive or start/stop and set a default state
- Defining whether codes are nominal (e.g. road types) or rating scales (e.g. observer ratings of drowsiness)
- Defining whether codes are compulsory or optional, logging freely written comments created by a human analyst (no coding scheme)
- Support for interrater and intrarater reliability analyses: how consistent different individuals are at measuring the same phenomenon (e.g. annotating the same video) and how consistent an individual is, when measuring the same phenomenon.

8.4.2 Time history and event analysis with video

This section describes the basic functionality of tools for viewing numerical time history data and the associated environment sensing data which includes video data, map data (e.g. GPS), and traffic state data (e.g. radar).

Recommended functionalities for visualisation and interaction with data

It should be possible to replay single-participant data (numerical time-history data, video data, map data, and traffic state data) simultaneously. Multiple windows for different plots and illustrations provide maximal flexibility to arrange and resize, often spread out on multiple computer screens. Recommended visualisation functionality:

- Video recordings synchronised with other raw data plots
- Continuous variables and performance indicators which can be plotted (and zoomed) on graphs
- General information (FOT reference, subject ID, event lists, etc.).

For visualisation of collaborative driving functions, it can be necessary to aggregate and visualise multiple participants' data at once to compare event flows.

Recommended functionality to support data analysis

Being able to query an FOT database is a basic requirement. In many cases, statistical analyses are performed using only processed indicators and summary data (e.g. event and trip summaries). However, for refining such indicator processing or visualising a single event to understand its details, analysts need to be able to view time history data in an easy way.

Other requirements for functionality to support data analysis include:

- Signal processing of numerical data (see also [Chapter 9](#))
- Fully customisable mathematical computation, analysis, and algorithm development functionality
- Automated 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)
- Being able to export results to tabular format or statistical packages.

Some software tools may require a fairly high level of knowledge to use, so it may be advantageous to develop FOT-specific configurations for them or even proprietary easy-to-use graphical user interfaces.

8.5 Database usage

Further usage of the database in other FOTs and analysis projects should be considered. Data acquisition should also take into account this potential further use: selected additional data (e.g. traffic information or extra vehicle signals), potentially useful for foreseen analyses, could be stored for reducing time and cost constraints for upcoming research projects. Further discussion on data sharing and minimum recommended datasets can be found in the [Data Sharing Framework](#) (FOT-Net Data and CARTRE, 2019).

8.6 Databases and analysis tools for FOTs of automated driving

The rapid development of AD functions is matched by the transformation of the IT industry. IT infrastructure that previously was almost certain to be hosted in-house are now moved to cloud or hybrid solutions. The drivers are flexibility, scalability and cost both on storage but also from data processing perspective. For the specific case of simulations, development and training of AI algorithms, there has been a clear shift from CPU- to GPU-based processing.

The fact that it is costly to store all raw data, from all tests, with the original resolution, re-introduces the concept of data reduction. Lidar or other point clouds and video data consume large volumes of data storage and this needs to be addressed by techniques on how to reduce data. The techniques include down-sampling, identifying segments of interest for the specific analysis, and/or converting rich data into structures (e.g. point clouds to objects, video to annotations).

Relational databases are not considered state-of-the-art when hosting data from automation pilots. The benefits of precise data access are less important when it comes to processing large datasets from tests of AVs. Requirements such as easy replay of log data, development and validation of event detection algorithms, and use and development of AI algorithms have challenged relational databases and instead distributed or flat file systems are favoured. However, after the raw data has been finally processed into aggregated indicators and spreadsheet-like lists of events, evaluation can continue in databases and statistical tools.

The driver identification is still valid for AV pilots. It must be clearly recorded who is actually driving at any given time; the driver or the AD function. Also, for future projects it may be of interest to study the passengers' activities in an AV. If so, the passengers must be identified and consent must be given.

9 Data Analysis and Modelling

9.1 Introduction

The strategy and the steps of data analysis need to be planned in order to provide an overall assessment of the impact of a system from the experimental data. Despite the appearance of data analysis as one of the latter stages of FESTA V, experts have stressed that planning for the later analysis work should be done from the start, and analysts need to be involved in the research questions, data needs, and in planning of the experimental design to ensure the integrity of the study. Data analysis is not an automatic task limited to some calculation algorithms. It is the place where hypotheses, data and models are confronted. There are three main difficulties:

- The huge and complex amount of data coming from different sensors including questionnaires and video, which needs to be processed
- The potential bias about the impact of the system(s) on behaviour, which may arise from sampling issues including the location of the study, the selection of a relatively small sample of drivers, etc.
- The resort to auxiliary models, such as simulation models, to extrapolate from the behavioural effects estimated and tested within the sample to effects at the level of the whole transport system.

To be confident of the robustness of the outputs of the data analysis, one has to follow some strategic rules in the process of data analysis, and apply to the whole chain and to its five links (Figure 9.1) the required techniques such as appropriate statistical tests or data mining to uncover hidden patterns in the data.

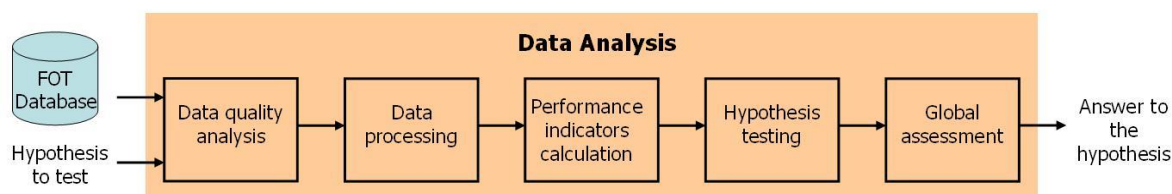
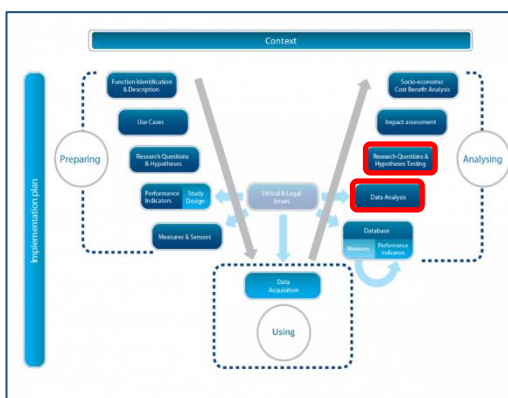


Figure 9.1 Block diagram for data analysis

Some specific actions are required to tackle the difficulties mentioned above and to ensure the quality and robustness of the data analysis:

1. A pilot of the study is a prerequisite to check the feasibility of the chain of data collection and evaluation and to achieve a pre-evaluation of the usefulness of the system.



2. The data flow has to be monitored in detail but also overall. One of the strategic rules to follow is to ensure local and overall consistency in the data processing and data handling and analysis.
3. The sources of variability and bias in the performance indicators have to be identified, where feasible, in order to control for them in the data analysis.
4. There is a crucial need for an integrative assessment process which should ideally combine within a meta-model information gathered on the usability, usefulness and acceptability of the system with the observed impacts of the system on behaviour. The estimated effects obtained from the sample of drivers and data have to be extrapolated using auxiliary models to scale them up.
5. Appropriate techniques have to be applied for each link of the chain – data quality; data processing, data mining and video analysis; performance indicator calculation; hypothesis testing; and generalisation of results. The techniques come from two set of statistical and informatics tools belonging to two main kinds of data analysis: exploratory (data mining) and confirmatory or inferential (statistical testing).

9.2 Large dataset handling

An FOT often collects so much data that there are not enough resources and time to analyse all of it in the timeframe of the FOT project. There are different choices when it comes to selection of data for analysis.

An option is to collect all data available, because an FOT may provide a unique opportunity (and funding) to collect data which may be hard to collect later on. This approach gives a rich dataset that enhances the probability that the data is useful also in future projects. However, before starting data collection, it is recommended to develop a plan on how to store the data after the original project and how to make it available for later analysis or analysis by others. This plan should specify detailed data dictionaries, open software formats, rules for data access and other relevant information such as metadata.

Analysis later on and by others (in other words, re-using data from other projects) is a good idea, thus reducing the need for an expensive and time-consuming data collection phase. The researchers should, however, be aware that data may become outdated because traffic, vehicles, driver support and information systems change. Therefore, data collected today may not be of much relevance in 10 years' time for certain analyses. However, although the context may change, the fundamentals of driving behaviour do not. Therefore, whether it is possible to re-use data fruitfully depends on what is wanted to be known. New projects should be aware that sponsors and stakeholders may want to have fresh data, but the interest for global re-use of data is increasing. The opposite approach to the "all data" approach is to collect only the minimum set of relevant data needed to answer the research questions in the project or to trigger data collection for the specific events of interest. Limiting data to specific events may have the consequence that it is not possible to look at generalised behavioural side-effects. Selection of data should be driven in the first place by the research questions that need to be answered. With limited resources it may be useful to find a compromise between an explorative study with naturalistic driving

and a stricter experimental study in which the expected behaviour of drivers and systems are evoked in a more condensed manner, requiring less time and providing more focused data. Usage of this selected data for other purposes and projects might not be feasible, as the selected data has been collected for certain research questions. Even for later analysis, the specification of the relevant data can be changed (e.g. threshold for an event) because of new findings within the analysis. An adaptation of such selected data will not be possible, because of missing data.

To make analysis more efficient, it is recommended to take a layered approach to data analysis, making sure that first the data is selected that is needed to provide information on the research questions before going into a detailed analysis. Moreover, it needs to be checked whether the selected data is appropriate for performing the analysis before actually starting.

It should be kept in mind that, very often, less data with higher quality is more, since in the end all data in the complete dataset often cannot be analysed due to

- Delays
- Missing data
- Poor data quality
- Budget restrictions
- Limited time
- Restricted access to gathered data in the database.

The lack of resources to analyse all data is usually the lack of human resources, and not a problem of computational resources. Thus, methods for automation of the analysis are needed in order to increase especially the processing of data (e.g. recognition of events). The analysis of video data is generally a time-consuming task, which should be considered from the beginning with respect to planning. Data mining methods are important to tackle this problem. An additional problem with resources is that data analysis comes late in a project. If delays occur in the data collection phase, which is often the case, the phase of data analysis may have to be shortened and resources will be diminished. It is therefore important to plan the data analysis from the beginning of the project.

The processed data for analysis is generally stored in databases. The performance of databases decreases with the amount of stored data. Thus, intelligent approaches on data storage need to be applied to avoid unnecessary processing time. Datasets for the analysis may be defined in advance as part of the data acquisition scheme and then processed before they are stored in databases.

9.3 Consistency of the chain of data treatment

There will be a lot of computations and data flows, starting from measurements collected into the database, through calculation of performance indicators (PIs), to the testing of hypotheses and on to impact assessment. This process, in the form of a chain of operations, has to be monitored in detail but also overall. There are five operations linked together in terms of data treatment. In addition, three kinds of models are needed as support to carry out the three top operations: probability models for justifying the calculations of the PIs, integration models to interpret in a systematic way the results of

the test, and auxiliary models to assess the effects on a larger scale (scaling up). Moving from the data to an overall assessment is not only a bottom-up process; it also has to include some feedbacks (Figure 9.2). There are two movements along this chain: a data flow going up and a control feedback loop from the top which concerns the consistency of the evaluation process and which mainly depends on control of uncertainty.

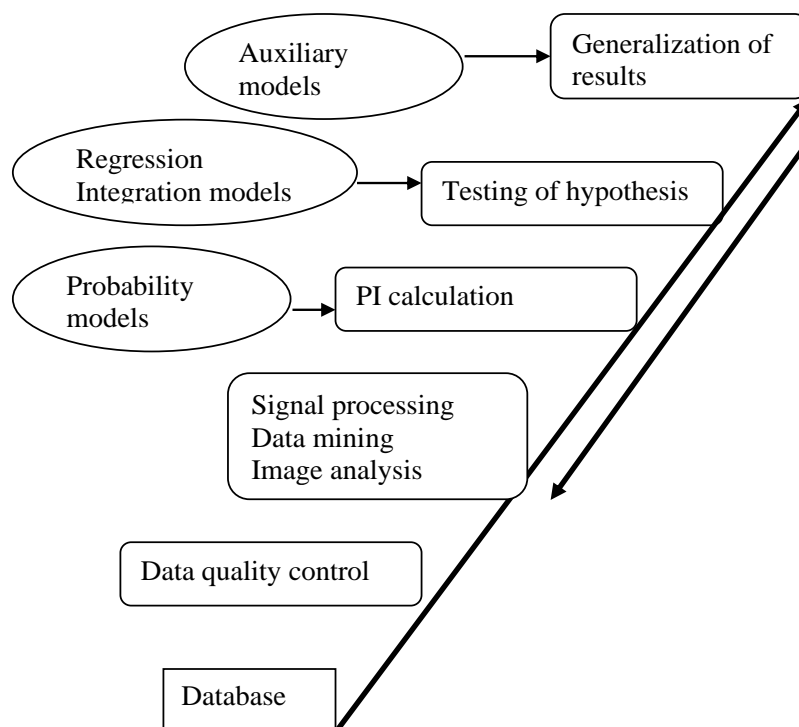


Figure 9.2 Deployment of the chain with feedbacks and additional models

In moving up the chain, the consistency of each operation can be checked locally according to the specifications, which are governed by the nature of the PIs that correspond to a set of hypotheses related to the use cases of the system. For each PI, there are some rules which ensure the validity of the calculation procedures. For example, it is important to sample data which can change rapidly at a high data rate. The sampling rate must fit the variability of the variable.

As a complement to local consistency, a general criterion is to have sufficient sample size to get enough power to carry out the test of a hypothesis or to make an overall assessment with enough precision. This is a feedback loop coming from the top to control the uncertainty of the estimations. The precision required for measurements depends on the uncertainty of the auxiliary models, of the regression models and of the probability models.

9.4 Precision in sampling

The aim of the FOT is to measure the effect of an intervention or treatment – the use of a system or systems – on a sample of subjects and in various driving situations while controlling for external conditions. From the sample, we have to infer the effect on the population by aggregating the values obtained through the sensors without and with the system to get an estimate on the effect on the chosen PI. How do we ensure that this inference is valid, in other words that the estimation is very near, and variance which could

be combined to get a measure of the sampling error (Wannacott and Wannacott, 1990)? To control the bias and variance, one has to rely on a well-defined sampling plan using appropriate randomisation at the different levels of sampling: driver, driving situations and measurement.

Consideration should be given to identifying the possible sources of (unintended) bias and variance in the sample, and either attempt to minimise or account for these in the data analysis. This is one of the most fundamental principles of statistical methods.

1. **Driver variation.** The simple fact of the matter is that drivers vary. The range of behaviours that drivers exhibit (in terms of speed selection, headway preference, overtaking behaviour) is immense, but fortunately the variation obeys some probability laws and models. Strict randomisation procedures ensure that only the outcome that is being varied (or the outcome whose variation we are observing) is working systematically. However, strict randomisation is not usually possible or desirable in an FOT, particularly when the sample sizes are relatively small. It may not be desirable, for example, to waste sample size by recruiting drivers who only drive small amounts each week. Many FOTs have for good reasons used a quota sampling procedure, in which equal numbers of (say) males and females are recruited. This can create bias when scaling up the observed data to estimates of effects at a national or European scale. The theoretical best method is to stratify the population of drivers according to some variables or factors related to the outcome and to sample proportionally to the size of the sub-population and to the a priori variance of the outcome (e.g. speed choice). For practical reasons, a different sampling or selection procedure may be followed. In either case, it is important to be able to compare the sample to the overall driver population in order to identify what are the main discrepancies and to assess possible sources of bias.
2. **Driving situation variation.** There will be variation within and between the journeys and the driving situations within these journeys. For example, a particular journey may be affected by congestion part-way through, or weather conditions may change from day to day. This type of variation cannot be controlled and is considered to be random. The observation period should be sufficiently long to allow for these random effects. One example here is that seasonal effects should be considered.
3. **Measurement variation.** Once in a driving situation, by means of the sensors, we get a series of measurements at a certain frequency. Their size is not fixed but varies. Each set of measurements within a driving situation constitutes a sample of units taken from a cluster, according to sampling theory. Usually, there is a correlation between the measured outcomes. The information coming from this sample of measurements is not as rich as expected from an independent sample. One such cluster is at the driver level – the data collected from one driver is not independent.

How do we quantify the variance of the estimate of an outcome from the experimentation taking into account these three sources of variations? The total variance of the average of the indicator on the sample breaks down into an inter-individual, intra-individual and intra-situational variance. If the inter-individual variance is strong, an increase in number

of situations observed and in the measurement points per situation will not bring any precision gains (Särndahl et al., 1992). However, it may help to ensure a reduction in bias from e.g. seasonality.

9.5 Pooling of data

To get more reliable results, it is beneficial to pool data from different sources. This may include data from different drivers, different locations at a single test site, several test sites of a single FOT and data from several FOTs.

To keep the process transparent, pooling is recommended to be made stepwise. How to construct these steps depends on the data and tested function. For example, pooling can be done by analysing the data from a single location, then combining data from all locations within the same test site with the same speed limit, analysing the data pooled over the whole test site and over all test sites with the same speed limit, and in the end pooling the data from all test sites. Utilising this stepwise approach, local or condition-specific impacts can be separated from more general ones.

The benefit of letting one analyst (or team of analysts) analyse data from multiple sources also supports a harmonised approach in assessment. As analysis consists of many little choices that analysts make along the way, there is a risk that different datasets have been analysed with different assumptions if only the results of several FOTs or analyst teams are pooled. Therefore, the analysis of a pooled set of data at a single analysis site is more secure from this perspective.

9.6 Requirements for integration

Having treated and aggregated the data by means of statistical models, there are two kinds of problems to solve, related first to synthesis of the outputs and second to scaling up of the results from the sample to a larger population. Integration of the outputs of different analysis and hypothesis testing requires a kind of metamodel and the competences of a multidisciplinary evaluation team (Saad, 2006). Scaling up relies upon the potential to extrapolate from the performance indicators to estimates of impact at an aggregate level.

It is often necessary to employ quantitative models from previous studies to estimate the effect of an indicator in question. It is, however, important to note that individual models have usually been developed for particular purposes, from particular data and with specific assumptions. However, in the absence of appropriate models available for the purpose of study, it is usually necessary to apply the “least bad” model available with appropriate weighting or adjustment.

It is also important to consider the constraints, assumptions, and implications behind the design of the study in mind when interpreting the analysis results. Behavioural adaptation may lead to side-effects (i.e. indirect effects) and also result in a prolonged learning process. However, the study period may, for practical reasons, not be sufficiently long to fully explore this.

Extrapolating from the sample to the population depends on the external validity of the experiment. The power of generalisation to the population of the estimates of impact is

related to their precision, which is composed of two parts – bias and variance. We can use three approaches:

1. If the required performance indicator is available in the sample (e.g. if journey time is an indicator of choice for efficiency, and journey time has been collected), the impact at the population level can be calculated directly, although sometimes a correction factor or other form of extrapolation adjustment may have to be introduced (Cochran, 1977).
2. If neither a performance indicator nor a proxy indicator is available, then it is necessary to adopt an indirect approach through models which provide an estimate of the output from the behavioural performance indicator estimated from the sample. Speed changes can be translated into changes in crash risk by applying statistically derived models from the literature which have investigated the relationship between mean speed, speed variance or individual speed and crash risk. Emissions models can be used to calculate the instantaneous emission of a car as a function of its recorded speed and gear selected.
3. Finally, a macroscopic or microscopic traffic simulation model can be applied to translate the effects observed in the sample to an effect on the traffic network. The outputs from such a simulation can be used e.g. to calculate journey time effects or fuel consumptions effects at the network level.

9.7 Scaling up

Scaling up refers here to a process where findings from an FOT, such as certain safety improvements, are considered against what would happen in the whole society, e.g. EU-wide, if the tested functions would be gradually reach widespread use. This process usually involves many steps such as considering time factors of take-up and related future scenarios. It is also necessary to clarify where the change would happen, such as for what types of environments, processes and statistics. Unfortunately, accurate statistics of all such topics is not usually readily available.

The calculations of scaling up could become rather direct, if causality between certain types of changes and their impacts has been established in previous work and is supported by statistical studies. Also, when there exist detailed simulation models of similar environments or processes – as well as statistics on how many similar cases exist country-wide.

It is good to acknowledge, that most of the time scaling up must rather be based on indirect methods such as partial models, incomplete statistics and expert assessment. In case it is very much based on indirect and incomplete evidence, scaling up can become somewhat speculative. Therefore, the steps of scaling up need to be kept transparent for the readers, so that the assumptions made and the references are clear and could be refined, later, with new or more accurate information. The methodology for scaling up needs to be systematic and scientific.

For scaling up, the representativeness of the FOT and its sampling are very important. For example, if the FOT was carried out with mainly male participants between the ages of 25

and 45, the results can in principle not be extrapolated directly to the whole population of drivers. Or, if most test rides took place outside peak hours, only in summer time and only in suburban or motorway environment, the results may not be scalable for all traffic, road and weather conditions. However, getting a representative sample of the whole population and of all traffic conditions is impossible. Therefore, it is acceptable to have an imperfect sample, as long as the limitations of the sample are known and they are described together with their implications in the result report. It is still desirable to try to have a sample as representative as possible. Another common challenge for scaling up is related to data only representing small penetration rate (e.g. in cooperative or automated systems FOTs). Specific considerations are needed in those cases for extrapolation of impacts to larger penetration rates.

There are, in principle, two types of methods for scaling up. The first is scaling up by using statistical data. Here the results from assessment, where different positive and negative effects and their magnitude have been considered, are reflected against available statistics. For example, if a certain accident type could now be avoided thanks to a new function, how many are these accidents yearly.

The second method is scaling up by using detailed models of the impacted systems. In FOT context, the modelling usually refers to microscopic or macroscopic multimodal traffic models. Such models can provide estimates e.g. on what would be the frequency of a situation in traffic, after system introduction. In large-scale FOTs, some incidences might directly visible from log data. Still, e.g. accidents are very rare in successful tests and thereby simulations of millions of dangerous scenarios are likely necessary to clarify the elements of most critical cases and their frequencies in traffic.

These methods described in the following explain how scaling up can be applied theoretically. In practice, scaling up is often a big challenge. It is important to consider the goal one wants to achieve and datasets and tools available to make a feasible scale up plan.

9.7.1 *Scaling up using statistics*

The scaling-up method using statistics initiates from the impacts measured in field experiments on e.g. CO₂ emissions at a local level as distinguished for different situations (such as traffic state, vehicle type, etc.) coming from the FOT. The definition of situations depends on:

- The system characteristics
- The situational variables that are expected to have the largest impact (e.g. a night vision system will only be active during driving in the dark)
- The possibility of measuring the different situations and the model capabilities.

Statistics on how frequent is the condition for which the result is (which part of accidents, mileage or CO₂ is linked to the scenario) needed on the large-scale level that is targeted. Then, the impact on a local scale is scaled up using these statistics (for example kilometres driven on different road types) under the specific situations like weather conditions or driving scenarios.

Scaling up using statistics is applicable when interaction and second-order effects (i.e. latent demand induced by improvement of the service level, caused by a system) can be expected to be insignificant, or when there is a clear effect at certain traffic situations for which data on a higher level is available, or even at the mere event that no appropriate macroscopic model is available to perform the model-based methodology. A drawback of this method is that a clear indication of traffic flow level impact needs to be measurable in the experiment and datasets need to be available for the countries one wants to scale up to, and at present there is very limited measurement data for some countries in Europe and statistics do not cover all parameters relevant for scaling up.

9.7.2 Scaling up using (multimodal) traffic models

Traffic modelling can be used either directly to scale up impacts if macroscopic model of the target network is available/used. It can also serve as an intermediate step between (small scale) field experiments and regional statistics.

In this first approach, the network of a macroscopic (multimodal) traffic model determines the level on which the results are calculated. Ideally, the model is available at country or EU level. Scaling up using such a model can be done in two different ways:

- The calculation of the impact is done with a model other than the macroscopic traffic model. The local effects of the system are in this case determined (e.g. via a microscopic simulation tool). These effects can be used as input for the macroscopic model at city/region/country/EU level. One run is performed for deriving the direct effect to a larger scale.
- The calculation of the impact is performed directly with a macroscopic traffic model. In this case, should the model be at the required level (city/region/country/EU), the direct effect of the system is calculated. This can be done performing a run of the macroscopic model. A limitation to this approach is that microscopic effects of ITS cannot be taken into account, e.g. changes in driver behaviour. Therefore, it can only be used to determine the effects of ITS that mainly affect macroscopic mechanisms in the network, such as mode or route change. Additional limitation is the lack of available macroscopic models for large regions.

In the second approach, the results of microscopic simulation model serve as intermediate step to convert first e.g. the local single vehicle level results to flow level results in different traffic scenarios. Finally, statistics can support the final scaling up if a macroscopic model of the whole region is not available.

Optionally (for both cases), the economic effect can be calculated with an appropriate model. Then a second run is performed with the macroscopic model to account for the second-order effect.

Scaling up using a macroscopic model is a good method to apply when second-order effects are expected and/or when the effects of the ITS system can be used directly as an input parameter for the macroscopic model. Also, this method can be used only if such a large-scale model is available. Being a more elaborate method than scaling up using statistics, it allows taking into account specific circumstantial differences especially if there are interaction effects. A downside of scaling up with a macroscopic traffic model is that

urban roads are usually not part of the network on such a large scale, and also that it requires more effort than scaling up using statistics. Thus, in practice a combination of modelling and statistics can be a practical solution.

9.8 Appropriate techniques at the five links of data analysis

The five links follow the right branch of the development process of an FOT, from data quality control to generalisation of results. Different techniques of data analysis and modelling which could be used at each step are presented here.

9.8.1 Step 1: Data quality analysis

Data quality analysis is aimed at making sure that data is consistent and appropriate for addressing the hypothesis of interest. . A general data quality analysis is also important if the data is to be re-used after the project, as the researchers might not have been involved in the collecting project and are not aware of the data quality issues. Data quality analysis determines whether the specific analysis that the experimenter intends to perform on the data to address a specific hypothesis is feasible. Data quality analysis can be performed by following the four sub-steps reported below (and shown in Figure 9.3). A project report detailing the quality of the data to be used to test the hypothesis of interest should perhaps be created.

The sub-steps for data quality analysis are:

- **Assessing and quantifying missing data** (e.g. percentage of data actually collected compared to the potential total amount of data, or frequency of changes in signal values)
- **Ensuring that data values are reasonable, within expected range, and units of measure are correct** (e.g. a mean speed value of 6 may be unreasonable unless speed was actually recorded in m/s instead of km/h)
- **Checking that the data dynamics over time is appropriate for each kind of measure** (e.g. if the minimum speed and the maximum speed of a journey are the same, then the data may not have been correctly sampled)
- **Guaranteeing that measures' features satisfy the requirements for the specific data analysis** (e.g. in order to calculate a reliable value for standard deviation of lane offset, the lane offset data segment should be at least 10 s long; additionally, this time length may depend on the sampling rate (Östlund et al., 2005)).

Note that the first three sub-steps refer to general quality checks; thus if any of these fails, data analysis cannot proceed. If a failure is encountered, it should be reported to those responsible for the dataset, so that the possible technical error behind it can be tracked down and solved. However, the last sub-step is different, and is related to the specific analysis or to a specific PI to be used in the subsequent data analysis. As a consequence, if step 4 fails, it may not be due to a technical issue that needs to be solved, but to intrinsic limitations in the collected data.

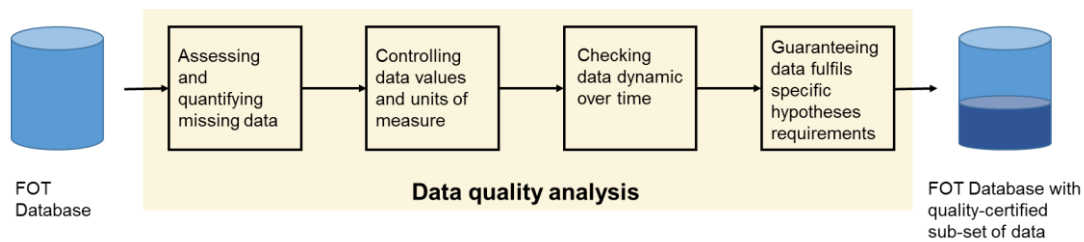


Figure 9.3 Block diagram of data quality analysis

Data quality analysis is handled differently with regard to data from in-vehicle sensors (generally CAN bus data and video data) and subjective data (generally from questionnaires). Subjective data, once collected, is hard to verify unless the problem stems from transcription errors.

Once basic data quality has been established, the next step in data analysis is data processing. Complex quality issues may present themselves also at this and later steps, when calculations using several inputs do not produce plausible results. Therefore, comprehensive data quality checking benefits from piloting most processing and evaluation steps.

9.8.2 Step 2: Data processing

Data processing aims to “prepare” the data for addressing specific hypotheses which will be tested in the following steps of data analysis. Data processing includes the following sub-steps: filtering, deriving new signals from the raw data, event annotation, and reorganisation of the data according to a different time scale ([Error! Reference source not found.](#)). Not all the above-mentioned sub-steps of signal processing are necessarily needed for all analyses. However, at least some of them are normally crucial.

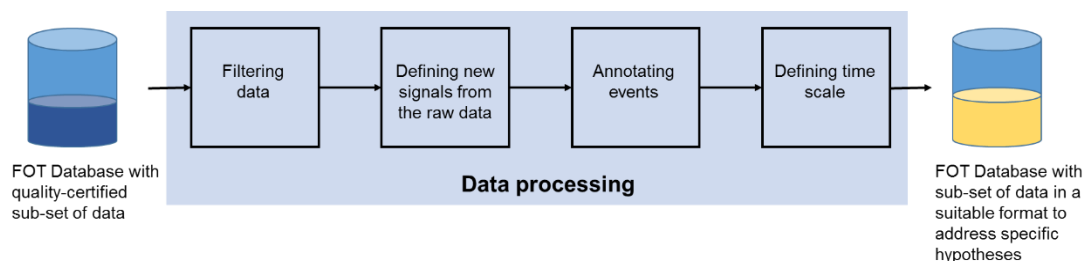


Figure 9.4 Block diagram for the procedure of data processing

Data filtering can involve a simple frequency filter, e.g. a low-pass filter to eliminate noise, but also any kind of algorithm aimed at selecting specific parts of the signals. Very often a new signal more suitable for the hypothesis to be tested has to be elaborated by combining one or more signals. Marking specific time indices in the data to identify events of interest is fundamental to individuate the part of data which should be analysed. Ideally, an algorithm should be used to go through all FOT data and mark events of interest. However, especially when the data to be annotated is from a video and requires the understanding of the traffic situation, writing a robust algorithm can be very challenging even with advanced image analysis techniques, and manual annotation from an operator may be preferable. Re-organising data into the most suitable time scale for the specific hypothesis to be addressed has to be considered in the following steps of the data analysis. It is

important to note that also this derived, calculated and annotated data should be documented properly in the same way as the collected data, to allow researchers during and after the project to understand how this new data was created.

9.8.3 Step 3: Performance Indicator calculation

There are five kinds of data which provide the PIs: direct measures, indirect measures, events, self-reported measures and situational variables. The scale of the dataset and the uncontrolled variation in driving situations that occurs from driving freely with vehicles become a seriously limiting factor unless an efficient calculation methodology is implemented. The choice of which performance indicators to calculate is clearly dependent on the amount of effort required. Efficient calculation methods need to anticipate that:

- (a) PIs will be calculated on imperfect data – there is a strong need to create special solutions for “exceptions to perfect data”
- (b) PI calculation requires situation or context identification – a “denominator” or exposure measures to make a measure comparable is required to determine how often a certain event occurs per something (e.g. km, road type, manoeuvre).

The fact that test exposure is largely uncontrolled (not tightly controlled as in experiments) means that analysis is largely conducted by first identifying the important contextual influences, and then performing the analyses to create a “controlled” subset of data to compare with.

The ability to find and classify crash-relevant events (crashes, near-crashes, incidents) is a unique possibility enabled by FOTs to study direct safety measures. As reported in [Section 5.4](#), this possibility should be exploited by using a process of identification of critical events from review of kinematic trigger conditions (e.g. lateral acceleration $> 0.2 g$). The definition of these trigger values and the associated processes to filter out irrelevant events are of particular importance for enabling efficient analyses.

Care should be taken to use appropriate statistical methods to analyse the PIs. The methods used must consider the type of data and the probability distribution governing the process. Categorical or ordinal data, such as that from questionnaires, needs to be analysed appropriately. Data on the degree of acceptance of a system (e.g. positive, neutral, negative) can be applied in multivariate analysis to link it to behavioural indicators so as to create new performance indicators.

9.8.4 Step 4: Hypothesis testing

Hypothesis testing in an FOT generally takes the form of a null hypothesis: no effect of the system on a performance indicator such as 85th percentile speed, against an alternative such as a decrease of x% of the performance indicator. To carry out the test, one relies on two samples of data with/without the system from which the performance indicator is estimated with its variance. Comparing the performance indicators between the two samples with/without intervention is done using standard techniques such as a t-test on normally distributed data. Here the assumption is that there is an immediate and constant difference between the use and non-use of the system, i.e. there is no learning function, no drifting process and no erosion of the effect.

However, the assumption of a constant effect is often inappropriate. To get a complete view of the sources of variability and to handle the problem of serially correlated data, multi-level models are recommended (Goldstein, 2003).

With such models, drivers or situations with missing data have generally to be included. Elimination of drivers or situations because of missing data in order to keep a complete dataset may cause bias in the estimation of the impact.

It is assumed that data will have been cleaned up in the data quality-control phase. Nevertheless, to be sure that the estimation will be influenced minimally by outliers, one can use either robust estimates such as trimmed mean and variance or non-parametric tests such as a Wilcoxon rank test or a robust Minimum Mean regression (Gibbons and Chakraborti, 2003; Wasserman, 2007; Lecoutre and Tassi, 1987). Such tests provide protection against violation of the assumption of a normal distribution of the performance indicator.

9.8.5 Additional Step 4: Data mining

Data mining techniques allow the uncovering of patterns in the data that may not be revealed with the more traditional hypothesis-testing approach. Such techniques can therefore be extremely useful as a means of exploratory data analysis and for revealing relationships that have not been anticipated. The data collected in an FOT is a huge resource for subsequent analysis, which may well continue long after the formal conclusion of the FOT. One relatively simple technique for pattern recognition is to categorise a dataset into groups. Cluster analysis tries to identify homogeneous groups of observations in a set of data according to a set of variables (e.g. demographic variables or PIs), where homogeneity refers to the minimisation of within-group variance but the maximisation of between-group variance. The most commonly used methods for cluster analysis are k-means, two-step, and hierarchical clusters (Lebart et al., 1997; Everitt and Dunn, 2000).

Whilst data mining is an efficient and powerful tool for exploring large volume data and automatically detecting and identifying hidden patterns in the data, one may have to do several preparation tasks to make sure that the discovered patterns are useful and reliable. For the first criterion, one should be sure that the specification and configuration of the data-mining tool capture the analyst's needs (e.g. what does the tool search for?).

It is important to note that many data-mining methods are integrated with state-of-the-art computing facilities such as the Internet of Things (IoTs), Cloud Computing, and Big Data Analytics. What this means is that data mining is capable of analysing data from all available sources, and discovering some new information which may be personal, even though the original datasets were anonymised. Therefore, the researcher needs to make sure that the new information discovered is compliant with data privacy policy.

9.8.6 Step 5: Generalisation of results

This section deals with the issue of identification of models and methodologies to generalise results from a certain FOT to a road network, national or international level in terms of traffic safety, environmental effects and traffic flow. One problem when generalising results from an FOT is to know how closely the participants in the FOT

represent the target population. It is often necessary to control for usage, market penetration and compliance (the system might be switched off by the driver) and reliability of the system. The process of how to go from the FOT data to safety effects, traffic flow and environmental effects is illustrated in Figure 9.5. In this process, two steps need to be taken. One is scaling up the FOT results, for example to higher penetration levels or larger regions. The other is to translate the results from the level of PIs (e.g. time headway distribution) to the level of effects (e.g. effect on the number of fatalities). For each type of effect there are (at least) two different ways to generalise the results: through microsimulation or directly.

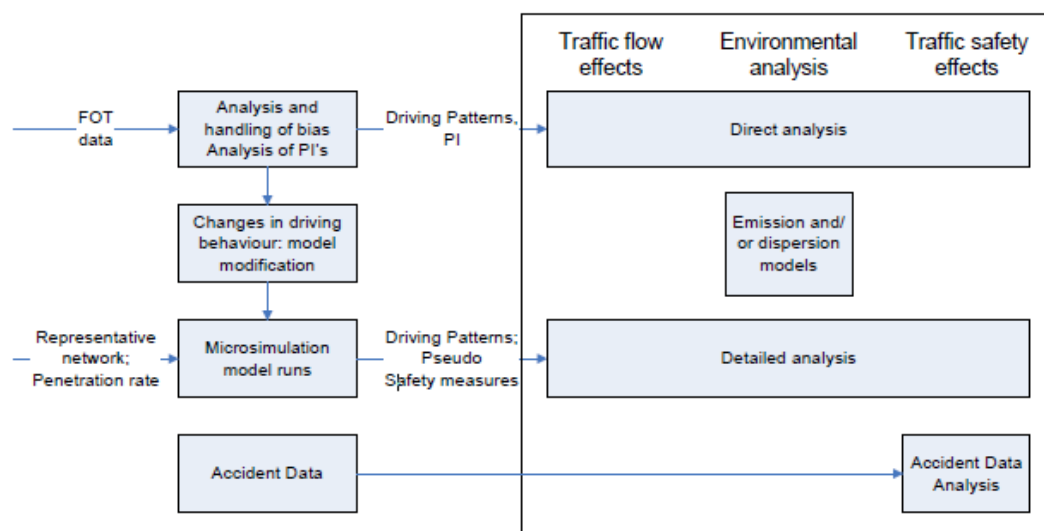


Figure 9.5 Block diagram of translating FOT indicators to large-scale effects

The direct route includes both estimation directly from the sample itself and estimation through individual or aggregated models. Some advantages of the direct route are that it could be rather cheap and quick, if underlying calculation models are clear and inherited from previous studies. The alternative is to use a traffic microsimulation model which represents the behaviour of individual driver/vehicle units. The advantages of microsimulation are that they can be more reliable and precise and can incorporate indirect effects (such as congestion in the network at peak times), the drawback naturally being that building and maintaining a simulation model for national or European scale requires substantial resources.

The assumptions used in both calculation methods should be made transparent for readers of the assessment results. Transparency could allow even repetition of the calculations using modified assumptions and values.

The simulation of indirect effects allows studying interaction effects between equipped and non-equipped vehicles, for example changes in traffic flow. Interaction is of major importance in any FOT. Effects due to interaction with other road users can be difficult to analyse directly from log data, especially beyond environmental sensor range, because data collection is mostly only carried out by means of the test vehicles. Hence traffic simulations are a useful tool to investigate indirect effects in different driving situations. The indirect effects are also of interest when testing connected vehicles. The interactions between connected and non-connected vehicles provide necessary information for the

assessment of safety and traffic flow effects. However, microsimulations for connected vehicles require further information to provide a model for the simulation environment: information on the communication range of the connected vehicles, information on when vehicles are connected to another vehicle, and integration of the communication infrastructure. Constructing the simulations might require more effort than originally foreseen.

Microsimulations also allow studying certain effects or functionalities before the data-collection phase starts, in order to get an idea of the function and its effect. Here different settings of the function can be tested. Tuning the function based on the simulations is also a development option. The simulations need input from real traffic in order to understand how certain situations are evolving. Especially data on the relevant network and the traffic density, e.g. whether the vehicle is driving in free flow or in a car-following situation, is needed for traffic simulations. But subjective data is of importance as well. For instance, the questions why certain drivers did not follow a certain function warning or advice might be interesting when implementing the driver model. Interviews with drivers can be an efficient approach to understanding certain driver behaviour patterns. Moreover, traffic simulations are an important tool for interpretation of the results or determined effects. By means of traffic simulations, more details on the relevant network can be observed and provided, in order to understand a certain behaviour or effect.

Since traffic microsimulation models consider individual vehicles in the traffic stream, there is consequently the potential to incorporate FOT results in the driver/vehicle models of the simulation. Impacts on the traffic system level can then be estimated through traffic simulations, including varying levels of system penetration into the vehicle population. Here it is of importance to provide a detailed driver model. The driver model becomes more realistic if required details are determined by means of data gathered in real traffic. This requires in addition the provision of a driver model for the baseline behaviour as well as the adapted behaviour in the treatment phase, due to the usage of a tested function.

Microsimulation does not necessarily yield the impact variable that is of interest. Various aggregated and individual models are necessary to convert e.g. speed to safety effects (e.g. via the Power Model, which considers the relationship between driving speed and the risk of an accident at different levels of severity). In addition, the modelling detail of traffic microsimulation places restrictions on the practical size of the simulated road network. Macroscopic or mesoscopic traffic models combine the possibility to study larger networks with reasonable calibration efforts. These models are commonly based on speed-flow or speed-density relationships. Large area impacts of FOT results can therefore be estimated by applying speed-flow relationships obtained from microsimulation for macro- or mesoscopic traffic modelling.

9.9 Data analysis in FOTs of automated driving

The main challenge for automated driving FOTs in data acquisition, data processing and data analysis is the establishment of automated processes. Because of the large amount of environmental sensor data and needs to analyse various driving strategies and manoeuvres in detail, manual analysis may not be feasible. Ideally, the data is automatically classified into common driving situations (e.g. car-following on a certain road type) and related events (e.g. small distance to another road user), which can be further

processed in terms of parameter analysis, correlation, etc. This detection of context or “driving scenarios” is enabled by the comprehensive sensing capabilities of automated vehicles and their detailed maps.

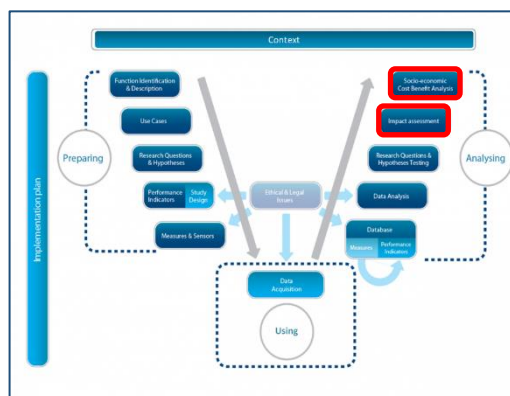
The automated classification of scenarios/situations needs a classifier in which each type is defined. In simple environments (e.g. a specified test route), such a rule-based classifier can be determined manually. In this case, rules are derived from environmental sensor and map data, tested and then applied in the classifier. Such a classifier would be able to classify at least simple cases, but correct classification on various road types remains a topic for research. By means of e.g. machine-learning techniques on large datasets, further classifiers could be developed and validated for complex traffic situations.

Automated video analysis also needs to be established. Automated vehicles produce large amounts of video data, which is time-consuming to analyse manually. In past FOTs, videos have mostly been viewed from the periods where other data (sensor, map or communication) has suggested an interesting event. In the future, automated video analysis tools will be able to determine relevant information directly from the video data and thus provide another important source of data.

10 Impact Assessment and Socio-Economic Cost-Benefit Analysis

10.1 Introduction

All FOTs require an assessment of impacts of the tested function or functions. Many times, especially if the FOT is supported by funding from the European Commission, the requirement is that the assessment should be performed at European level, also with details on the socio-economic evaluation.



The goal of this chapter is to provide concise advice on how to carry out an impact assessment. Many parts of the chapter will also be relevant to FOTs conducted at a national or regional level within Europe or to FOTs conducted outside Europe. A consistent methodology for carrying out this assessment for EC-funded FOTs will maximise the comparability of the results across projects.

Another goal is to address the possible breadth of impacts that can be considered and the available resources for carrying out the assessment, like references to examples of good practice. Although impact assessment comes at the end of the FOT chain, it should also play a role and be planned at the beginning of a project; indeed some of the choices made in setting up the FOT are linked to the Cost-Benefit Analysis (CBA). For example, choices about performance indicators and scaling up are directly linked to what can be analysed in terms of scope and impacts for the CBA. Furthermore, the expected impact and a feasible way to measure evidence of it is a major driver for the decisions to be made in the design of the FOT. People who are responsible for the impact assessment should be involved from the beginning of the project.

It should also be considered that policy makers expect concrete results from an FOT, such as the gain in terms of reduction of accidents or CO₂ emissions, and scaled-up costs and benefits (on a European level). However, these are results that may be difficult to assess and there can be various reasons why the FOT does not deliver these results in the end with the completeness or certainty that policy makers desire. Managing expectations is therefore important, throughout the project, explaining any apparent failures or uncertainties that will remain early, whilst emphasising lessons learned or any secondary benefits found. Understanding the strengths and weaknesses of the approaches taken is vital in this.

Our advice will be useful for a variety of parties: the organisations conducting the FOTs, including also the impact assessment specialist; the client commissioning the FOTs; and the consortia drawing up proposals for the FOTs. This chapter assumes that a “professional” in the area of impact assessment and of socio-economic cost benefit analysis will carry out the analysis. This individual we will refer to as the “analyst”. This information on impact assessment is not meant as a “tutorial”.

Broadly, there are three research angles where the evidence comes from:

- **Impact assessment studies** typically investigate the impacts of a system for a future time horizon. These prospective studies make use of an ex-ante impact assessment, often based on literature review, simulation work and expert estimation. They are often comprehensive in scope but they do not involve, or only to a limited extent, empirical data from real-life conditions.
- **Transport appraisal guidelines** or **scoping studies** in this area are very much focused on the appraisal part of the impacts. They dig deep into the methodology and practice of appraisal. They also involve proposals for standardisation of appraisal. Their detriment is that they are not developed for specific use in the field of safety evaluation.
- **FOT assessment studies** typically assess the impacts of one or more system functions. FOT evidence can lead to a quantum leap in the impact assessment because FOT produce measured data about direct effects. Therefore, the assessment can rely on ex-post measurement data.

All different research angles have their specific strengths and weaknesses. In preparing this guidance document, we found it useful to combine the strengths of the different perspectives.

10.2 Considerations

The impact assessment investigates the impacts of a technology on society. Ideally an impact assessment provides the decision maker with relevant information in a concise format. The relevant comparison is between the benefits and costs between a base case, e.g. a scenario without the ICT system (“without-case”) compared to those of the scenario with the ICT system (“with-case”). In preparing to carry out an impact assessment, the analyst is faced with making choices about the deployment scenarios, the geographical scope of the assessment and the analyses to be carried out. This section will go deeper into the issues surrounding these choices. The chapter will conclude with guidance on how to make the choices and carry out impact assessment and the socio-economic CBA.

10.2.1 Re-use of data in impact assessment

Impact assessment studies the impact that the full introduction or different penetrations of the tested system(s) have on a variety of impact areas such as driver behaviour, mobility, efficiency, safety and environment. Data re-use is often a prerequisite for comprehensive impact assessment in FOTs. Usually, an FOT focuses on a set of functions and use cases that it tests. The tests are usually carried out in a single country – or in a large-scale FOT project, a similar system is tested in a few countries. Nevertheless, the variety of testing environments and conditions and in people used as test drivers is limited.

Public authorities and the EU have an interest in the impacts of the introduction of a variety of ITS and services. They want to get reliable knowledge of impacts in different driving contexts all over the Europe. Such wide information needs mean that impact assessments can benefit from the (re)use of data from several FOTs. To achieve this, the study design of these FOTs should be similar enough and it must be possible to calculate the same key PIs.

The use of a pooled, larger dataset combining several FOTs supports reliable findings and makes it possible to generalise the results to a European (or global) level. When similar functions have been tested in many countries, a pooled dataset helps to cover impacts measured from a large variety of conditions and driver population. Pooled databases also effectively enable comparisons of driving styles across Europe, e.g. speed behaviour, or over driving conditions.

In addition to re-use of data, also the results of another FOT can be utilised (re-used) as data for impact assessment in another FOT. Especially, data utilised for scaling up is such that the detailed statistics and forecasts made should be re-used in other FOTs to avoid overlapping work.

10.2.2 Deployment scenarios

Predictions of the future, particularly over the medium or long term, cannot be precise. This argues for a scenario-based approach when developing forecasts of how future deployment of a system might turn out. This approach permits alternative scenarios to be evaluated in the CBA. Very likely one scenario will emerge as more favourable with the highest benefit-to-cost ratio, although that scenario may not be the most probable. The scenario analysis will also enable obstacles, to the pursuit of that scenario, to be identified. This in turn can identify public policy needs and other stakeholder requirements.

For the socio-economic impact analysis, deployment scenarios are required, whether they are just implicit or explicit. One approach is not to make any specific predictions but simply examine the socio-economic impacts of different levels of penetration. But even for such a simple set of scenarios, there has to be an assumed growth in penetration over a period of years. However, it is more appropriate to build up some alternative scenarios with different “futures”. There are a number of potential inputs into creating those scenarios:

- Policies of public authorities at a European, national, regional and local level. Policies and strategies for transport in general, road safety, environment, accessibility, traffic management and general ICT deployment are all relevant
- Other quasi-regulation such as Euro NCAP and standards
- Stakeholder plans – the strategies of OEMs, telecoms operators, road operators, large fleets and so on
- Likely developments in the various relevant markets, including costs, technical issues (such as synergies between different systems) and competitive pressures
- Public-private partnerships and their influence on the deployment process
- The attitudes and willingness to pay of the general public to identify potential purchasing and usage decisions.

Deployment can be pushed or even mandated by the public authorities, which can be termed “regulatory deployment”. It can be more voluntary, with a push from major stakeholders, e.g. by formal Voluntary Agreement. Or it can be purely market-driven, i.e. totally voluntary, depending entirely on the public’s willingness to purchase systems and use functions.

There are a number of tools for scenario development. At the most basic level there are forecasts on the growth of the vehicle market, changes in mobility and changes in the road network. Government strategies provide information about policy goals and targets. Stakeholder questionnaires and analysis provide further details on willingness to promote and invest. For the views of end-users, feedback from FOT participants on acceptance and willingness-to-pay is an important source of information and should be routinely collected in an FOT. More general information on public attitudes can be collected by means of focus groups, household surveys and stated preference studies. The last are useful as a tool to reveal trade-offs between alternative choices.

10.2.3 General issues of the socio-economic impact analysis

As at the hypotheses formulation stage, consideration needs to be given to the potential bundles of systems to be handled in the impact assessment. Indeed, there can be a large number of permutations of market penetration of different bundle sizes and not all can be covered, and some combinations of functions may be more likely than others. Some expert judgement has to be applied here, make reference back to market surveys, etc. More information about disentangling the effects of combinations of functions can be found in (Faber et al., 2011).

An impact assessment can only be as good as the data on which it is based. Hence those carrying it out should also be involved in the performance indicator definition stage in order to ensure that the relevant indicators are being collected. These may be the obvious indicators, such as speed, route choice etc., but there may be occasions where the FOT cannot provide the desired data directly and other methods (e.g. surveys, workshops etc.) may have to be applied. It is also important that the analysts are involved in planning of the study design to ensure that the experimental set up supports impact assessment work.

It is difficult to provide a definitive guide for conduction of an impact analysis. The process depends on the research questions set for the FOT project and on the tested functions. For example, in terms of scaling up, the area of interest may be a particular city, country or group of countries. For a function that operates e.g. on highways it makes sense to scale up according to those networks only. There is the added problem that scaling up data from one country to another may be inappropriate for a wide variety of reasons (driver types, weather, cultural aspects etc.) or not to exist at all. More empirical data and information and reliable statistics with sufficient level of details are needed to increase certainty on effects and conclusions.

A major step of the impact assessment is scaling up from single users to the traffic flow level by using simulation or other tools. Simulation can determine indirect effects (i.e. reduction of congestion due to less accidents, implications due to changes in interaction among users and non-user, etc.). Simulation tools generally require modelling of driver behaviour. This modelling relies not only on specific indicators which are being collected in the FOT from equipped vehicles, but also on information about the interactions of those vehicles with non-equipped vehicles and other road users. The modelling can be performed in micro-simulations, which are able to provide input to the impact analysis. These interactions between non-equipped and equipped vehicles should be addressed also in the hypotheses and in measurement processes. Often this can only really be

addressed by undertaking observational studies (video data analysis could be a possibility).

Current micro-simulation methods have their own limitations. The most generally used software packages do not properly represent vehicle dynamics in terms of interaction with the road surface. They therefore are not properly capable of covering lateral dynamics and collision avoidance manoeuvres of the vehicle. The networks covered are often geographically small and have often been created for purposes other than the evaluation of new vehicle-related technologies. They may not be very representative of overall national road networks, and origin-destination matrices (i.e. traffic flows) are generally lacking for night-time and weekend periods. In addition, the set of parameters that are used to set the driver behaviour is limited.

Business analysis in a private environment serves the same purpose as socio-economic impact assessment in the public world. It provides crucial information for use in the decision-making process on further steps towards deployment. Where socio-economic analysis tries to rule out double counts on the cost side, the business analysis focuses more on the feasibility of the concept and the roles and consequence for each of the actors required in the value web. The process of balancing the transfer of value (hardware, software, money, data, permits etc.) needs to be investigated, and usually there are a number of options for the business model (private, public-private and public only), each with a specific constellation of the roles of the actors involved and each with a specific risk pattern.

10.2.4 Assessment scope and process implication

At the start of the socio-economic assessment, a view will need to be taken on the scope of the analysis. Ideally the assessment would include all impacts of the system no matter how small, long-term or indirect that impact is: safety, mobility, efficiency and productivity, environmental, user acceptance and human factors, performance and capability, legal and implementation issues, and costs. However, setting such a broad scope for a socio-economic assessment will result in excessive long-term data collection and analysis in terms of expense and time. Some of them are also hard to assess, even though their potential existence can be acknowledged. Given that the purpose of the assessment is firstly to ensure that the implementation of the system is socio-economically beneficial, and secondly to aid the choice between alternatives, the scope of the assessment often can be narrowed by excluding minor or insignificant impacts, as long as the exclusion of these impacts will not bias the appraisal.

The same consideration goes with the costs. It may be hard to assess all possible societal costs, especially the indirect ones. Thus, the socio-economic impact assessment team should ensure that a fair comparison is made and acknowledge any limitations of the assessment. An *impact table* (such as developed in the L3Pilot project (Bjorvatn et al., 2021)) is extremely useful at the start to clarify which impacts have been considered and for which numerical estimates of the impact could be assessed and for which only indication of the magnitude could be given. An example from L3Pilot is given in Table 10.1. In addition, the reporting should cover impacts – if any – have been ruled out as negligible or impossible to assess.

Table 10.1 An example of summary table of the annual social value of impacts of Motorway Automated Driving Functions (ADF) for different penetration rates, and the annual social costs of implementing the system. Source: Bjorvatn et al., 2021.

	Penetration rate #1	Penetration rate #2	Penetration rate #3
QUANTIFIED IMPACTS in Million € ₂₀₂₀			
<i>Accidents</i>			
<i>Travel time</i>			
<i>Accident-induced congestion</i>			
<i>CO₂ emissions</i>			
<i>Fuel costs</i>			
∑ NET QUANTIFIED BENEFITS			
COST OF MOTORWAY ADF			
NET BENEFIT/COST-RATIO			
NON-QUANTIFIED IMPACTS: <i>Safety impacts outside ODD</i>			
<i>Perceived risk of accidents</i>			
<i>Travel time reliability</i>			
<i>Cost of travel time</i>			
<i>Obedience to speed limits</i>			

10.2.5 Geographical scope of assessment

The issue related to geographical scope is the ability to translate the findings of the FOT to a larger geographical area. The FOT is usually carried out at one or more locations, on a regional or national scale. However, the number of equipped vehicles and, if relevant, equipped roads, as well as the number of “equipped” kilometres driven, is usually a small percentage of the total vehicle fleet and the kilometres of roads. Therefore, in order to draw conclusions about the impacts and effectiveness of the system tested, a “scaling up” of the results is needed in order to draw conclusions and in order to ensure transferability of the results. [Section 9.7](#) addresses the scaling-up issues, which is to the national or European level. The availability of data plays a role in the decision to what level to scale up the results. [Section 10.4](#) goes into more detail to explain how to deal with this issue.

10.3 Analysis of impacts

The impact assessment framework involves the distinction between direct and indirect effects (in safety but also with respect to other effects, see below). It also implies the distinction between effects on internal and external costs. Positive efficiency effects typically lead to lower internal costs of transport (i.e. time, fuel consumption) but also

external costs (e.g. pollution, CO₂). The reduction of external costs is flagged out separately under environmental benefits because of its importance on the political agenda. The assessment can of course also consider wider economic effects (e.g. growth and employment effects of new technologies). However, given limited time and budget, it is useful to concentrate on the main impacts. In Figure 10.1 the scope of the impacts within socio-economic impact assessment are shown.

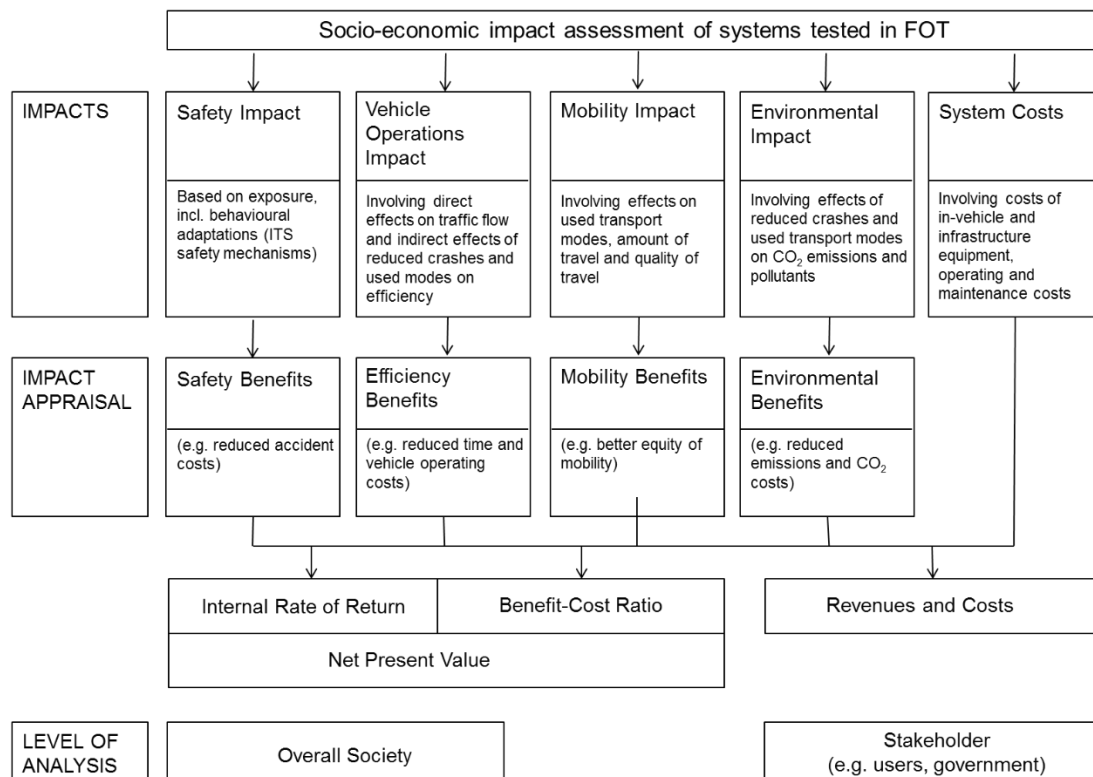


Figure 10.1 Scope of the impacts within socio-economic impact assessment. Source: Bjorvatn et al., 2021.

10.3.1 Safety benefits

The most direct and easiest way to calculate safety benefits would be to compare the number of accidents (and their consequences) happening during the baseline and treatment phase in an FOT. However, usually – luckily – not enough accidents happen in an FOT to make this approach feasible. Therefore, other methods have to be used. Traffic safety is regarded as a multiplication of three factors, namely exposure, accident risk and injury risk (Nilsson, 2004). The assessment of safety impacts has to consider these three effects which can be combined to predict the overall safety benefit, while taking driving conditions into consideration as well. Strategic decisions are highly relevant for exposure, and driving behaviour (on tactical and operational level) is relevant for accident and injury risk. A change in exposure may be possible to measure directly in the FOT: do people drive more or less with the system, do they drive on other road types, do they choose other routes, do they travel more by car? A change in mileage has a direct effect on exposure so on the number of accidents. Translating a change in driving behaviour into accident and injury risk is less straightforward. There are gaps in knowledge: the relation between changes in driving behaviour and (number of) accidents is often not known.

Therefore, there is not one method that is recommended to use. In this section a number of approaches to calculate the safety benefits of ITS applications are mentioned, with a (brief) explanation, references and information about in what situations they can be used.

Speed (variance) – accident relationships

Speed has a close relation to safety. The speed of a vehicle will influence not only the likelihood of a crash occurring, but will also be a critical factor in determining the severity of a crash outcome. This double risk factor is unique for speed. The relationship between speed and safety can be estimated by various models such as the Power Model (Nilsson, 2004; Elvik et al, 2004), that estimates the effects of changes in mean speed on traffic crashes and the severity of those crashes. The Power Model suggests that a 5% increase in mean speed leads to approximately a 10% increase in crashes involving injuries and a 20% increase in those involving fatalities. More examples of models for speed-safety relationships are reviewed in Aarts and van Schagen (2006). In the ISA UK project this is elaborated, different relations (models) for different road types can for example be found in Table 15 in Carsten et al. (2008).

The Power Model is valid under the assumption that mean speed is the only factor that has changed. Therefore, these models are more suitable for FOTs with systems mainly dealing with speed, and even then they fail to consider changes in the distribution of speed (shape of the speed distribution and changes in speed variance). The model is not suitable for systems that for example influence lateral behaviour.

Event-based analysis

Crashes are very rare events, thus there is a strong interest and need for the use of crash surrogates or “crash-substitute” events. The basic idea is that less severe events can be used instead of crashes to estimate safety benefits, because there is a systematic and well-understood relationship with crashes. Event-based analysis (EBA) uses events to estimate safety benefits. As reported in [Section 5.4.1](#), the basic principle of EBA is to identify short driving segments (typically in the order of 5–10 seconds), during which the risk of crashing is judged to be higher compared to other driving in the dataset, and then to analyse these events further. These events are often referred to as Crash Relevant Events (CRE), since their occurrence is thought to be indicative of actual crash risk in one way or another. EBA can be used for functions that warn of a certain event (e.g. FCW), not for functions that work continuously.

One way to gain insight in changes in safety include e.g. looking at the frequency of certain events (e.g. hard braking) and speed violations. More on the definition of events can be found in [Section 5.4](#). From a policy point of view, events are better usable than TTC, increase of mean speed, etc.

eIMPACT method

The eIMPACT method was used in the eIMPACT project for the safety assessment of in-vehicle safety systems (IVSS). The complete methodology can be found in (Wilmink et al., 2008). In short it works as follows: Effects of IVSS on traffic safety may appear in many, both intended and unintended ways. It is not possible to define in advance the group of accidents affected by the system, although system developers typically have as a starting point a target group of accidents for a system. Therefore, it is highly important that the analysis of IVSS covers all possible effects in a systematic manner. The approach was based on the system nature of transport. When one element of the system is affected, the consequences may appear in several elements and levels of the system, both immediately and in the long term, due to behavioural modification. Road safety is regarded as a multiplication of three orthogonal factors: (1) exposure, (2) risk of a collision taking place during a trip, and (3) risk of a collision resulting in injuries or death. In the analyses, the three main factors of traffic safety were covered by nine behavioural mechanisms as first described in (Draskóczy et al., 1998). Five mechanisms are mainly connected to the accident risk, three mechanisms deal with exposure, and there is one mechanism that deals with changes in accident consequences.

Every mechanism may result in either positive or negative impacts on road safety. In summary, the analysis aims to cover not only the direct intended effects of systems but also the indirect and unintended effects, including behavioural adaptation in long-term use. In addition, it was taken into consideration that the effects will vary according to road conditions and circumstances. This should ensure that all effects on safety are covered by the analyses.

The starting point for the safety impact assessment were the system specifications, including detailed safety function definitions. Figure 10.2 presents an overview of the phases in the analysis. An important part of the analysis is the use of accident data. For further details the reader is referred to (Wilmink et al, 2008).

Overview of safety impact assessments

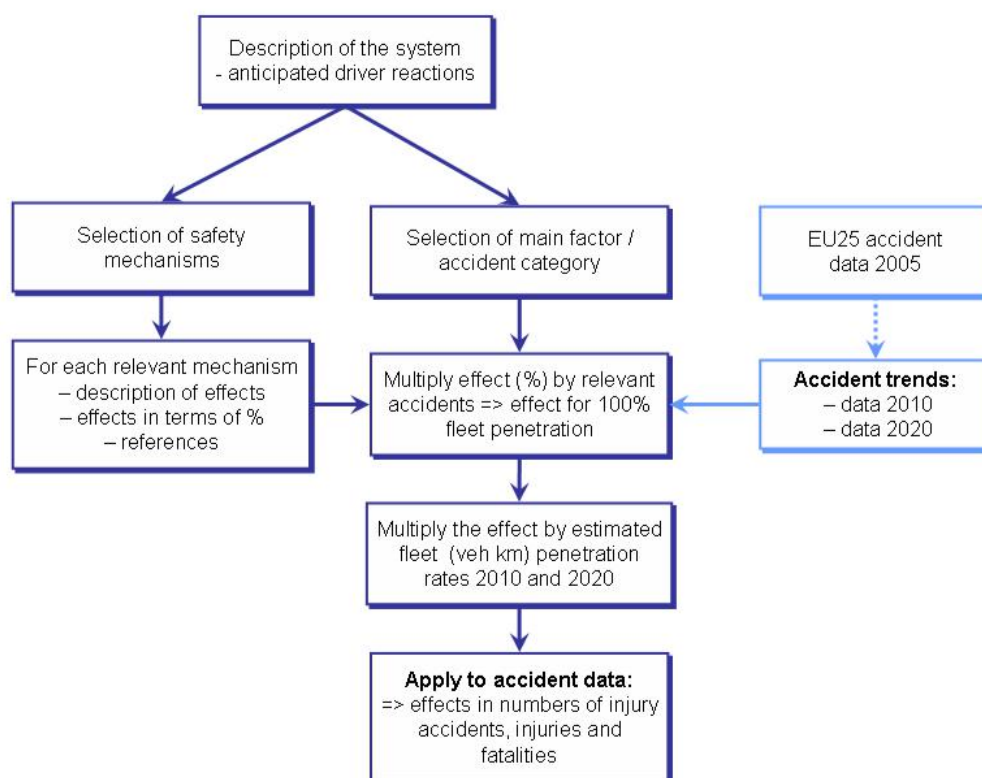


Figure 10.2 An overall schematic picture of the safety impact assessment in eIMPACT

Risk matrix approach

The risk matrix approach (RMA) was developed in the EuroFOT project (Malta et al., 2012). The RMA is developed for systems that function continuously (e.g. ACC) and that address an isolated accident type (e.g. rear-end collisions). The RMA associates a risk to each data point, by assuming a hypothetical accident scenario, developing from this data point. Separately there is a risk calculation from FOT data. Risks are pre-calculated once, and the application of FOT data is quite simple. This method is a variant of a method developed by NHTSA (Najm et al., 2006, NHTSA, 1996). The RMA does not rely on video data and is usable without in-depth accident statistics.

Expert judgment

This 'method' can be used when quantitative methods (as described above) do not work for some reason, or in addition to or to supplement them. Expert judgment usually produces qualitative results. Expert judgment can be done in different ways, for example by organising a workshop, or by having experts fill in a questionnaire. Expert judgment should be based on the data that are available from the FOT (surrogate safety measures, e.g. speed, speed variance, headways) but also from previous studies. Cutting the assessment into small, well-defined pieces that contribute to the overall estimate is recommended to allow the experts to make as reliable estimates as possible with this approach.

Choosing methods

The methods mentioned above are not perfect or applicable in all situations (for all functions, all types of FOTs, all impact areas, all regions, all demographic groups, all time scales). In the end, independent of the method one chooses, it is important to clearly write down the assumptions that were made and the consequences of these assumptions. When there is a lot of uncertainty in the safety impact assessment, an option is to work with a bandwidth and not deliver one fixed number as a result, but a range. A sensitivity analysis can also help with this.

As an example and representing best practice, the Mack FOT stated the goals of the safety analysis as follows (Orban et al., 2006):

1. Determine if driving conflict and crash probabilities will be reduced for drivers using the system
2. Determine if drivers drive more safely using the system
3. Determine reduction in crashes, injuries, fatalities if all fleets operating in the observed area were equipped with the system
4. Determine if drivers using the system have less severe crashes than drivers without the system.

The first step collects sensor data from each vehicle within the FOT (e.g. brake force, steering angle). Based on earlier definitions the number of driving conflicts can be determined. Thus, two numbers for the driving conflicts – reflecting the with-case and the without-case – are available to calculate the exposure ratio. This ratio reflects the number of driving conflicts in the with-case compared to the without-case. To provide an example: given a system which maintains the safe distance to a predecessor vehicle, the number of driving conflicts due to close following will be reduced from 10 conflicts per 1000 km to five conflicts per 1000 km. Thus, the exposure ratio equals 0.5, which indicates that driving with the system is safer than without it. In general, an exposure ratio below 1 indicates a safety benefit.

The benefit of lower exposure to accident risk will likely be modified based on adaptations of individual behaviour due to psychological reasons (second step). Behavioural adaptations can comprise e.g. adapting the following distance, adapting the speed variance, adapting the lane change behaviour (risky cut-ins or changing the lane without signalling it in advance). Examples of such behavioural changes can be found in the ITS safety mechanisms from the eIMPACT project (Wilmink et al., 2008). In this project, nine mechanisms have been introduced which lead to positive or negative safety effects. In most cases, the motivation for behavioural adaptation is that the driver wants to avoid “public” warnings (noticeable to all passengers) and “education” by the system.

The third step deals with scaling up from the FOT to a wider area (EU, country, region). This process is subject to the procedure proposed in scaling up.

The last step leads to the prevention ratio. In-depth information on accidents is used to calculate the mitigation effects of using the system. Maybe the system cannot avoid the accident, but it can mitigate the accident consequences. This issue has to be considered

in determining the effects for casualties. For systems affecting speed, the Power Model can be applied to calculate changes in severity.

Combining steps 2 to 4, it is possible to calculate the prevention ratio. For this ratio the probability of having a crash (casualty) when having a driving conflict in the with-case is compared to the same probability in the without-case. In the above example, the number of driving conflicts in the with-case was five and in the without-case 10. Let us assume that out of the five driving conflicts one accident occurs, and out of the 10 driving conflicts three accidents occur. Thus, the probability of having an accident due to a driving conflict is 0.2 in the with-case and 0.3 in the without-case. These values reflect the prevention ratios.

10.3.2 Mobility benefits

Many in-vehicle systems or systems related to driving impact not only the efficiency of the transport network or system, but also mobility on a personal level. Mobility is the potential for movement. It consists of means of travel and networks one has access to, knows about and is willing to use (Kulmala and Rämä, 2010; Spinney et al., 2009). Mobility is concluded to be the willingness to move along with potential and realised movement rather than just physical movement of vehicles, people and goods (Innamaa et al., 2011). Along with transport and infrastructure, it encompasses people's and road users' intentions, opinions and choices in their daily travel and movement. The concept of mobility is versatile. However, it is often reduced to transport or confused with accessibility or efficiency (Innamaa et al., 2013).

When the basic transportation infrastructure and services are functioning well and people have a choice in their means of travel, the quality of travel often becomes more important than the simple ability to get somewhere. Safety, comfort, reliability, privacy, continuity, and even "greenness" may be important choice parameters in everyday mobility (Nilsson and Küller, 2000). Mobility in itself also includes people's preferences of travel and choices of time, mode and route, their feelings, and also entails the ease of travel itself (Button et al., 2006; Gudmundsson, 2005).

People are reluctant to consider new modes of travel and new routes, because they have already gathered lots of information about their normal route with the vehicle most used, and travelling thus seems "easy". The relationship between how strongly past behaviour or habit and intention determine behaviour is assumed to be reciprocal. The more frequently a choice is made, the more habitual or script-based it becomes (Gärling and Axhausen, 2003). Thus, if the route or vehicle were to be changed, the person would need to seek information and construct new routes and evaluate alternatives, bringing psychological stress. This makes mobility impact assessment interesting, i.e. whether a new ITS provides the opportunity to experience new alternatives and thereby changes behaviour or not (Innamaa et al., 2013).

Gärling and Axhausen (2003) have addressed the problem of why private car use cannot be easily suppressed. The car is an attractive alternative to many, and there are often obstacles that prevent switching to other modes. Thus, drivers may be unable to switch even though they are motivated to do so. Unavailability of alternatives is of course a main obstacle in many cases, as is having a mobility impairment that prevents switching car trips to traditional bus services, cycling or walking. Yet, inertia or habit may also play an

important role. They increase transaction costs, since switching to another mode makes it necessary to learn new routines. Furthermore, searching and processing information about alternatives are reduced. Hence important changes may go unnoticed, such as for instance attractive alternatives becoming available (Innamaa et al., 2013).

Impacts on personal mobility have been assessed in FOTs from three points of view:

- Amount of travel: number of journeys, their length and duration
- Travel patterns: Timing of journeys, used modes and routes
- Quality of travel: feeling of safety and comfort, user stress and uncertainty.

As an example of how the mobility impact assessment was structured, see Figure 10.3.

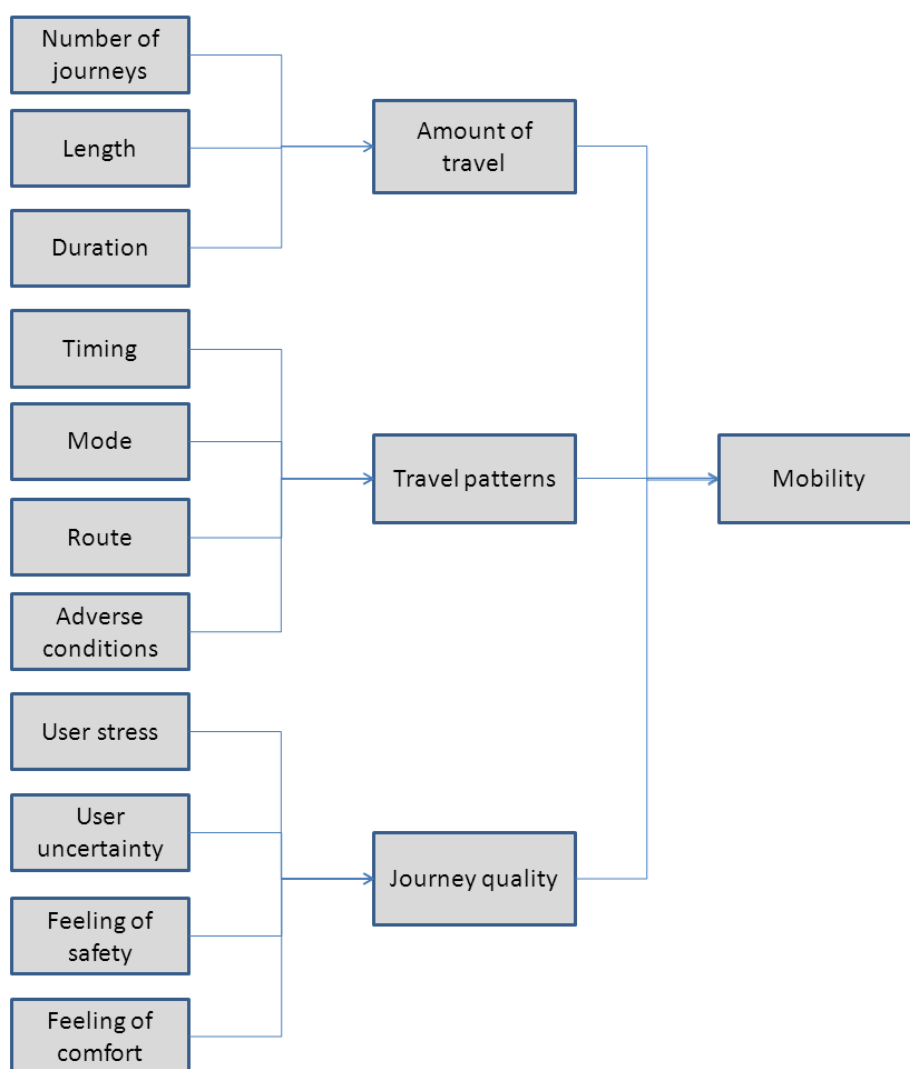


Figure 10.3 TeleFOT mobility model (Innamaa et al., 2013)

Mobility impact assessment is closely related to user assessment. It can be made as a qualitative assessment to indicate in which areas of mobility there are impacts and whether the impacts are beneficial or not. The point of view may be personal or of society, preferably both. The benefits in personal mobility can be seen as societal benefits on improvement in e.g. the equity of mobility. The choices made on timing and modal use

affect transport network efficiency, and the impact on overall mileage affects other impact areas.

Mobility impacts are tightly linked with the indirect impacts of changes on safety and other areas as they lead to changes in exposure (of different transport modes, road types, mileage). Thus, this is an impact area that should not be overlooked. However, it must be noted that the changes in mobility patterns often take time to change; thus, they may not be fully visible in the lifetime of an FOT. In addition, if the FOT consists of a set of controlled tests, impacts on mobility can be measured only in a limited way.

10.3.3 Efficiency benefits

Efficiency benefits are typically composed of two effects. They involve:

- Direct efficiency effects resulting from impact on vehicle operations (car-following, speed selection, etc.) and smoother traffic flow, e.g. where the system allows traffic to re-route to avoid current congestion, or improves mean speeds by encouraging safe car-following behaviour
- Indirect efficiency effects resulting from reduced crashes, e.g. reduced delays at incidents and accidents, or changes in exposure.

Direct efficiency effects can play an important role in the socio-economic impact assessment. On the appraisal level, direct efficiency effects are reflected in changes of time costs, fuel consumption costs and reliability changes. Because socio-economic impact assessment identifies quite commonly reductions of time costs as a major driver of the results, direct efficiency effects are generally worthwhile to explore.

The investigation of direct efficiency effects typically involves microscopic traffic flow simulation. A number of models (e.g. ITS Modeller, VISSIM, Paramics, DRACULA) have been applied to assess these impacts. Best practices, including on cross-validation of models, can be found in e.g. eIMPACT Deliverable 4 (Wilmink et al., 2008). Typically, when traffic flow becomes more homogeneous, the standard deviation of the vehicle speed becomes lower. As a result, the average vehicle speed may increase or the infrastructure capacity improves. As a consequence, time costs and vehicle operating costs will decrease.

However, the realisation of those benefits is closely related to the likely market penetration. Mature ICT systems typically can produce such effects, ICT systems in the phase of market introduction typically cannot. For internal efficiency, it is therefore important to figure out at the beginning of the FOT assessment (when the scope is defined) whether direct mobility effects will be likely to appear or not.

Compared to the direct efficiency effects, experience suggests that indirect efficiency effects are not restricted by conditions of market penetration. They can be realised in any case, as an add-on to the safety and mobility impacts. Indirect efficiency effects occur when the number – as well as the severity – of crashes is reduced and transport network-wise more efficient transport modes are used. The benefits result from less crash-caused congestion, therefore reducing journey times and fuel consumption. Typically, indirect traffic effects add up to about 10% of the safety benefits.

Given the state of the art in traffic modelling, indirect efficiency effects are assessed more frequently than direct efficiency effects. Good practice on the appraisal of indirect efficiency effects can be found, however, in European-scale assessment studies (eIMPACT; COWI, 2006) and US American FOT assessments (Batelle Memorial Institute, 2003; Volvo Trucks North America Inc., 2005). Some countries have methods specifically to address these effects (e.g. INCA in the UK).

10.3.4 Environmental benefits

Environmental benefits comprise lower CO₂ and air pollutants emissions. Noise also fits into this category, but we would caution that noise should only be analysed where ICT systems are expected to make a significant difference between the two scenarios (with/without the system). CO₂ and pollutants emissions are both speed dependent, with CO₂ emissions directly linked to fuel consumption.

Exhaust emission from road traffic is a complex process to describe. In [Section 5.6.3](#) a simplified formula is reported for calculation:

$$\Sigma(\text{Traffic activity}) \times (\text{Emission factor}) = \text{Total emissions}$$

where Traffic activity data includes: mileage, engine starts and parking. In addition to Traffic activity data, one needs data for: the vehicle fleet; road network; meteorological conditions; fuel quality, etc. If the driving pattern is influenced by the traffic situation, such data for the FOT vehicles are directly available. In order to estimate driving pattern changes for all vehicles by traffic situation, microsimulation models could be used. In order to estimate emission factors for these alternative driving patterns, there is need for exhaust emission measurements or exhaust emission models on an individual level. The recorded speed traces from the FOT vehicles can also be post-processed through a fuel consumption and emissions model to produce data on environmental effects.

The impact of CO₂ emissions is on a global scale, and is not linked to the particular country or area type where the CO₂ is emitted. The impact does, however, vary according to the year in which the reduction (or increase) in emissions takes place – the impact becoming greater further into the future. Actually, mobility effects have impacts on both efficiency and environmental benefits. However, because they are transmitted through the environment, and because they are largely externalities (i.e. their incidence is mostly on individuals other than the emitter), environmental benefits fall into a special category.

10.3.5 Integration of results

FOTs should make sure that there is enough time for the integration of results. At the end of the analysis phase, there are results for traffic efficiency, behaviour, safety, environment, acceptance, etc. This has to result in overall conclusions about a system, where the different impact areas are interwoven. However, this requires sitting together with the experts, and having time to let the results 'sink in'. Often this time is not available. Often iterations are needed, as impacts on e.g. safety lead to impacts on efficiency and environment in addition to the direct impact of the effects on driving style.

It is recommended to make a template for the reporting of results early on (for both external and internal reports), so researchers know what is expected of them. Reporting in a clear and systematic way helps with the integration of results and the overall view on the effects of the system.

It is very useful to make room for analysts to add insights discovered during the analysis phase (internal reports). It is a pity if these lessons were to be lost, as they could prove highly valuable to future FOTs.

10.4 Socio-economic Cost–Benefit Analysis

The analyst faces choices in setting up and carrying out the analyses. The choices will be influenced by the priorities identified by those setting up the FOT as well as budget and time constraints. The list below summarises these choices.

Methods:

- The basic choice is CBA, which summarises benefits and costs at a societal level
- Stakeholder perspectives: Makes use of the same input data as the CBA, but considers stakeholder-specific benefits, costs and financial analyses.

Identification of impacts:

- The basic choice is to use the costs incurred and the main expected benefit(s), as identified by use of the impact table
- Other impacts – both direct and indirect – can also be included, depending on the stakeholder perspective as well as the choices made elsewhere in the project, e.g. in hypothesis formulation, measurement methods and equipment and modelling capability
- Willingness-to-pay and use evidence, if also collected during the FOT, can be used to supplement the analysis methods above.

Scope of geographical assessment:

- The basic choice is the country level. In this case, the generic data needs (see [Section 10.4.4](#)) are limited to the country in question.
- EU-level analyses are preferred. These require substantially more general data from individual countries. Extra challenges in execution can be encountered due to differences in definitions or classifications.

10.4.1 System costs

System cost estimation is an element within FOTs which is quite often neglected. System promoters may not see costs as an impact. However, from a socio-economic point of view, they are a (negative) part of the impact of systems. Cost estimation should take care of the following aspects:

Cost elements to include: The system costs comprise the costs of in-vehicle, (roadside) infrastructure equipment and nomadic devices. Besides that, operating and maintenance costs have also to be considered.

Relevant size of costs: CBA applies a resource-based view. This means looking at potential savings of productive resources and, on the other hand, at the resources necessary to achieve this effect. The implication for cost estimation is that only the input of productive resources is relevant and not potential market prices. The convention proposed e.g. by eIMPACT is to use the cost price (the price of the ICT system paid by the manufacturer to its supplier) plus a mark-up which is allowed for in-vehicle implementation. However, market prices are relevant for user-centred analyses. Generally, in the face of limited evidence it is useful to apply the “Factor 3” rule of thumb, which means that in the automotive industry market prices for ICT systems differ from the cost prices by a factor of 3.

Process of cost estimation: Typically, cost estimation will be carried out by an expert group comprising of FOT internal staff and external industry experts. To avoid conflicts with confidentiality and the like, it appears sometimes helpful to introduce rough estimations to the group instead of working from blank sheets.

10.4.2 Classification of assessment methods

Figure 10.4 gives a classification of socio-economic assessment methods, based on which of the elements are included, and in particular:

- Whether a full set of impacts is addressed; e.g. if a significant CO₂ reduction can be anticipated, has it been included?
- Whether the assessment is from the social perspective only, or whether financial and stakeholder analyses are also provided.

The recommendation is that the FOTs should be designed to be as complete as possible, both in terms of impacts and stakeholder views. The assessments in the FOTs reviewed are examples of good practice. However, they differ in the types of analyses carried out, as well as in the scope of the effects examined, with the exception of safety impacts.

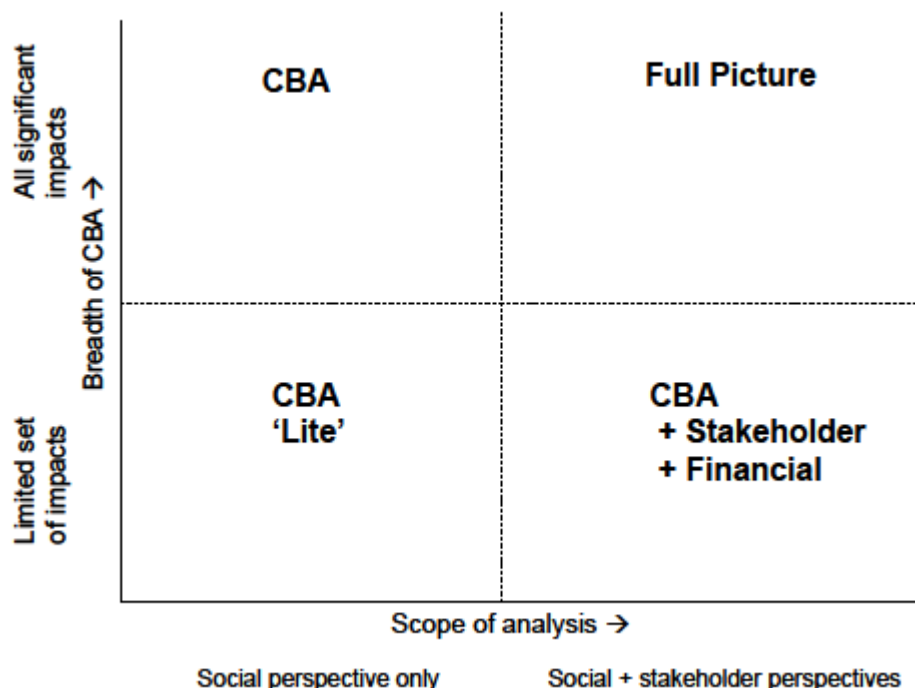


Figure 10.4 Classification of socio-economic assessments

Another dimension in which assessment methods can be classified is whether or not they make use of case-specific Willingness-to-Pay (WTP) evidence. In the design of future FOTs, we recommend that clients and analysts consider WTP studies as a way of getting better evidence on the users' likely demand for the products. WTP can provide uniquely useful evidence on the value of the ICT system to consumers and producers. In absence of this, FOTs can refer to evidence in the literature (market-based). WTP studies will, however, add to the cost and skill set required for FOTs, so the advantages and disadvantages will need to be weighed in each case.

We note that past FOTs have generally relied on market-based values (e.g. the US ACAS and Mack FOT), although the US ICC FOT did make use of specific WTP evidence, and as such is a useful reference. Also, we note that most previous assessment guidelines, including eIMPACT, assume that literature-based values will be used. Here, we leave the option open and recommend that clients and analysts decide at the inception phase of the FOT whether or not to go down the WTP route.

10.5 How to carry out the CBA assessment

The socio-economic impact assessment of a system within an FOT should be based on a CBA, since it is the most widespread, commonly accepted and practised method for analysing socio-economic impacts. It is clear that CBA accounts for all benefits and all costs on a society level, including benefits and costs to all groups. CBA follows a four-step-process involving framework and preparatory work, measuring impacts, appraising impacts in a common monetary value, and confronting the discounted society benefits with the costs of the policy measure. However, this process leaves also some room for shaping the individual steps of the process. We recommend considering the following issues:

10.5.1 CBA framework

Definition of the cases to be compared: Looked at is the with-case (ICT-system equipped) against the without-case (without the system).

Base year and time horizon of the assessment: CBA can be performed for the whole life cycle of the considered system or only for selected target years. This decision depends on information needs.

Geographical scope: Because of data availability the geographical scope should be congruent to existing statistical reporting ICT systems. Reference only to the local area where the FOT takes place is insufficient for this reason, and because the results of different FOTs need to be compared. This implies, however, that the socio-economic impact assessment has to undergo a scaling up procedure before the CBA in order to project the impacts from the FOT itself onto a larger area. The most practical appears to be assessment at the national level (assuming “nationwide deployment”). However, it is even more useful to provide results on a European level. The European perspective is important when the effects of FOTs in different Member States should be compared or when policy measures are planned or considered to ensure a European-scale deployment (e.g. eCall).

Discount rate: The discount rate ensures that benefits and costs are expressed for a common base year. A discount rate of 3% (real) is recommended as a default (see 'Other economic parameters').

Deployment scenario: It has to be estimated which share of new vehicles or which share of the total vehicle fleet will be equipped with the system in the target years and over the assessment period as a whole (depends on answer to 'Base year and time horizon' issue above). For lifecycle assessment it is also necessary to estimate the development of the equipment (technical capabilities, costs).

Impact table: The impact table serves as an instrument to expedite identification of impacts. It is aimed to ensure that the FOT team and the group responsible for the socio-economic impact assessment are fully aware of the complete impacts of the system. For efficiency reasons and likely budget constraints (competing FOTs and competing assessment issues within an FOT), it is necessary to concentrate the analysis on the significant impacts – those expected to be negligible, or impossible to analyse within the resources available, should be flagged as such in the impact table. Concerning the system, safety is often a relevant impact. Direct and indirect mobility impacts and environmental impacts are typically also addressed. System costs will always be relevant.

10.5.2 Inputs for impact assessment (including cost estimation)

Impact measurements represent an essential input to the CBA. We would normally expect most of these to feed through from the FOT experiment to the scaling-up procedure ([Section 9.7](#)) and to the CBA inputs. In particular, accident prevention and system costs at the national/EU level should be delivered this way. Impacts on efficiency and environment may require additional analysis before or at the CBA stage. Although in a well-designed FOT experiment, it may be possible to gather data specifically on any expected sources

of benefit, e.g. reduced variability of traffic speeds or reduced fuel consumption. However, e.g. simulation may still be required for traffic flow level impacts in different scenarios. The analysis of different FOT assessments has revealed some evidence on best practice for impact measurement. The requirements for CBA can be provided as a sort of output specification. This makes sure that the socio-economic impact assessment will be provided with the appropriate input data for carrying out the assessment. In terms of an output specification the following elements have to be put in place:

- *Accident and traffic performance database*: See [Section 10.5.3](#).
- *Effectiveness of the system*: These values represent a key output of the FOT which has to be provided to the socio-economic impact assessment
- Procedure for *scaling up the effects* to nationwide/European level
- *Cost estimations*: See [Section 10.4.1](#) on system costs.

10.5.3 Data needs

The data needed to carry out a socio-economic assessment for an FOT are extensive, and fall into two broad categories:

- FOT-specific data which will be gathered during the FOT itself
- Generic data, which plays a role in:
 - Scaling up the results from the experimental situation of the FOT to the national or EU level
 - Reaching a socio-economic assessment, based on the FOT data scaled up to national or EU level.

The following section outlines the FOT-specific and generic data likely to be needed. Thereafter recommendations on ensuring data quality and validity are given. Management of the data for socio-economic assessment follows next.

FOT-Specific Data

The key items of FOT-specific data likely to be needed are:

Accident rates (or risks) with and without the ICT system in place for the FOT sample. These will need to be differentiated by all the key drivers of accident rates (risks) in the FOT sample (e.g. road type; driver type; traffic conditions) so that accurate extrapolations can be made to the whole network. Accident rates (or risks) will be needed with and without the ICT system in place for the FOT sample. These may need to be derived from data on unsafe behaviours if the sample is too small to contain a significant number of actual accidents, although this is likely to be done as part of the PIs in any case. See [Section 10.3.1](#) on Safety Benefits.

One approach to estimating the impact on accident rates uses the effectiveness rate (% of relevant crash type avoided) as in the Collision Avoidance Systems (CAS) Benefits Study (NHTSA Benefits Working Group, 1996).

A more sophisticated approach can produce data on accident severity in addition to accident rates. Since accident severity is determined by the severity of the most serious casualty only, a complementary item of data would be any expected change in the number

of casualties per accident. Regan et al. (2006) measured time spent buckled up and time before buckling up to produce injury severity estimates.

Examples of how data is produced for accident severity and accident rates can be found in Regan et al. (2006), LeBlanc et al. (2006), and Volvo Trucks North America (2005).

Whichever approach is used to estimate accident rates and severities, the analysis will need to take account of any options in the implementation path. For example, in the Freightliner FOT study (Batelle Memorial Institute, 2003), there were four possible deployment groups (hazardous materials tankers; all tankers; tractor trailers; all large trucks); input data will be required for each of these options.

Multiple scenarios may also be needed to enable sensitivity testing. That is, where there is uncertainty over accident rates/severity or other key variables, this can be handled through 'what if' scenarios based on combinations of the possible outcomes (Batelle Memorial Institute, 2003).

There may also be some value in having spatially differentiated data, and being able to link behaviour to traffic conditions, e.g. urban/non-urban and traffic congestions, Volvo Trucks North America, 2005).

Market penetration forecasts: In the literature, SEiSS (Abele et al., 2005 and Baum et al., 2006) gives particular attention to market penetration.

Usage, reliability and compliance: Although the CAS Benefits Study (NHTSA, 1996) made assumptions about usage, reliability and compliance rather than gathering data, it did draw attention to these important factors in the out-turn effectiveness of ICT systems. Usage refers to the percentage of drivers, driving time or kilometres driven for which ICT systems installed in the vehicle will be switched on and active. Reliability refers to the likelihood that ICT systems will operate without failure, technically. Compliance refers to the percentage of occasions on which the driver's behaviour complies with a warning or indication provided by the system.

Attitudinal and acceptance data: Many FOTs gathered attitudinal and acceptance data (Regan et al., 2006; LeBlanc et al., 2006; USDOT, 1999; Dingus et al., 2006; and Volvo Trucks North America, 2005).

Costs of the ICT systems: See [Section 10.4.1](#). In some FOTs, data has been gathered which inputs directly into the maintenance and operating cost calculations (Volvo Trucks North America, 2005). Where the assessment period is longer than the expected service life of the equipment, replacement costs should be included (e.g. in the Freightliner FOT, one round of replacement was included since the service life was 10 years and the assessment period 20 years; Battelle Memorial Institute, 2003).

Generic Data

The key items of generic data likely to be needed are:

National and EU-level network, fleet and traffic data, which are used in scaling up the findings from the FOT to the level of political interest: The International Road Traffic and

Accident Database, IRTAD (<http://www.itf-oecd.org/IRTAD>), contains traffic data for the EU27. This includes vehicle kilometres on the total road network, vehicle kilometres on motorways, and vehicle kilometres on urban roads. Vehicle kilometres on rural roads can be derived; some data are missing.

Accident data (accidents, fatalities, severe and slight injuries) for base scenario: National databases are available. At the EU level, the collection and compilation of accident data as a basis for the safety impact assessment is a challenge, especially when specific target accidents are going to be explored. Several EU projects are dedicated to harmonising accident databases (e.g. TRACE project or SafetyNet). Forecasts of road safety are needed. An example can be found in eIMPACT “Impact Assessment of Intelligent Vehicle Safety Systems”, in which road safety predictions (made in 2008) for 2010 and 2020 for the EU25 were presented.

More detailed network specifications (e.g. infrastructure equipment) may be required for some systems: the presence/absence of beacons, signalisation. Basic figures (e.g. share of Trans-European Road Network equipped with dynamic traffic management) are available from the eSafety Forum Implementation Roadmap Working Group (available on request from ERTICO).

Speed-flow relationships or network models, which allow journey times and costs to be derived from changes in flows: Although these are strictly much more than just ‘data’, it is worth highlighting the key role they play in socio-economic assessment of transport ICT systems. Many of the effects of new ICT systems will be mediated through changes in traffic flow on the network – for example, advanced warning ICT systems allow drivers to change route to avoid hazards, but the net effect on travel times and costs is dependent not only on the behaviour of the individual, but also on the behaviour of large numbers of individuals and interaction with the limited capacity of the network. Hence, at the very least, knowledge of speed-flow relationships is needed to understand the consequences of shifting traffic across the network.

HCM (2000) and FGSV (2001) are sources of speed flow relationships. Network models or strategic transport models incorporate this data and have much wider functionality. The fact that these models are very expensive to develop and maintain means that they tend not to be developed for one socio-economic assessment in isolation. Instead, part of the socio-economic assessment process is usually to identify models already existing which can provide the necessary functionality.

Evidence on accident costs, used to measure the benefits of accident reduction and changes in accident severity: The HEATCO project (Bickel et al., 2006) was designed specifically to provide harmonised cost estimates for socio-economic assessment in Europe. We recommend that the HEATCO accident cost values are used in the FOTs, and we provide one additional piece of evidence to fill a gap in HEATCO which is a generic dataset on the costs of ‘damage only’ accidents.

Two of the main issues in this field are:

- An apparent inconsistency between ‘willingness-to-pay’ (WTP) methods and ‘cost of damage’ or ‘human capital’ methods as a basis for values – empirically, WTP

methods can produce significantly higher values for fatalities in particular (see Assing et al., 2006: Table 16);

- Double counting of casualties' lost future consumption, which is included in both lost future output and WTP to reduce accident risk.

HEATCO addressed these issues by specifying a common framework in which the different elements of accident costs measured by each method can be reconciled. For example, 'human capital' methods do not capture people's full valuation of safety risk, whilst WTP-based values do not capture the external resource costs of accidents (e.g. healthcare costs borne by the state) but often do double-count lost future consumption, as already noted. The HEATCO framework includes:

- Property damage
- Medical costs
- Administration costs
- Lost output
- Welfare losses due to casualty reduction.

As a result, the HEATCO values for fatalities are neither as high as the US NHTSA's WTP values cited by Assing et al. (2006), nor are they as low as the NHTSA's cost-of-damage values. They are broadly in line with 'best practice' European values used in CBA, and the differences can generally be understood by examining the differences in the underlying measurement methods.

Other important functions of the HEATCO values are to provide:

- A common unit of account in the face of taxes and subsidies – HEATCO values are provided at the factor cost unit of account (Bickel et al., 2006, p. 52)
- A common price base year
- A common currency, €, for European-level assessments.

In HEATCO, accident values are listed in a table, expressed as values per casualty saved. These values do include the full set of accident costs, per casualty.

To apply these values, analysts will require further data:

1. Forecasts of accidents with and without the technology in place – based on the FOT findings and the results of the scaling-up process.
2. If these forecasts do not address unreported accidents, then factors for the number of unreported accidents given the number of reported accidents can be found in HEATCO (Bickel et al., 2006, Table 5.1).
3. Growth in the values over time – an elasticity of 1.0 with respect to GDP per capita, thus a 2.0% annual increase in GDP per capita would imply a 2.0% annual increase in the values of accident reduction.
4. Damage-only accident values. As Baum et al. (2007) shows, savings in damage-only accidents can make up a large proportion of the benefits from ICT safety systems. Damage-only accident costs may be approximated at 17% of the cost of a Slight Casualty (Nellthorp et al., 1998).

National level assessments may wish to take advantage of the most recent safety valuation evidence at the national level. For multinational assessments it will be important to ensure that any national evidence is checked for consistency across boundaries, and conversions made if necessary (e.g. in terms of base year, unit of account, cost elements included, measurement methodology, etc.).

For EU-level assessments, consistency across countries and comparability between assessments is important, which makes the use of a harmonised set of values (as above) more attractive. If the harmonised values are found not to provide the detail which the analyst needs – e.g. if differentiated accident costs by road type or user type are expected to be a key requirement for a particular assessment – then it may be appropriate to vary the values above, based on more detailed information (for example, the accident cost data included in national level assessment guidelines).

Evidence on values of time savings and vehicle operating cost savings, used to measure the benefits of changes in traffic flow: Values of travel time savings will be needed to assess the benefits of improved traffic flow due to the ICT systems. HEATCO Tables 4.6-4.8 provide suitable values for working and non-working passenger trips, and for freight transport (Bickel et al., 2006, pp. 73-75). These values increase with GDP per capita, at an inter-temporal elasticity of 0.7.

Sometimes the impact on reliability is addressed, not only expected (mean) travel times, and in these cases we recommend using the reliability ratios set out in HEATCO (Bickel et al., 2006, Table 4.3) to value changes in the standard deviation of journey time.

Vehicle operating cost savings are also likely to arise from changes in traffic flows. The traffic models used to predict traffic flow responses to ICT systems will typically be capable of predicting changes in vehicle operating costs, and the fine network detail in these models usually makes it more logical to calculate these cost savings within the model, rather than attempting to do so based on model outputs. As a result, standard values are not offered for these impacts by HEATCO (Bickel et al., 2006, pp.135-140).

Emissions factors and values for the damage caused by emissions of greenhouse gases, air pollutants and noise: HEATCO provides values for both sources of emissions (Bickel et al., 2006, Tables 6.2, 6.4). Values for particulate (smoke) emissions are differentiated between urban and non-urban locations, due to their much localised impact pathway. Other air pollutants are valued uniformly at country level. HEATCO provides a shadow price of CO₂ by year of emissions (Bickel et al., 2006, Table 6.12), which should be applied to all forecast changes. The impact of noise changes may be quantified using the HEATCO values for road, rail and aircraft noise in each Member State (Bickel et al., 2006, Table 6.9).

Other economic parameters such as the social discount rate: Discount rates are required for socio-economic assessment. In line with HEATCO, we recommend using a risk-free social time preference rate for the countries to which the assessment would apply. If a default discount rate at the EU level is required, we would recommend using 3% per annum (real). GDP growth data for the members of the EU27 (required for updating values of accidents, etc, over time) is available from [Eurostat](#).

Data quality and validity

The EC ROSEBUD project provided the following guidance as part of a “professional code for analysts” (BASt et al., 2005, p. 46):

“Data has to be attributed correctly to its sources, especially when different data sources like national or international accident databases or in-depth databases are used. Where and how estimations were made to fill data gaps needs to be documented. Regression models should be used to generate future time series; trend extrapolations can replace them where available data are insufficient for regressions”.

In addition, we would recommend that:

- The principles of statistics apply – statistical tests should be used wherever possible to determine if hypotheses about ICT system impacts are supported by the FOT evidence, and sample sizes should be chosen to obtain statistically significant results;
- When scaling up from the FOT to the national or EU level, a methodical approach based on the key drivers of safety/other significant outcomes identified in the FOT should be used (cross reference);
- Confidence intervals as well as mean data should be recorded for key variables – note that confidence intervals are given in HEATCO for the various economic parameters recommended;
- We have noted the need to recognise the uncertainty in the data using sensitivity analysis – if analysts wish to take a more advanced approach and use Monte Carlo simulation or related techniques (for example to derive a probability distribution on NPV (Net Present Value) or BCR (Benefit Cost Ratio)), that would be welcome, as it simplifies the outputs seen by the decision makers, although it does place an additional burden on the analysts;
- Known problems with the data should be acknowledged and acted upon, e.g. LeBlanc et al. (2006) excluded a proportion of drivers whose trials were invalidated (in that case nine out of 87 drivers), and some trips by the remaining drivers. Well-known problems with the omission of unreported accidents from data have prompted Bickel et al. (2006, Table 5.1) to provide adjustment factors for different accident severities and types.

Finally, the US NHTSA observes that “the validity of any experimental test results depends on the experimental condition effects that were placed on the drivers” (NHTSA, 1996: p. 36). Care is needed, therefore, when extrapolating data from short-term experiments to long-term term adjustments in behaviour and demand for ICT systems – e.g. the CAS Benefits Study (NHTSA, 1996) proposed that “a better estimation of the safety benefits...can be achieved as more relevant test data are gathered especially from long-term, large-fleet field operational tests” (p. C-8).

10.5.4 Impact valuation

Methodological base for impact valuation: The general objective of this step is to provide unit values for the physical impacts. Several methods compete in the field of impact appraisal. They can be subdivided into objective approaches (e.g. damage costs, avoidance costs) and subjective approaches (e.g. WTP). In EU Member States, different practices and preferences exist for impact appraisal. A lot of surveying and standardisation efforts have been made by projects like HEATCO (Bickel et al., 2006) to reach a common European base. As a general recommendation, it can be stated that unit values for CBA should be based on objective approaches. However, WTP information can largely contribute to a higher quality of the assessment when analyses for the users are carried out.

Good practice on unit values: See eIMPACT (Assing et al., 2006), HEATCO (Bickel et al., 2006) and the handbook on external costs of transport.

National or European unit values: This decision corresponds to the geographical scope. Assessment on national level will typically make use of national cost unit rates. For European-scale assessment we recommend using the harmonised values contained in HEATCO – note that these are still differentiated by country, but are on a harmonised theoretical basis.

10.5.5 Results

CBA can produce different summary measures of performance. It represents good practice to calculate the Net Present Value (NPV) by summing up all discounted values of benefits (plus sign) and costs (minus sign). Moreover, Benefit–Cost Ratios (BCR) are a very common expression of system profitability which can be calculated by dividing the total benefits by the total costs. It is also practical (see "Base year and time horizon") to calculate "snapshot" BCR for target years. In this case, the costs will be transformed to annual values (using the discount rate) and will be compared to the target year benefits. For FOTs, we recommend the calculation of both figures, NPV and BCR.

For the **social CBA**, we recommend reporting:

- Safety benefit (€M)
- Other benefits to road users (€M) – mainly time savings, operating cost savings and reliability gains
- Environmental benefits (€M) – including climate change, regional and local air quality effects; noise; and other impacts
- Revenue to operators (€M) – there may be multiple operators, including infrastructure and service operators; each will want to know the impact on themselves (financial), although for the social CBA these revenues may be aggregated
- Costs to operators (€M) – including capital, maintenance and operating costs
- Revenue to automotive OEMs (€M)
- Costs to automotive OEMs (€M)
- Revenue to government (€M) – including tax revenue changes

- Costs to government (€M) – including investments in R&D and in implementation of ICT systems.

An example of tabulation of the social CBA is shown in Table 10.2. All entries are at Present Values. A common base year (for prices and discounting) aids comparison across different technology options. RAILPAG (EIB & EC, 2005) has a more detailed breakdown by stakeholders (an 'SE Matrix'), which some analysts may find helpful in presenting the social CBA.

Table 10.2 Social CBA tabulation

Group	Impact	€M (2008 base)		
		2015	2025	Present Value (Total)
Consumers	Safety benefits	289	299	3715 a
	Other road user benefits	574	606	603 b
	Environmental benefits	63	66	58 c
	...			d
Producers	Revenue	723	780	8520 e
	Costs	248	233	-8311 f
Government	Revenue	3	4	34 g
	Costs	12	14	-379 h
Net Present Value (NPV) = $\sum a..h$				4240

Notes: Sign: all negative impacts on the Group affected are shown with a negative sign, thus costs appear with a negative sign; 2008 base: indicates appraisal at constant general prices using 2008 Consumer Price Index (CPI), and with 2008 as the base year for discounting in the Present Value column.

In cases where the public sector expects to contribute to the development or implementation of the system, we recommend also presenting a BCR with respect to public sector support, which HEATCO (Bickel et al., 2006, pp. 41–42) identifies in use by the EU and Switzerland:

$$BCR = \frac{NPV}{PV(\text{PublicSectorSupport})}$$

The calculation of the BCR is a delicate issue in CBA. On one hand, the BCR is a very powerful measure, because it applies to the common situation where investment budgets are limited and maximum value for money is required (making best use of a scarce resource). On the other hand, the definition of 'costs' (the denominator) can be problematic.

As a general rule, the BCR is useful when the denominator is defined in the same way for all options being compared – for example, NPV per unit of central government budget (which would be a BCR of interest to central government). Our recommendation of a BCR with respect to public sector support broadens this to the budget for public expenditure as

a whole. This avoids creating an incentive to manipulate the BCR by shifting costs to local and regional government.

In the example shown in Table 10.2, the BCR with respect to public sector support will be $4240/(379-34) = 10.3$, which indicates a high social return from each € of public funds contributed.

Carrying out a stakeholder analysis

In contrast to CBA, only particular benefits and costs are relevant for particular stakeholders. The reduction of exhaust and CO₂ emissions are not benefits to users, unless they are charged for it (through vehicle taxes or tolls). The costs of in-vehicle equipment do not represent costs to the government, unless the government agrees to pay for a share of this. The consequence is that ICT systems which are profitable at society level (NPV, BCR) will not be deployed when a relevant stakeholder group is economically impaired. Hence, it is necessary to include stakeholder perspectives in the FOT socio-economic assessment.

Practically, stakeholder analyses also make use of accounting costs and benefits, but on the level of the individual stakeholder group. This implies the following for *users*, but also in general:

- Costs and benefits must be investigated according to their stakeholder relevance. Safety benefits (reduced accident and casualty risks), for instance, are relevant to *users* (and to *insurance companies* as well).
- The appraisal of the impacts can be different. *Users* face market prices when considering the investment in a system (see factor 3 rule of thumb). For benefit evaluation the implication is to use market values if available (e.g. fuel consumption: station prices (incl. taxes) instead of net prices). Otherwise, WTP approaches have their justification here because they are better suited to reflect individual preferences.

Further adaptations to the CBA approach involve the use of a different discount rate (reflecting private sector interest rates) and the use of a different result measure (fair market price for a pre-defined annual vehicle mileage or the critical (break-even) mileage for a given market price).

The stakeholder analysis reporting will vary with the analytical methods used. For example, in the TAC Safe Car FOT Monash University used subjective questionnaire methods to investigate users' acceptance of several ICT systems including ISA (see Regan et al., 2006).

Another useful form of stakeholder analysis from the user perspective is WTP evidence, as shown in eIMPACT Deliverable 3 (Assing et al., 2006 p. 119).

For the *vehicle OEMs* and both *infrastructure and service operators*:

- Where they are commercial bodies, a *financial analysis* will provide the most important stakeholder information

- Where they are *public sector* agencies, a *financial analysis* may need to be combined with an assessment against their public service objectives – however, in some cases the overall social CBA will serve this purpose, depending on the approach taken by the agencies involved.

Carrying out a financial analysis

The internal rate of return (IRR) of a project is the interest rate that will generate an NPV of zero. In an equation, this is:

$$\sum_{t=0}^T \frac{B_t - C_t}{(1 + i_{IRR})^t} = 0$$

where IRR is internal rate of return.

The stakeholder for whom the IRR is calculated compares the IRR with a target rate. This target rate depends on each stakeholder. For public authorities as a stakeholder, the target rate will be less than for private investors.

In any case, a calculated BCR or IRR should be accompanied by an NPV. We recommend that financial IRRs are reported for all FOTs.

The IRR concept can be modified for comparison purposes. For this approach, the cash flow streams are subtracted. With the new cash flows the modified IRR is calculated. If the IRR is above the trigger rate, the project with the larger cash flow is the better project.

Of key interest will be the IRR from the point of view of *specific stakeholders* (or stakeholder types). The IRR for vehicle OEMs will influence their decision about investing in the technology. Similarly, the IRR for infrastructure operators and service operators will influence their decisions – particularly where these are commercial operations.

Hence the key information will be in the form:

$$IRR_{OEMs} = \dots \%$$

$$IRR_{RoadAuthorities} = \dots \%$$

Further IRRs should be reported where there are other stakeholders with a commercial interest, for whom significant impacts are expected. Tables such as those used by [WebTAG](#) also provide a useful series of snapshots of the financial impact. In this case, in order to be meaningful the tables should relate to specific stakeholders or stakeholder types, e.g. vehicle OEMs or road authorities.

The financial results can be taken a stage further by reporting the break-even point in terms of sales or market penetration, or the target price, down to which the system must be engineered in order to achieve financial viability. Graphical presentations may be useful in these cases.

10.6 Socio-economic impact assessment in CCAM FOTs

Due to regulation and safety reasons, most of the automated driving tests in H2020 programme were pilots where the vehicles are being tested with controlled tests or in specific areas, mostly having a safety driver on-board. Consequently, one cannot scale up the results of the field measurements directly to large penetrations. An additional step of scaling up and future scenario work is needed to go from prototype behaviour and limitations to market-ready CCAM behaviour. Naturally, an agreement of the likely features of this market-ready CCAM need to be reached and evaluation results presented in transparent manner. Similar considerations regarding market-ready behaviour have been used in past ADAS FOTs, but with less uncertainties.

For the scaling up from single vehicle in certain conditions to a European level all year round, the Operational Design Domain (ODD) of the tested systems is very important. We need to identify on which roads and when these systems could be used, and, in addition, to identify the trips and conditions when those people who have the system decide to use it. Statistics and databases that have been used previously for scaling up, may not all include relevant elements of ODD. Work is needed to link ODD with the traffic and driving conditions in the whole of Europe.

If the socio-economic evaluation is made to a certain future year, the impact of CCAM will be mixed with the impact of other trends affecting transport, like electrification, shared economy, urbanization, etc. This creates challenges for definition of baseline, future without automation, i.e. what are meaningful points of comparison.

11 Sharing knowledge, lessons learned and data

In order to learn from FOTs, it is crucial to collect lessons learned. Lessons learned may concern many aspects of FOTs, and each step in the FESTA methodology may lead to lessons learned. FESTA strongly recommends to document all steps of an FOT, as this is important for both interpreting the results and learning from the (good and bad) experiences.

Lessons learned may have four elements:

- What went well, and why?
- What went wrong, why, and what was the consequence?
- What could have been done better and how?
- What is the relevance of the lesson learned for other projects?

Lessons can be collected on all levels of detail, and all can be useful. Very detailed ones are useful as recipes of a kind (e.g. “software tool X is handy for the analysis”), or helping to avoid similar mistakes. If the level is too vague, the lesson may become an open door (e.g. “use sufficient resources”). However, even if a stated lesson looks rather generic, it may be useful as a pointer to a project document, where more details can be found. In addition to the tested best/good practices, also hints on potential improvements that could be tried out are useful.

We may distinguish several types of lessons learned:

1. Set-up and execution of field test
2. The systems and functions under evaluation, their description
3. Data gathered in the project and external data, data processing and storage
4. The evaluation process itself
5. The evaluation methodology applied
6. Evaluation tools and sensors
7. The experiences and acceptance of the users
8. The communication with stakeholders
9. Sharing of data with project and external partners.

An example template for gathering and documenting lessons learned is given in Table 11.1.

Table 11.1 Template for gathering and documenting lessons learned

Short name of the lesson learned	
Summary	
Category of lesson learned	
Project	
Related deliverable/document	
FESTA V step	
Optional: relevance for specific projects	
Keywords	

1. Set-up and execution of a field test

In setting up the test many choices and decisions have to be made, for example, selection of research questions and experimental environment. Practical issues have to be solved, such as the availability of vehicles and participants, use of resources, and permissions needed from authorities and ethical committees. Not all problems can be foreseen and new issues may arise during the execution of the field test. Learning from the experiences from other projects may be very valuable in avoiding and mitigating problems.

2. The systems and functions under evaluation, their description

A large variety of systems and functions may be evaluated, and a wealth of experience is available from projects, on how to deal with the detailed technical aspects. With fast-paced development, lessons learned may be too specific and become outdated soon. However, it may be very useful to search for experiences from other projects evaluating systems with similar functions.

3. Data gathered in the project and external data, data processing and storage

For both data gathered by the project and data from external sources there may be many issues e.g. with content and quality of the data. Some of them are not known before the analysis process. Increasing the common knowledge on data sources with sufficient quality or at least the weaknesses of different (especially public) data would save time and effort.

4. The evaluation process

Lessons learned concern how the evaluation was set up in an FOT, starting from setting goals to reaching the final conclusions. Examples are the way in which certain issues in the project were managed and structured, goals were set and realised, resources managed, and results were interpreted and justified.

5. The evaluation methodology applied

As the FESTA methodology is regularly updated and, in the future, a Common Evaluation Methodology for CCAM will be developed, lessons learned for the evaluation methodology are very valuable. Both for the purpose of improving the application of the methodology and for improving the methodology itself.

6. Evaluation tools and sensors

New tools and sensors are becoming rapidly available, so lessons about specific tools may become obsolete at some point. However, it may be useful to search for experiences from other projects before acquiring hard- and software tools and sensors or developing them by the project itself.

7. The experiences and acceptance of the users

Performing user tests with complex systems such as automated vehicles in a natural environment may be complex and challenging, and especially so, if the systems are not yet mature and are still under development. Users may even not be allowed to operate the systems themselves, or systems can only be tested in artificial circumstances. Sharing best practices on how to enhance the user experience may be very helpful.

8. Communication with stakeholders

The goal of an FOT is usually to produce results that are useful for stakeholders. Lessons learned about stakeholder analysis, managing the expectations of stakeholders, communicating with stakeholders and translating technical results into information that can be used by different groups of stakeholders, are very useful. These lessons are not always easily to share for confidentiality and commercial reasons.

9. Sharing data with project and external partners

When working with external partners, some data, results and methods can be shared, but not all. Even within a project, some data will remain confidential. Practical solutions found for data sharing within the project and outside the project are useful for others.

Finally, FESTA encourages FOTs to share their methods, tools, knowledge, and data. Re-use of existing methods and tools can make a big difference in the time that is needed to prepare an FOT. Sharing knowledge and data can offer interesting collaboration and dissemination options. Also, consulting evaluation experts can help to clarify available options regarding study design and how to make the best out of available resources. For the information-sharing, the European Knowledge Base for Connected and Automated Driving is a good starting point: <https://www.connectedautomateddriving.eu/>.

References

- Aarts, L. and van Schagen, I. (2006). Driving speed and the risk of road crashes: a review. *Accident Analysis and Prevention*, 38, pp. 215-224.
- Abele, J., Kerlen, S., Krueger, S., Baum, H., Geißler, T., Grawenhoff, S., Schneider, J. and Schulz, W. H. (2005). Exploratory Study on the potential socio-economic impact of the introduction of Intelligent Safety Systems in Road Vehicles (SEiSS). Final Report to the European Commission DG Information Society and Media, VDI/VDE Innovation + Technik GmbH and University of Cologne. https://trimis.ec.europa.eu/sites/default/files/project/documents/20101208_184626_84759_SEISS%20final_2005.pdf
- Ajzen, I. and Fishbein, M. (1980). *Understanding attitudes and predicting social behavior*. Englewood Cliffs, NJ: Prentice-Hall.
- Armstrong, M. (1999). *A handbook of Human Resource Management Practice*. London, Kogan Page.
- Antin, J., Lee, S., Hankey, J. and Dingus, T. (2011). Design of the In-Vehicle Driving Behavior and Crash Risk Study. In Support of the SHRP 2 Naturalistic Driving Study. SHRP2 Report S2-S05-RR-1. ISBN: 978-0-309-12895-7, National Academy of Sciences.
- Assing, K., Baum, H., Bühne, J.-A., Geißler, T., Grawenhoff, S., Peters, H., Schulz, W.H. and Westerkamp, U. (2006). Methodological framework and database for socio-economic evaluation of Intelligent Vehicle Safety Systems. Deliverable D3, eIMPACT: Socio-economic impact assessment of stand-alone and co-operative Intelligent Vehicle Safety Systems (IVSS) in Europe, EU 6th Framework Programme Contract no.: 027421. TNO, Delft.
- Barnard, Y., Koskinen, S., Wilmink, I., Sanz, L. and Mesones, A. (2021). Micro-FESTA for CCAM pilot projects. ARCADE document. <https://www.connectedautomateddriving.eu/wp-content/uploads/2021/08/Micro-FESTA-v2.pdf>
- BASt, DITS, KuSS, TRL, UOC and CDV (2005). Road safety and Environmental Benefit-Cost and Cost-Effectiveness Analysis for Use in Decision Making. Rosebud Deliverable D7.
- Battelle Memorial Institute (2003). Evaluation of the Freightliner Intelligent Vehicle Initiative Field Operational Test, Final Report FHWA-OP-03-179. Battelle, Columbus, Ohio. <https://rosap.ntl.bts.gov/view/dot/37217>
- Baum, H., Geißler, T., Grawenhoff, S. and Schulz, W.H. (2006). Cost-Benefit Analyses of Intelligent Vehicle Safety Systems – some empirical case studies. *Zeitschrift für Verkehrswissenschaft*. 77(3), pp. 226-254.

- Baum, H., Grawenhoff, S. and Geißler, T. (2007). Cost-Benefit-Analysis of the Electronic Stability Program (ESP). Summary Report. Institute for Transport Economics, University of Cologne.
- Bickel, P., Friedrich, R., Burgess, A., Fagiani, P., Hunt, A., De Jong, G., Laird, J., Lieb, C., Lindberg, G., Mackie, P., Navrud, S., Odgaard, T., Ricci, A., Shires, J. and Tavasszy, L. (2006). Developing Harmonised European Approaches for Transport Costing and Project Assessment, HEATCO Deliverable 5. Germany.
<https://www.scienceopen.com/document?vid=93e4f422-c66f-4098-845b-e2a5f3b0d09b>
- Bjorvatn, A., Page, Y., Fahrenkrog, F., Weber, H., Aittoniemi, E., Heum, P., Lehtonen, E., Silla, A., Bärgrman, J., Borrack, M., Innamaa, S., Itkonen, T., Malin, F., Pedersen, K., Schuldes, M., Sintonen, H., Streubel, T., Hagleitner, W., Hermitte, T., Hiller, J., & Torrao G. (2021). L3Pilot Deliverable D7.4 / Impact Evaluation Results.
<https://ec.europa.eu/research/participants/documents/downloadPublic?documentIds=080166e5e32107ea&appId=PPGMS>
- Boyle, L.N., Hallmark, S., Lee, J.D., McGehee, D.V., Neyens, D.M. and Ward, N.J. (2012). Integration of analysis methods and development of analysis plan. SHRP2 report S2-S02-RW-1. Transportation Research Board, Washington, D.C.
<https://www.nap.edu/read/22847>
- Button, K., Stough, R.R., Bragg, M. and Taylor, S. (2006). Telecommunications, Transportation and Location. Massachusetts: Edward Elgar Publishing, Inc.
- Carsten, O., Fowkes, M., Lai, F., Chorlton, K., Jamson, S., Tate, F. and Simpkin, B. (2008). ISA-UK: Final Report. Institute for Transport Studies, University of Leeds.
<http://webarchive.nationalarchives.gov.uk/20120215001100/http://www2.dft.gov.uk/pgr/roads/vehicles/intelligentspeedadaptation/fullreport.pdf>
- Castells, M. (2001). The Internet Galaxy: Reflections on the Internet, Business and Society. Oxford University Press, 2001.
- Cochran, W.G. (1977). Sampling techniques. 3rd edition. Wiley & Sons, New York.
- COWI, ECN, Ernst & Young Europe, ECORYS (2006). Cost Benefit assessment and prioritization of vehicle safety technologies. Final Report.
https://ec.europa.eu/transport/road_safety/sites/default/files/pdf/projects_sources/vehicle_safety_technologies_final_report.pdf
- 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. and Knippling, R.R. (2006). The 100-Car Naturalistic Driving Study, Phase II – Results of the 100-Car Field Experiment. Virginia Tech Transportation Institute, Blacksburg, VA, NHTSA DOT HS 810 593. <https://www.nhtsa.gov/sites/nhtsa.gov/files/100carmain.pdf>

- Draskóczy, M., Carsten, O. and Kulmala, R. (1998). Road safety guidelines. Deliverable B5.2 of CODE project (TR1103). Atkins Wootton Jeffreys, Birmingham, UK.
- EIB (European Investment Bank) and EC (European Commission) (2005). RAILPAG, Rail Project Appraisal Guidelines. http://www.eib.org/attachments/pj/railpag_en.pdf
- Elvik, R., Christensen, P. and Amundsen, A. (2004). Speed and road accidents: an evaluation of the Power Model. TOI Research Report 740/2004, Institute of Transport Economics, Oslo. <https://www.toi.no/publications/speed-and-road-accidents-an-evaluation-of-the-power-model-article17882-29.html>
- Everitt, B. and Dunn, F. (2000). Applied multivariate data analysis. Arnold Publication.
- Faber, F., Jonkers, E., Ljung-Aust, M., Benmimoun, M., Regan, M., Jamson, S. and Dobberstein, J. (2011). EuroFOT Deliverable 6.2, Analysis methods for user related aspects and impact assessment on traffic safety, traffic efficiency and environment. https://www.eurofot-ip.eu/download/library/deliverables/eurofotsp620110628v10dld62_analysis_methods.pdf
- Featherman, M.S. and Pavlou, P.A. (2003). Predicting e-Services Adoption: A Perceived Risk Facets Perspective. International Journal of Human-Computer Studies, 59(1).
- FESTA (2008a). A Comprehensive Framework of Performance Indicators and their Interaction. FESTA Deliverable D2.1
- FESTA (2008b). Performance indicator matrix table. FESTA Deliverable D2.1.
- FESTA (2008c). Data Requirements for FOT Methodology. FESTA Deliverable D2.2.
- FESTA (2008d). Data analysis and modelling. FESTA Deliverable D2.4.
- FESTA (2008e). Socio-economic impact assessment for driver assistance systems. FESTA Deliverable D2.6.
- FESTA (2008f). Common vision regarding vehicle systems FOTs. FESTA Deliverable D3.
- FESTA (2008g). Common vision regarding cooperative systems FOTs. FESTA Deliverable D4.
- FESTA (2008h). Common vision regarding nomadic devices FOTs. FESTA Deliverable D5.
- FGSV (2001). Forschungsgruppe für Strassen- und Verkehrswesen. Handbuch für die Bemessung von Straßenverkehrsanlagen. FGSV, Köln. <http://www.fgsv.de>
- Fitch, G.M., Rakha, H.A., Arafeh, M., Blanco, M., Gupta, S.K., Zimmermann, R.P. and Hanowski, R.J. (2008). Safety Benefit Evaluation of a Forward Collision Warning System, US DOT, NHTSA, Report No DOT HS 810 910, Washington, D.C.

- <https://one.nhtsa.gov/DOT/NHTSA/NRD/Multimedia/PDFs/Crash%20Avoidance/2008/DOT-HS-810-910.pdf>
- FOT-Net Data and CARTRE (2019). Data Sharing Framework Version 1.1. <https://www.connectedautomateddriving.eu/data-sharing/>
- Gibbons, J.D. and Chakraborti, S. (2003). Nonparametric statistical inference. 4th edition. CRC.
- Goldstein, H. (2003). Multilevel statistical models. 3rd edition, Arnold, London.
- Grudin, J. (1992) Utility and usability: research issues and development contexts. Interacting with Computers. Vol 4 (2), pp. 209-217.
- Gudmundsson, H. (2005). Mobility as a Policy Concept. In: Thomsen T.U., Nielsen L.D. and Gudmundsson H. (Eds.), Social Perspectives on Mobility. Hampshire, Ashgate Publishing Limited, pp. 107-126.
- Hanowski, R. J., Blanco, M. Nakata, A., Hickman, J.S., Schaudt, W.A., Fumero, M.C., Olson, R.L., Jermeland, J., Greening, M., Holbrook, G.T., Knipling, R.R. and Madison, P. (2008). The Drowsy Driver Warning System Field Operational Test: Data Collection Methods. Virginia Tech Transportation Institute. <https://vtechworks.lib.vt.edu/handle/10919/55072>
- HCM (2000). Highway Capacity Manual. Transportation Research Board, Washington D.C.
- Innamaa, S., Hakonen, M., Engelbrektsson, P., Fruttaldo, S., Mascolo, J., Morris, A., Pagle, K., Schröder, U., Vega, H. and Welsh, R. (2011). Impacts on Mobility – Preliminary Results. TeleFOT deliverable D4.4.2, European Commission, Information Society and Media, 7th Framework Programme. <https://cordis.europa.eu/docs/projects/cnect/7/224067/080/deliverables/001-TeleFOTD442Impactsonmobilitypreliminaryresultsv11.pdf>
- Innamaa, S., Axelson-Fisk, M., Borgarello, L., Brignolo, R., Guidotti, L., Martin Perez, O., Morris, A., Paglé, K., Rämä, A., Wallgren, P. and Will, D. (2013). Impacts on Mobility – Results and Implications. TeleFOT deliverable D4.4.3, European Commission, Information Society and Media, 7th Framework Programme. https://cris.vtt.fi/ws/portalfiles/portal/52616773/TeleFOT_D443_Impacts_on_mobility_results_and_implications_v10.pdf
- Innamaa, S., Smith, S., Barnard, Y., Rainville, L., Rakoff, H., Horiguchi, R. and Gellerman, H. (2018). Trilateral Impact Assessment Framework for Automation in Road Transportation. https://connectedautomateddriving.eu/wp-content/uploads/2018/03/Trilateral_IA_Framework_April2018.pdf

- Jacobson, I., Bylund, S., Jonsson, P. and Ehneboom, S. (1995). Modeling with Use Cases: Using contracts and use cases to build pluggable architectures. *Journal of Object Oriented Programming*, 8(2), pp. 18-24.
- Jamson, S., Chorlton, K., Gelau, G., Schindhelm, R., Johansson, E., Karlsson, A-S, Metz, B., Tadei, R., Benmimoun, M., Val, C., Regan, M., Wilschut, E. and Brouwer, R. (2009). Experimental Procedures. Deliverable 4.2 of EuroFOT. Ford Research & Advanced Engineering Europe, Aachen, Germany.
- Kulmala, R. and Rämä, P. (2010). Mobility Data Analysis Plan. TeleFOT deliverable D4.4.1, European Commission, Information Society and Media, 7th Framework Programme.
<https://cordis.europa.eu/docs/projects/cnect/7/224067/080/deliverables/002-TeleFOTD441MobilityDataAnalysisPlan.pdf>
- Lebart, L., Morineau, A. and Piron, M. (1997). *Statistique exploratoire multidimensionnelle*. Dunod, Paris.
- LeBlanc, D., Sayer, J., Winkler, C., Ervin, R., Bogard, S., Devonshire, J. Mefford, M., Hagan, M., Bareket, Z., Goodsell, R., and Gordon, T. (2006). Road Departure Crash Warning System Field Operational Test: Methodology and Results. Report UMTRI-2006-9-1. University of Michigan Transportation Research Institute, Ann Arbor, MI.
- Lecoutre, J.P. and Tassi, P. (1987). *Statistique non paramétrique et robustesse*. Econometrica, Paris.
- Ljung-Aust, M. L., Barnard, Y., Chen, H., Gellerman, H., Jonkers, Thomasson, Westhoff, D., Zlocki, A., Bärman, J. and Svanberg, E. (2014). D3.2 – Working Group Additions to FOT Methodology. FOT-Net 2 Deliverable.
<https://cordis.europa.eu/docs/projects/cnect/3/269983/080/deliverables/001-FOTNet2D32AdditiontoFOTMethodologyv9.pdf>
- Michon, J.A. (1985). A critical view of driver behaviour models: what do we know, what should we do? In: L. Evans and R. C. Schwing (Eds.). *Human Behaviour and Traffic Safety*, pp. 485-524. New York: Plenum Press.
- Malta, L., Ljung-Aust, M., Faber, F., Metz, B., Saint Pierre, G., Benmimoun, M. and Schäfer, R. (2012). Final results: Impacts on traffic safety. EuroFOT Deliverable 6.4, EuroFOT consortium.
https://www.eurofot-ip.eu/download/library/deliverables/eurofotsp620121121v11dld64_final_results_impacts_on_traffic_safety.pdf
- Najm, W.G., Stearns, M.D., Howarth, H., Koopman, J. and Hitz, J. (2006). Evaluation of an automotive rear-end collision avoidance system. DOT HS 810 569. DOT VNTSC-NHTSA-06-01. Washington, DC: U.S. Department of Transportation, National Highway Traffic Safety Administration.
<https://www.nhtsa.gov/DOT/NHTSA/NRD/Multimedia/PDFs/Crash%20Avoidance/2006/HS910569.pdf>

- Nellthorp, J., Mackie, P.J. and Bristow, A.L. (1998). Measurement and Valuation of the Impacts of Transport Initiatives, Deliverable D9, EUNET Project, EI Fourth Framework RTD Programme. ITS, University of Leeds, Leeds, UK. https://ec.europa.eu/environment/archives/tremove/pdf/measurement_maintext.pdf
- Nielsen, J. (1993). Usability engineering. Boston, MA: Morgan Kaufmann Publishers.
- Nilsson, G. (2004). Traffic safety dimensions and the power model to describe the effect of speed on safety. Bulletin 221. Lund Institute of Technology, Lund University. <https://portal.research.lu.se/ws/files/4394446/1693353.pdf>
- Nilsson, M. and Küller, R. (2000). Travel behaviour and environmental concern. Transportation Research Part D 5, pp. 211–234.
- NHTSA Benefits Working Group (1996). Preliminary Assessment of Crash Avoidance Systems Benefits. NHTSA, Washington D.C. http://webpages.eng.wayne.edu/~ad5781/CrashAvoidance/CrashAvoid_PreBenefit.pdf
- Olson, R. L., Hanowski, R. J., Hickman, J. S. and Bocanegra, J. (2009). Driver Distraction in Commercial Vehicle Operations. FMCSA-RRR-09-042. Center for Truck and Bus Safety Virginia Tech Transportation Institute, Blacksburg, VA. <https://www.fmcsa.dot.gov/sites/fmcsa.dot.gov/files/docs/DriverDistractionStudy.pdf>
- Orban, J., Hadden, J., Stark, G. and Brown, V. (2006). Evaluation of the Mack intelligent vehicle initiative Field Operational Test; Final Report. FMCSA-06-016. US Department of Transportation, Federal Motor Carrier Safety Administration. Battelle Memorial Institute, Columbus, Ohio. <https://rosap.ntl.bts.gov/view/dot/51>
- Östlund, J., Peters, B., Thorslund, B., Engström, J., Markkula, G., Keinath, A., Horst, D., Juch, S., Mattes, S. and Foehl, U. (2005). Driving performance assessment methods and metrics. AIDE Deliverable 2.2.5. http://www.aide-eu.org/pdf/sp2_deliv_new/aide_d2_2_5.pdf
- Polit, D.F., Beck, C.T. and Hungler, B.P. (2001), Essentials of Nursing Research: Methods, Appraisal and Utilization. 5th Ed., Philadelphia: Lippincott Williams & Wilkins.
- Regan, M., Triggs, T., Young, K., Tomasevic, N., Mitsopoulos, E., Stephan, K. and Tingvall, C. (2006). On-road evaluation of intelligent speed adaptation, following distance warning and seatbelt reminder systems: Final Results of the Australian TAC SafeCar Project. Volume 1: Report 253. Monash University Accident Research Centre, Melbourne. https://www.monash.edu/_data/assets/pdf_file/0020/217262/On-road-evaluation-of-Intelligent-Speed-Adaptation,-Following-Distance-Warning-and-Seatbelt-Reminder-Systems-final-results-of-the-TAC-SafeCar-project.pdf
- Saad, F. (2006). Some critical issues when studying behavioural adaptations to new driver support systems. Cognition, Technology and Work, 8, pp. 175-181.
- Sagberg, F., Eenink, R., Hoedemaeker, M., Lotan, T., Van Nes, N., Smokers, R. and Winkelbauer, M. (2011). Recommendations for a large-scale European naturalistic

- driving observation study. PROLOGUE Deliverable D4.1. Oslo, Norway: TØI Institute of Transport Economics.
https://prologue.kfv.at/fileadmin/content/Dokumente/PROLOGUE_D4.1_01.pdf
- Särndahl, C.-E., Swensson, B. and Wretman, J. (1992). Model assisted survey sampling. Springer Verlag.
- Shakel, B. and Richardson, S.J. (1991) Human factors for informatics usability. Cambridge University Press.
- Spinney, J.E.L., Scott, D.M. and Newbold, K.B. (2009). Transport mobility benefits and quality of life: A time-use perspective of elderly Canadians. *Transport Policy* 16, pp 1–11.
- Vanclay, F., Bains, J. and Taylor J. (2013). Principles for ethical research involving humans: ethical professional practice in impact assessment Part 1: Impact Assessment and Appraisal (31) 4: 243-253.
- Victor, T. W., Bärngman, J., Hjälm Dahl, M., Kircher, K., Svanberg, E., Hurtig, S., Gellerman, H. and Moeschlin, F. (2010). Sweden–Michigan naturalistic field operational test (SeMiFOT) phase 1: Final report. SAFER Report 2. Gothenburg, Sweden: SAFER.
- Volvo Trucks North America, Inc. (2005). Volvo Trucks Field Operational Test: Evaluation of Advanced Safety Systems for heavy truck tractors. Final Report. DTFH61-99-X-00102. Volvo Trucks North America, Inc., Greensboro, NC.
<https://rosap.ntl.bts.gov/view/dot/3686>
- Wannacott, T. and Wannacott, R. (1990). Introductory statistics for business and economics. 4th edition. John Wiley & Sons.
- Wasserman L. (2007). All of nonparametric statistics. Springer.
- Wilmink, I., Janssen, W., Jonkers, E., Malone, K., van Noort, M., Klunder, G., Rämä, P., Sihvola, N., Kulmala, R., Schirokoff, A., Lind, G., Benz, T., Peters, H. and Schonebeck, S. (2008). Impact assessment of Intelligent Vehicle Safety Systems. Deliverable D4 eIMPACT: Socio-economic impact assessment of stand-alone and co-operative Intelligent Vehicle Safety Systems (IVSS) in Europe, EU 6th Framework Programme Contract no.: 027421. TNO, Delft.

Annex A FOT Implementation Plan (FOTIP)

(To be read in conjunction with [Chapter 2](#) of the FESTA Handbook)

FOT Teams and People

1. Research Institute contracted to run FOT
2. Project Manager
3. Research Team
4. Technical Support Team
5. Administrative Support Team
6. Project Steering Committee
7. Project Management Team
8. Accounting/Auditing Advisor
9. Legal and Ethical Advisors
10. Sub-Contractors
11. Public Relations and Communications Advisor
12. Project Sponsor(s)

Activity 1: Convene FOT teams and people		
Tasks and Sub-Tasks	Person/Team/Organisation Responsible for Activity	Done
1.1 Appoint the FOT project manager	Research Institute contracted to run FOT	<input type="checkbox"/>
1.2 Appoint the research team	Project Manager	<input type="checkbox"/>
1.3 Appoint the technical support team	Project Manager, Project Steering Committee	<input type="checkbox"/>
1.4 Appoint the administrative support team	Project Manager	<input type="checkbox"/>
1.5 Appoint team leaders in each of the research, technical and administrative teams	Project Manager	<input type="checkbox"/>
1.6 Appoint the project steering committee	Project Manager, Project Steering Committee, Project Management Team, Public Relations and Communications Advisor, Project Sponsor(s)	<input type="checkbox"/>
1.7 Appoint the project management team (for day-to-day management	Project Manager	<input type="checkbox"/>
1.8 Appoint the accounting/auditing advisor	Project Manager, Project Management Team	<input type="checkbox"/>
1.9 Appoint a legal and ethics advisor	Project Manager, Project Management Team	<input type="checkbox"/>
1.10 Appoint sub-contractors	Project Manager, Project Management Team	<input type="checkbox"/>
1.11 Appoint a public relations/communications advisor	Project Manager, Project Management Team	<input type="checkbox"/>
1.12 Sign off on the agreed research and support structure	Project Manager, Project Management Team, Administrative Support Team, Accounting/Auditing Advisor, Project Sponsor(s)	<input type="checkbox"/>

Critical Considerations (the “dos” and “don’ts”)*(Italics emphasise the most important items)*

- ✓ While the project manager must have knowledge of all activities, ensure that critical knowledge is not vested in just one person. Personnel, including the project manager, may leave the project. *Ensure that there is “standby” for all key research and management roles within the FOT.*
- ✓ Appoint early someone to deal with human participants/ethics committee issues.
- ✓ Include in the research team someone who is a “gizmo” expert – who has up-to-date knowledge about current ICT/ITS developments and capabilities. Database and geographical information system (GIS) expertise is also critical.
- ✓ Ensure the project management team meets regularly (about once a month) to resolve research issues, monitor timelines and budgets, and resolve administrative, technical and other issues.
- ✓ *Choose contractors that can guarantee that, if a staff member leaves or is ill, there is sufficient expertise and capacity to maintain project continuity.*
- ✓ Maintain good relations with other partners involved in the FOT.
- ✓ *Ensure that the FOT evaluation process will be, and be recognised as, independent.*
- ✓ It is not necessary to appoint all teams/people at the same time – appointments should coincide with project needs, It is, however, necessary that the project management team is able to ensure coverage of the different aspects, eventually asking for support from specific experts in specific domains.
- ✓ Identify a final internal arbiter, acceptable to all parties, who can resolve scientific, administrative, legal and other disputes.
- ✓ Decide early in the project the frequency and timing of project Steering Committee meetings.

General Advice

Although this Activity precedes Activity 2, the choice of teams and people will be determined to some extent by the aims and objectives of the FOT.

- ✓ Appoint a project manager with excellent research, project management and communication skills. (Note: In some FOTs, the FOT project manager is responsible for both the administrative and scientific management of the FOT. In other FOTs, a senior researcher may be responsible for the scientific, but not the administrative, management of the FOT. This requirement will depend on the scale of the FOT.)
- ✓ The research team should be multi-disciplinary and would typically include psychologists, civil, mechanical, electrical and electronics engineers, statisticians, human factors experts, traffic safety experts, and socio-economic modelling experts.
- ✓ The technical support team would normally include computer software engineers, communications engineers, mechanical, traffic, civil and electronic engineers, and GIS experts.
- ✓ The project Steering Committee sets the strategic direction of the project and keeps it aligned with the project aims and objectives. Normally it would include the FOT project manager, selected members of the research and project management teams (e.g. team leaders), along with key stakeholders and the sponsor(s). Members should have authority to commit their organisations to the aims, objectives and implementation of the FOT. For smaller FOT projects, the stakeholder committee may not be necessary.

- ✓ The project management team is led by the FOT project manager and includes selected members of the research (e.g. team leaders), technical and administrative teams.
- ✓ A legal advisor should support the FOT over the full duration of the project (a lawyer's office providing advice whenever needed is sufficient). Legal knowledge must be available on the legal situation in the country in which the FOT is conducted.
- ✓ Define all necessary steps including all activities as soon as possible. Even little things can cause big problems.
- ✓ Adapt regularly the planning of analysis tasks in accordance with deviations from the original project plan; elaborate on the consequences.
- ✓ Open communication with regards to project monitoring to reduce expectations, should things go wrong in the implementation and data acquisition phase.
- ✓ It is recommended that experts of phases not immediately operating (e.g. data analysis) are involved in the overall FOT/NDS process from the very beginning, not only focusing on the specific task.
- ✓ The same is also recommended involving the right people in the interactions and iterations between planning, using, and analysing.
- ✓ The status of the project should be diffused to all relevant team members, including those not yet directly involved.

Activity 2: Define aims, objectives, research questions and hypotheses

Tasks and Sub-Tasks	Person/Team/Organisation Responsible for Activity	Done
2.1 Define the aims and objectives of the FOT, in conjunction with relevant stakeholders	Project Manager, Research Team, Technical Support Team, Project Steering Committee	<input type="checkbox"/>
2.2 Identify systems and functions to be tested	Project Manager, Research Team, Technical Support Team, Project Steering Committee	<input type="checkbox"/>
2.3 Identify use cases/situations in which systems and functions are to be tested	Project Manager, Research Team, Technical Support Team, Project Steering Committee	<input type="checkbox"/>
2.4 Define the research questions and prioritise them	Project Manager, Research Team, Technical Support Team, Project Steering Committee	<input type="checkbox"/>
2.5 Formulate hypotheses to be tested, deriving from research questions	Project Manager, Research Team, Technical Support Team, Project Steering Committee	<input type="checkbox"/>
2.6 Determine the constraints which may prevent the aims and objectives from being met	Team Project Manager, Research Team, Technical Support Team, Project Steering Committee, Project Management Team	<input type="checkbox"/>

<p>2.7 Define the final aims and objectives of the FOT, and seek agreement from relevant stakeholders</p>	<p>Project Manager, Research Team, Technical Support Team, Project Steering Committee</p>	<p><input type="checkbox"/></p>
<p>2.8 Sign off on the aims and objectives of the FOT</p>	<p>Project Manager, Research Team, Technical Support Team, Project Steering Committee, Project Management Team, Public Relations and Communications Advisor, Project Sponsor(s)</p>	<p><input type="checkbox"/></p>

<p>Critical Considerations (the “dos” and “don’ts”)</p> <ul style="list-style-type: none"> ✓ Be prepared for the potential for FOT aims and objectives to change when new administrations come in. ✓ Be prepared for potential conflict in objectives by different stakeholders; e.g. a car manufacturer wants a deep understanding of product use and driver behaviour and acceptance, while public authorities are more interested in determining the impact of system use on traffic and on the transport system. ✓ Check that use cases may be practically tested and reproduced with the requested statistical relevance in order to avoid that unfeasible use cases (and related research questions and hypotheses) are discarded in the successive steps. ✓ Ensure that all terms and phrases making up the research questions and hypotheses are clearly defined and unambiguous. This will facilitate interpretation of the FOT outcomes and comparisons with previous and future FOTS.
<p>General Advice</p> <ul style="list-style-type: none"> ✓ See the FESTA Handbook for further advice on defining the aims, objectives, research questions and hypotheses for an FOT. ✓ Constraints which may prevent the aims and objectives from being met might include cost, lack of supporting infrastructure, time, willingness and commitment of key stakeholders to cooperate in providing supporting infrastructure, their likely support in promoting the aims and objectives of the FOT, the availability of appropriate data, etc. ✓ Commonly cited aims are: <ul style="list-style-type: none"> - Evaluate system(s) effectiveness in changing behaviour and performance - Evaluate driver acceptance of system(s), including willingness to purchase - Evaluate system technical operation - Stimulate societal demand for new technologies - Evaluate safety impacts - Evaluate environmental impacts - Evaluate impacts on traffic (e.g. congestion, mobility) - Evaluate socio-economic cost-benefits

<ul style="list-style-type: none"> - Evaluate commercial impacts (e.g. productivity, return on investment, direct cost savings, incremental revenues by getting more customers, customer loyalty, etc.) <p>✓ Defining the research questions and prioritising them at an early stage will ensure they stay at the focus of the FOT and help protect from subsequent “mission creep”.</p>

Activity 3: Develop FOT project management plan		
Tasks and Sub-Tasks	Person/Team/Organisation Responsible for Activity	Done
3.1 Define the project activities, tasks and sub-tasks	Project Manager, Project Management Team	<input type="checkbox"/>
3.2 Decide who is accountable for completion of activities, tasks and sub-tasks	Project Manager, Project Management Team	<input type="checkbox"/>
3.3 Determine timelines for completion of activities, tasks and sub-tasks	Project Manager, Project Management Team	<input type="checkbox"/>
3.4 Determine the budget for project activities, tasks and timelines	Project Manager, Project Management Team	<input type="checkbox"/>
3.5 Develop a project GANTT chart to guide project management	Project Manager, Project Management Team	<input type="checkbox"/>
3.6 Implement procedures for monitoring project activities, timelines, budgets and resources (e.g. project management team meetings)	Project Manager, Project Management Team	<input type="checkbox"/>
3.7 Undertake a risk assessment for the FOT and plan contingencies as required	Project Manager, Project Management Team, Risk Management Consultant	<input type="checkbox"/>
3.8 Determine sign-off procedures (meetings and documents) to ensure there is sign-off on all critical decisions and stages of FOT by all relevant parties	Project Manager, Project Management Team	<input type="checkbox"/>
3.9 Agree on project issues which are confidential, and implement mechanisms for safeguarding their confidentiality	Project Manager, Project Management Team, Project Sponsor(s)	<input type="checkbox"/>
3.10 Develop a manual for conducting the FOT that documents critical procedural knowledge	Project Manager, Research Team, Technical Support Team	<input type="checkbox"/>

Critical Considerations (the “dos” and “don’ts”)

- ✓ *Include in the total budget some “contingency” that can be used to pay for unforeseen activities and tasks (especially meetings) that cannot be anticipated. 5 -10 percent of the total project cost is recommended. Different elements of the project may require different proportions of this contingency. It should be held and allocated by the project manager, not sub-activity leaders or partners.*
- ✓ *Identify and document in the GANTT chart the dependencies that exist between different activities, tasks and sub-tasks.*
- ✓ *Anticipate the need and budget for specialist consultants with skills and expertise that does not exist within the project team (e.g. training experts, software developers, lawyers etc.)*
- ✓ *Where relevant, anticipate changes to 3rd party vehicle fleets (e.g. vehicle upgrades and changes in operating routes) during the course of the FOT.*
- ✓ *Be aware that technical efforts are most likely to incur risk in terms of time and budget (especially the hardening up/refinement of systems, where these are developed within the FOT)*
- ✓ *Don’t under-estimate the time required and the cost of designing, running, analysing and de-commissioning the FOT. It will be greater than you think.*
- ✓ *Assume that some further modifications to, and fine tuning of, the project management plan will be required. It is impossible to foresee everything that is required in running an FOT.*
- ✓ *Develop procedural manuals for those conducting the FOT to ensure that, if staffs leave, all procedural knowledge does not leave with them. These should be developed for each activity.*

General Advice

- ✓ *Documentation of all project meetings is critical to record critical decisions, document the lessons learnt and justify possible blowouts in budgets and timelines.*
- ✓ *A budgeting structure that accommodates the uncertainties associated with running FOTs is desirable; e.g. a series of prospective budgets for each critical stage of the FOT.*
- ✓ *Be aware that in some jurisdictions project papers from publicly funded projects are public documents and copies can be requested by members of the public.*

Activity 4: Implement procedures and protocols for communicating with stakeholders

Tasks and Sub-Tasks	Person/Team/Organisation Responsible for Activity	Done
4.1 Commission a communications advisor to design communications plan	Project Manager, Project Management Team	<input type="checkbox"/>
4.2 Develop and implement a communications plan	Project Manager, Project Management Team, Public Relations and Communications Advisor	<input type="checkbox"/>

4.3 Appoint media spokespeople	Project Manager, Project Management Team, Project Steering Committee	<input type="checkbox"/>
4.4 Sign off on agreed communication protocols	Project Manager, Project Management Team, Public Relations and Communications Advisor, Project Sponsor(s)	<input type="checkbox"/>

Critical Considerations (the “dos” and “don’ts”)

- ✓ *Assume that you will be mis-represented by the media. Try and limit media attention until the data collection is complete.*
- ✓ *Agree in the contract with the sponsor who is responsible for press releases and dissemination of information and results.*
- ✓ *FOTs attract a lot of media attention. Provide adequate time and budget for unsolicited communication with stakeholders, especially with the media.*
- ✓ *Ensure that the project steering committee has input to the communications plan.*
- ✓ *Ensure that there is appropriate control of communication with the media, through the appointed media spokesperson. For EU projects, involving multiple partners, it may be necessary to appoint more than one media spokesperson.*
- ✓ *Everyone involved in the project must know who the media spokesperson is.*
- ✓ *The media spokesperson should consult with the project management group before speaking to the media, especially on sensitive issues.*
- ✓ *Provide media training for appointed spokespeople.*
- ✓ *Build political support for the FOT early in the project, and maintain it during and after the FOT.*
- ✓ *Be aware that there may be some key stakeholders who believe that FOTs are an impediment to system rollout. These people, in particular, must be made aware of the rationale for FOTs.*
- ✓ *Plan to have some results available at early stages of the project. If desirable, they should be released to an informed audience (e.g. at a conference), but not to the media as they could contaminate subsequent data collection.*
- ✓ *Plan for annual public meetings, and a project website, to disseminate information and findings.*
- ✓ *Don't undermine the scientific integrity of the research programme by mis-timing communications with the media and other stakeholders.*
- ✓ *Have a response prepared in case of serious incidents, such as a crash involving a test vehicle. Anticipate media contact between the media and participant drivers.*
- ✓ *Be aware that fleet/truck drivers may be more inclined to disclose opinions to the media if asked.*

<p>General Advice</p> <ul style="list-style-type: none"> ✓ <i>Open communication with key stakeholders is important at an early stage of the FOT to ensure that the aims and objectives of the FOT are clear, that stakeholders are committed to the project, and that the aims and objectives of the FOT are not misquoted, misrepresented or misunderstood.</i> ✓ There should be an agreed minimum level of transparency and result sharing in the FOT – avoid “confidential FOTs”. ✓ It may be beneficial to engage a professional press office to handle external communications, particularly with the media. ✓ FOT drivers and FOT researchers are usually of most interest to the media. ✓ Decide in advance with stakeholders a minimum time for approval for statements released to the media. ✓ Be prepared for the possibility that politicians may at times want to veto communications between the FOT project team, the media and other stakeholders. ✓ Building political support outside the project can help provide protection against strong partners/sponsors. ✓ <i>Early negative media attention may have a significant impact on participant recruitment and/or colour participant expectations of system performance. Try to prevent any media awareness until after the recruitment phase is complete.</i>

Activity 5: Design the Study		
Tasks and Sub-Tasks	Person/Team/Organisation Responsible for Activity	Done
<p>5.1 Become familiar with the methods, measures and procedures of previous FOTs:</p> <ul style="list-style-type: none"> ○ Read the FESTA handbook ○ Attend FOT-Net webinars and networking events ○ Talk to experts who have conducted FOTs previously ○ Review the relevant literature 	Project Manager, Technical Support Team	•
<p>5.2 Identify the performance indicators necessary to test the hypotheses derived in Activity 1</p>	Project Manager, Research Team, Technical Support Team	•
<p>5.3 Select measures (objective and subjective) that allow performance indicators to be derived to test the hypotheses</p>	Project Manager, Research Team, Technical Support Team	•

<p>5.4 Identify the sensors and sensor requirements for obtaining the required measures</p>	<p>Project Manager, Research Team, Technical Support Team</p>	<p>•</p>
<p>5.5 Design the experimental methods, tools and procedures for testing the hypotheses</p>	<p>Project Manager, Research Team, Technical Support Team</p>	<p>•</p>
<p>5.6 Define methods, tools, requirements and procedures for acquiring, storing, transferring, de-coding, reducing/transcribing, filtering, backing up and verifying the data</p>	<p>Project Manager, Research Team, Technical Support Team</p>	<p>•</p>
<p>5.7 Define methods, tools and procedures for analysing the data</p>	<p>Project Manager, Research Team, Technical Support Team</p>	<p>•</p>
<p>5.8 Determine optimal sample size (conduct power analyses) to ensure sufficient statistical power</p>	<p>Project Manager, Research Team</p>	<p>•</p>
<p>5.9 Select models for estimating the potential safety, environmental and other benefits of the technologies tested</p>	<p>Project Manager, Research Team</p>	<p>•</p>
<p>5.10 Sign off on study design, methods and tools, questionnaires and associated procedures</p>	<p>Project Manager, Project Management Team, Project Steering Committee, Legal and Ethical Advisors, Project Sponsor(s)</p>	<p>•</p>

<p>Critical Considerations (the “dos” and “don’ts”)</p> <ul style="list-style-type: none"> ✓ <i>Ensure that necessary historic data (e.g. data on vehicle speeds on certain roads) is available for baseline comparisons or Cost Benefit Analysis.</i> ✓ <i>Where relevant, allow sufficient time between vehicle allocations for system maintenance and verification, servicing and repairs to be undertaken.</i> ✓ <i>Accept that it is impossible to design a perfect FOT. Many practical issues – including time and money – will constrain the final experimental design.</i> ✓ <i>Remember that an FOT is not an experiment – control is limited, and counterbalancing may not be possible.</i> ✓ <i>Design into the FOT a contingency plan, in case there is an unexpected requirement to reduce or increase the scope of the study (e.g. to save money or time).</i> ✓ <i>Employ a multidisciplinary team in developing hypotheses that includes researchers and people with expert knowledge about the systems to be tested.</i> ✓ <i>Design the study in a way that allows for direct comparisons to be made between objective data (logged by the platform) and participative data (collected through questionnaires, focus groups etc.).</i> ✓ <i>Keep to an acceptable minimum the number and size of questionnaires that must be completed by participants at different points of the study, to maximise the likelihood of them being completed. A sub-2-hour completion duration is a useful target, as longer sessions may tend to remind participants that they are part of a scientific study.</i>

<ul style="list-style-type: none"> ✓ <i>Don't be tempted to reduce the sample size in order to save money – conducting a study with too few participants leads to a lack of statistical power to detect effects, and may ultimately be a waste of time and money.</i> ✓ Make sure that everyone understands the FOT study design, so that they appreciate the timing issues and the consequences of wanting to make changes to it, e.g. if wanting to reduce the scope of the study. ✓ Delays in one area of the programme cannot necessarily be made up by making sacrifices to other areas. ✓ <i>Don't assume that FOT users are the only ones who will use the FOT platforms.</i> ✓ Don't be pressured into changing the design of the study if, in doing so, it compromises the scientific integrity of the study. ✓ <i>When performing the sample size calculations, allow for participant attrition; e.g. if using fleet drivers, some may leave the company during the FOT period.</i> ✓ Where hypotheses are not supported, consider conducting a process evaluation. This can help determine whether the system did not work, or whether any implementation issues may have impacted on the results. Plan for annual public meetings, and a project website, to disseminate information and findings.
<p>General Advice</p> <ul style="list-style-type: none"> ✓ See the FESTA Handbook for detailed advice on designing the research study. ✓ See the FESTA Handbook reference list for published reports on previous FOTs. ✓ Where it is not possible, for ethical, practical or safety reasons, to investigate an issue in an FOT, consider safe alternative means for doing the research (e.g. simulators, test tracks). ✓ <i>The level of driver familiarity with the test vehicle may influence driver performance during the early stages of the FOT.</i> ✓ Ethical incentives that can be given to discourage driver attrition from the study should be agreed on early in the project. ✓ The models for estimating safety and other benefits may need to be updated in response to recent literature when making the estimation. ✓ <i>For the business sector, the commercial impact of the technologies deployed (e.g. in terms of productivity, return on investment, cost savings, incremental revenues by getting more customers, customer loyalty, etc.) will be important to evaluate.</i>

Activity 6: Identify and resolve FOT legal and ethical issues		
Tasks and Sub-Tasks	Person/Team/Organisation Responsible for Activity	Done
6.1 Seek specialist advice to identify relevant legal and ethical issues	Project Manager, Accounting/Auditing Advisor, Legal and Ethical Advisors	•
6.2 Resolve all legal and ethical issues that can be identified in advance	Project Manager, Project Management Team, Accounting/Auditing Advisor, Legal and Ethical Advisors	•

<p>6.3 Create contracts and/or agreements with all relevant parties (e.g. vehicle leasing organisations, suppliers, road operators, traffic centres, consultants, fleet managers, researchers etc.) for all relevant issues (e.g. data collection, provision and usage, theft, insurance, privacy, duty of care, property, disposal of vehicles after the study, etc.)</p>	<p>Project Manager, Project Management Team, Accounting/Auditing Advisor, Legal and Ethical Advisors</p>	<p>•</p>
<p>6.4 Seek approvals to conduct the study and store and possibly share test data from relevant ethics and privacy committees</p>	<p>Project Manager, Research Team, Technical Support Team, Legal and Ethical Advisors</p>	<p>•</p>
<p>6.5 Seek expert advice regarding liability issues and to ensure insurance provision is adequate for all foreseeable eventualities</p>	<p>Project Manager, Accounting/Auditing Advisor, Legal and Ethical Advisors</p>	<p>•</p>
<p>6.6 Ensure that vehicle type approval and warranty requirements are adhered to in spite of the modifications (implementation of data logging equipment and possibly systems to be evaluated, etc.)</p>	<p>Project Manager, Research Team, Legal and Ethical Advisors</p>	<p>•</p>
<p>6.7 Obtain informed consent of participants before they are allowed to participate in the FOT</p>	<p>Project Manager, Research Team, Technical Support Team</p>	<p>•</p>
<p>6.8 Sign off on all aspects of the FOT design and procedures pertaining to legal and ethical matters</p>	<p>Project Manager, Project Management Team, Accounting/Auditing Advisor, Legal and Ethical Advisors, Project Sponsor(s)</p>	<p>•</p>

<p>Critical Considerations (the “dos” and “don’ts”)</p> <ul style="list-style-type: none"> ✓ In terms of the project timeline, legal and ethical issues need to be considered in parallel from beginning to end (and indeed afterwards in terms of data protection), especially if the data is planned to be re-used in new projects. ✓ <i>There must be mutual agreement on the relative risks to all parties before contracts are signed.</i> ✓ <i>Double-check that the final design and conduct of the FOT accords with ethical and legal requirements in all jurisdictions in which the FOT will physically occur.</i> ✓ Ensure that all intellectual property issues are identified and resolved “up front”. ✓ Ensure permission to drive (and necessary insurance cover) restrictions are understood by all parties, particularly participants. ✓ Identify the conditions under which a participant will be expelled from the study, and ensure these are made known to participants before the FOT commences. ✓ Ensure that all participating drivers are fully licensed to drive the test vehicles. ✓ <i>Don’t forget about the need to adhere to contractual obligations and confidentiality agreements. FOTs often extend over long periods, making it easy to lose sight of obligations and agreements.</i>
--

<ul style="list-style-type: none"> ✓ Clarify participant responsibilities and the study's obligations to the participants. Participant responsibilities should include routine vehicle maintenance activities, e.g. checking fluid levels. ✓ <i>Ensure all relevant health and safety requirements of participants and the study team are met.</i> ✓ All project staff must understand who has access to project data, especially video data. If personal data is to be passed on to third parties, it needs to be anonymised (Note: with GPS and video data it may be very difficult to guarantee anonymity). ✓ <i>All study team members must understand the agreed response should a major incident, such as an accident, occur. Any media comment should only be made by the spokesperson.</i> ✓ Don't underestimate the complexity and time commitment involved in identifying and resolving the legal and ethical issues associated with the conduct of an FOT. ✓ <i>Ensure that all methods, tools, procedures and materials used in the study that require legal and ethics approval are approved by the Ethics Committee at appropriate points in the study.</i>
<p>General Advice</p> <ul style="list-style-type: none"> ✓ See Chapter 3 of this FESTA Handbook for detailed advice on legal and ethical issues.

Activity 7: Select and obtain FOT test platforms (vehicles, mobile devices, road side units, ...)		
Tasks and Sub-Tasks	Person/Team/Organisation Responsible for Activity	Done
7.1 Specify functional requirements, performance specifications and user requirements for the test platforms needed for the study	Project Manager, Research Team, Technical Support Team	•
7.2 Specify functional requirements and performance specifications for the integration into platforms of all systems needed for the FOT (FOT technologies, support technologies and data collection technologies), if these are not already in the platforms	Project Manager, Research Team, Technical Support Team	•
7.3 Select test platforms (makes and models) that meet the above requirements	Project Manager, Technical Support Team	•
7.4 Where relevant, purchase, lease, hire or borrow (where the driver owns the vehicle) the test vehicles and/or platforms	Project Manager, Accounting/Auditing Advisor	•
7.5 Sign off on selection and obtaining of test platforms	Project Manager, Technical Support Team, Project Management Team, Project Sponsor(s)	•

Critical Considerations (the “dos” and “don’ts”)

- ✓ The choice of platforms may well impinge on the selection of participants which, in itself, will impact on the research questions. Choice of platforms must be undertaken at an early stage in the project’s planning.
- ✓ *Consider obtaining extra test platforms. These can be used as spare items in case of failure and as “showcasing” platforms. The latter can be driven at appropriate times by politicians and other high ranking officials in positions of authority to promote and deploy the systems on a wider scale.*
- ✓ Be aware that vehicle choice may affect participant response if the test vehicle is significantly better/worse than the vehicle they are used to driving. Choose a conservative model.
- ✓ *Do consider vehicle maintenance requirements and the dealer network that is available in the FOT area. If the FOT will take place in a limited area, consider advising the local dealer(s) of the study. This may be important if a participant takes a test vehicle to a dealer to fix a problem.*

General Advice

- ✓ Where used, the test vehicle will vary, depending on the nature of the FOT. In some FOTs, the test vehicles will already contain mature OEM systems. In others, the systems will need to be developed (fully or partly) and integrated into the vehicles. In some FOTs, the systems will be integrated into drivers’ own vehicles; in others, they will be integrated into company fleet vehicles.
- ✓ The test platforms must be capable of hosting the technologies to be evaluated (OEM, aftermarket and nomadic) and the data logging and support systems.
- ✓ Carefully evaluate the trade-off between owned and leased equipments and vehicles to be used in FOTs in order to choose the most convenient solution for the specific responsible organisation.

Activity 8: Select and obtain systems and functions to be evaluated during the FOT (if they are not already implemented in the test platforms)

Tasks and Sub-Tasks	Person/Team/Organisation Responsible for Activity	Done
8.1 Develop selection criteria for choosing systems and functions (OEM, aftermarket and nomadic) to be tested (if the technologies to be tested have not already been selected by the sponsor; see General Advice column)	Project Manager, Research Team, Technical Support Team, Project Steering Committee	•
8.2 Use the above selection criteria to select and obtain systems to be tested	Project Manager, Research Team, Technical Support Team	•

<p>8.3 If commercial systems do not exist that meet the above criteria, develop functional requirements and performance specifications for systems that do (including for HMI and security issues)</p>	<p>Project Manager, Research Team, Technical Support Team and (if appropriate) consultant</p>	<p>•</p>
<p>8.4 Develop functional requirements and performance specifications for the infrastructure needed to support deployment of the technologies to be tested (e.g. digital maps, roadside units)</p>	<p>Project Manager, Research Team, Technical Support Team, Project Steering Committee, and (if appropriate) Consultant</p>	<p>•</p>
<p>8.5 Source infrastructure that meets the above functional requirements and specifications</p>	<p>Project Manager, Technical Support Team and (if appropriate) Consultant</p>	<p>•</p>
<p>8.6 Where infrastructure is not commercially available, develop supporting infrastructure that meets the above functional requirements and performance specifications</p>	<p>Project Manager, Research Team, Technical Support Team and (if appropriate) Consultant</p>	<p>•</p>
<p>8.7 If appropriate, issue Expressions of Interest/Requests for Tenders for provision of systems and supporting infrastructure</p>	<p>Project Manager, Project Management Team, Accounting/Auditing Advisor, Legal and Ethical Advisors</p>	<p>•</p>
<p>8.8 If appropriate select preferred tenderers, negotiate contracts and award contracts</p>	<p>Project Manager, Project Management Team, Accounting/Auditing Advisor</p>	<p>•</p>
<p>8.9 Decide what will be done with the test platforms, and the equipment in them, once the FOT has been completed</p>	<p>Project Manager, Research Team, Technical Support Team, Administrative Support Team, Project Steering Committee, Legal and Ethical Advisors, Project Sponsor(s)</p>	<p>•</p>
<p>8.10 Sign off on selection and obtaining of systems and functions to be evaluated during the FOT</p>	<p>Project Manager, Technical Support Team, Project Management Team, Project Sponsor(s)</p>	<p>•</p>

<p>Critical Considerations (the “dos” and “don’ts”)</p> <ul style="list-style-type: none"> ✓ Do ensure that criteria for the selection of candidate systems (where this is appropriate) to be evaluated are developed in consultation with relevant stakeholders, in order to ensure that the systems to be tested meet the needs of all relevant stakeholders and are suitable for in-car use (this includes good interface design). ✓ Selection of systems must be undertaken with consideration of the data-logging system. If not, problems of interfacing may result. Ensure that data sharing and re-use are described in contracts. ✓ Beware of hidden costs of hardware and software development if these items are not originally designed for research purposes.
--

General Advice

- ✓ Criteria for selection of candidate systems in the FOT (if they have not been pre-selected by the sponsor) could include: likely safety or environmental benefit, likely benefit in increasing commercial productivity and efficiency, availability, compatibility with host vehicles, technical performance, cost, reliability, maintainability, likely acceptability to drivers, usability, compliance with relevant human factors/ergonomic guidelines, compliance with local legal requirements, compliance with relevant standards, crashworthiness etc.
- ✓ If prototype systems are tested, then estimates of durability, reliability, maintenance costs etc. of production systems will be difficult, and full Cost Benefit Analyses may not be possible.

Activity 9: Select and obtain data collection and transfer systems

Tasks and Sub-Tasks	Person/Team/Organisation Responsible for Activity	Done
9.1 Specify data to be logged (measures and sampling rate)	Project Manager, Research Team, Technical Support Team	•
9.2 Specify functional requirements and performance specifications for systems for collecting and transferring the data to be logged	Project Manager, Research Team, Technical Support Team	•
9.3 Source, purchase and/or develop systems for logging and transferring the data that meet the above functional requirements and performance specifications	Project Manager, Technical Support Team and (if appropriate) Sub-Contractors	•
9.4 Sign off on selection and obtaining of data collection and transfer system	Project Manager, Project Management Team, Legal and Ethical Advisors, Project Sponsor(s)	•

Critical Considerations (the “dos” and “don’ts”)

- ✓ Implement re-calibration procedures that will ensure accuracy of measurements/sensors over time and help prevent data drift issues.
- ✓ Plan for software upgrade and revision during the FOT and try to ensure that all software systems are updated together. Ideally, this should be possible remotely.
- ✓ Where used, in-vehicle data logging systems need to be unobtrusive, safe and secure – but they also need to be accessible to enable routine repairs.
- ✓ Where relevant, provide a location close to the participants for vehicle support.
- ✓ Data logging systems should offer continuous monitoring capabilities.
- ✓ Minimise user involvement in data download from test platforms.

<ul style="list-style-type: none"> ✓ Ensure boot-up time for test systems and data logging systems is sufficiently fast to prevent data loss at the beginning of each trip. ✓ Ensure that a common timestamp is used for all recorded data sources. ✓ Verify the definition of signals provided by 3rd parties (e.g. CAN message definitions by vehicle manufacturers) ✓ Do not allow data collection to proceed automatically without active confirmation of data capture and validity. This may include the generation of warning messages when out-of-tolerance data is recorded. ✓ Recognise that some data is much more important than others and should be given a relatively higher priority. ✓ Do keep a stock of spares for critical items, and anticipate that some components may become unobtainable during the study. ✓ Consider the opportunities for ad-hoc and post-hoc interrogation of raw data files to answer additional questions. This may not be possible if data collection is triggered.
<p>General Advice</p> <ul style="list-style-type: none"> ✓ The technologies fitted to test vehicles may also include supplementary technologies (such as sensor technologies; e.g. forward looking radars, GPS) that are needed e.g. to measure inter-vehicle following distances in order to determine whether speeds are free or constrained. ✓ See Chapter 3 of the FESTA Handbook on legal issues of data privacy to be aware of possible dangers and legal provisions.

Activity 10: Select and obtain support systems for FOT platforms		
Tasks and Sub-Tasks	Person/Team/Organisation Responsible for Activity	Done
10.1 Define the support systems needed (see General Advice column)	Project Manager, Research Team, Technical Support Team and (if appropriate) Consultant	•
10.2 Develop functional requirements and performance specifications for systems needed to support the study	Project Manager, Research Team, Technical Support Team and (if appropriate) Consultant	•
10.3 Where appropriate, develop functional requirements and performance specifications for the HMI, to ensure that the HMI for support systems is safe and user-friendly	Project Manager, Research Team, Technical Support Team and (if appropriate) Consultant	•
10.4 Source, purchase and/or develop support systems that meet the above functional requirements and performance specifications	Project Manager, Technical Support Team and (if appropriate) Sub-Contractors	•

10.5 Sign off on selection and obtaining of support systems for test platforms	Project Manager, Project Management Team, Project Sponsor(s)	•
---	--	---

<p>Critical Considerations (the “dos” and “don’ts”)</p> <ul style="list-style-type: none"> ✓ If possible, support systems should be capable of remote operation to allow e.g. remote system reboot. ✓ In the case of very large naturalistic studies it may not be practicable to intervene manually. In these cases do attempt to automate as much as possible. ✓ Anticipate data analysis requirements before specifying data to be logged (e.g. rates and resolution). ✓ Ensure that missing data are clearly indicated – e.g. if the data collection system malfunctions, missing data should NOT be indicated with a zero, where zero is a valid measure (e.g. speed). ✓ If in doubt about the final list of measures to be logged, log more parameters if performance of the data logging system or storage capacity are not affected. Consider the opportunities for ad-hoc and post-hoc interrogation of raw data files to answer additional questions. This may not be possible if data collection is triggered.
<p>General Advice</p> <ul style="list-style-type: none"> ✓ Support systems have multiple purposes: e.g. to display information to users; to automatically turn systems on and off where multiple systems are being tested and exposure to each is kept constant across drivers; for manually disabling systems in the event of malfunctions (i.e. “panic buttons”); for preventing use of systems by non-participants; for diagnosing system status and faults; etc.

Activity 11: Equip FOT test platforms with all systems		
Tasks and Sub-Tasks	Person/Team/Organisation Responsible for Activity	Done
11.1 Prepare a system installation/integration manual describing standardised procedures	Project Manager, Technical Support Team and (if appropriate) Sub-Contractors	•
11.2 Equip test platforms with the FOT systems to be evaluated (if not already installed)	Project Manager, Technical Support Team and (if appropriate) Sub-Contractors	•
11.3 Equip test platforms with data collection and transfer systems	Project Manager, Technical Support Team and (if appropriate) Sub-Contractors	•

<p>11.4 Equip platforms with FOT support systems (e.g. panic button, for turning systems off in a vehicle etc.)</p>	<p>Project Manager, Technical Support Team and (if appropriate) Sub-Contractors</p>	<p>•</p>
<p>11.5 Sign off on system integration activities, ensuring that all systems have been installed in accordance with the system installation/integration manual</p>	<p>Project Manager, Project Management Team, Project Sponsor(s)</p>	<p>•</p>

<p>Critical Considerations (the “dos” and “don’ts”)</p> <ul style="list-style-type: none"> ✓ Ensure that the computers running all systems (FOT, data collection and support) have sufficient computing power to avoid processing delays. ✓ <i>Ensure that all systems (FOT, data collection and support) operate identically across test platforms.</i> ✓ <i>Allow all new vehicles a burn-in period (around 1000km) so that vehicle faults, that could disrupt the FOT, can be detected.</i> ✓ Be aware that ‘identical’ platforms and sensors may perform differently due to variation in components, manufacturing variability and environmental conditions. Check for differences that may be critical for the FOT. ✓ Try and make all adaptations to test vehicles (e.g. fitment of novel display systems) invisible to reduce the likelihood of theft or behaviour modification by other drivers.
--

<p>Activity 12: Design and implement user feedback and reporting systems</p>		
<p>Tasks and Sub-Tasks</p>	<p>Person/Team/Organisation Responsible for Activity</p>	<p>Done</p>
<p>12.1 Design, develop and implement systems and procedures to allow users to report technical problems in a timely manner</p>	<p>Project Manager, Research Team, Technical Support Team</p>	<p>•</p>
<p>12.2 Design, develop and implement systems and procedures to allow users to provide feedback to researchers, in real time or retrospectively (e.g. usability problems, opinions of systems, confirmation that systems are operating as required etc.)</p>	<p>Project Manager, Research Team, Technical Support Team</p>	<p>•</p>
<p>12.3 Design, develop and implement systems and procedures that allow researchers to monitor participant progress (e.g. to ensure they are adhering to study requirements)</p>	<p>Project Manager, Research Team, Technical Support Team</p>	<p>•</p>

<p>12.4 Sign off on implementation of user feedback and reporting systems and procedures</p>	<p>Project Manager, Research Team, Technical Support Team, Project Management Team, Project Sponsor(s)</p>	<p>•</p>
---	--	----------

<p>Critical Considerations (the “dos” and “don’ts”)</p> <ul style="list-style-type: none"> ✓ Implement ‘user diaries’ to allow confirmation of user identity and trip details if this process cannot be automated. This may encourage users to behave less naturally. ✓ Implement a timetable for the timely collection of qualitative data so that participants don’t have to rely on their memories. ✓ Implement a timetable for periodic verification of collected log data and checking that systems/cars still operate correctly. ✓ <i>Anticipate that users may not complete diaries accurately or consistently and may fail to attend de-briefing interviews. Appoint user liaison staff as a single point of contact.</i> ✓ <i>Ensure that the project team can respond to emergencies and incidents on a 24/7 basis.</i> ✓ Do ask participants to announce when they are going on holiday or not using the platform for an extended period. ✓ Keep a record of all reported problems, and document these in relevant reports. ✓ Ensure that all feedback and reporting procedures are documented in a manual for quick reference by the research and technical support team as required. ✓ Consider whether you need to design, develop and implement a system to allow for the collection of fuel consumption information. ✓ Where fuel consumption is calculated manually, anticipate that drivers will not always use fuel cards, return fuel docketts or fill in the fuel logbook.
--

<p>Activity 13: Select, obtain and implement standard relational database for storing FOT data</p>		
<p>Tasks and Sub-Tasks</p>	<p>Person/Team/Organisation Responsible for Activity</p>	<p>Done</p>
<p>13.1 Design, develop and implement a database for storing data logged from the test platforms</p>	<p>Project Manager, Research Team, Technical Support Team</p>	<p>•</p>
<p>13.2 Design, develop and implement a database for storing the subjective data collected from participants (e.g. from questionnaires, focus groups, feedback lines etc.)</p>	<p>Project Manager, Research Team, Technical Support Team</p>	<p>•</p>
<p>13.3 Develop data navigation and visualisation tools</p>	<p>Project Manager, Research Team, Technical Support Team</p>	<p>•</p>

<p>13.4 Sign off on database for storing FOT data</p>	<p>Project Manager, Research Team, Technical Support Team, Project Management Team, Legal and Ethical Advisors, Project Sponsor(s)</p>	<p>•</p>
--	--	----------

<p>Critical Considerations (the “dos” and “don’ts”)</p> <ul style="list-style-type: none"> ✓ It is recommended to create a Data Management Plan describing the data management processes and e.g. access rights to the data. ✓ Before an FOT is launched, the database architecture should be reviewed by a system evaluator to ensure that all requirements are fulfilled. ✓ <i>Ensure copies are made of raw data, reduced raw data and all processed data files and store these securely, separate from the primary data store.</i> ✓ Ensure that raw and processed data items are well documented. ✓ Use an industry standard relational database to store the data. ✓ <i>Ensure that unauthorised access to the database is not possible. Preferably, do not give the database host an IP number.</i> ✓ Careful database design can reduce the need for post-collection manipulation if the database is designed to feed directly into a statistical package for data cleaning and analysis. ✓ <i>Decide early in the project how to manage post-project data. Issues to consider are: What happens to data when the project ends? Who will have data usage rights? Who can access it? Who pays for possible storage? In projects with large amounts of stored data (several terabytes), the cost to store and manage data is not insignificant, and all project partners might not have the means to handle it afterwards. Where data is taken off-line, determine what metadata should be kept, and how.</i> <p>General Advice</p> <ul style="list-style-type: none"> ✓ Basic legal advice on this issue is also provided in Chapter 3 of the FESTA handbook.

<p>Activity 14: Test all systems against functional requirements and performance specifications</p>		
<p>Tasks and Sub-Tasks</p>	<p>Person/Team/Organisation Responsible for Activity</p>	<p>Done</p>
<p>14.1 Develop “acceptance testing” protocols (see comment column)</p>	<p>Project Manager, Research Team, Technical Support Team</p>	<p>•</p>
<p>14.2 Test the systems for acceptance, using the acceptance testing protocol</p>	<p>Project Manager, Research Team, Technical Support Team</p>	<p>•</p>
<p>14.3 Develop a usability test plan for the purpose of assessing the systems and functions for usability</p>	<p>Project Manager, Research Team with consultant (if appropriate)</p>	<p>•</p>

<p>14.4 Conduct usability testing, using the usability testing plan, to ensure systems and functions are user-friendly and that they meet all usability assessment criteria</p>	<p>Project Manager, Research Team, with consultant (if appropriate)</p>	<p>•</p>
<p>14.5 Obtain or develop a valid and reliable ergonomic checklist</p>	<p>Project Manager, Research Team</p>	<p>•</p>
<p>14.6 Assess systems, using the ergonomic checklist, to ensure that they meet all relevant criteria</p>	<p>Project Manager, Research Team</p>	<p>•</p>
<p>14.7 Assess vehicles against relevant certification procedures to ensure that vehicles are safe, roadworthy and comply with all relevant National, State and Territory laws, treaties and other protocols</p>	<p>Project Manager, Technical Support Team with Consultant (if appropriate)</p>	<p>•</p>
<p>14.8 Ensure that all vehicle modifications that affect primary safety are signed off by a competent engineer or appropriate testing authority</p>	<p>Project Manager, Technical Support Team with Consultant (if appropriate)</p>	<p>•</p>
<p>14.9 Rectify all technical, usability, ergonomic and certification issues where deficiencies are noted</p>	<p>Project Manager, Research Team, Technical Support Team with Consultant (if appropriate)</p>	<p>•</p>
<p>14.10 Sign off on completion of all systems tests</p>	<p>Project Manager, Research Team, Technical Support Team, Project Management Team, Project Sponsor(s)</p>	<p>•</p>

<p>Critical Considerations (the “dos” and “don’ts”)</p> <ul style="list-style-type: none"> ✓ <i>Do not sign off on the outputs of any of the previous activities until all technologies have been tested and, where appropriate, refined.</i> ✓ <i>Be sure that all systems are designed so they do not drain the battery when the engine is not running.</i> ✓ <i>Be sure that retrofitted systems are properly secured and meet all relevant crashworthiness requirements.</i> ✓ <i>If sub-contractors are appointed to install or maintain test equipment, implement a quality assurance programme.</i> ✓ <i>Be aware that system clocks can drift significantly if left to run independently. Where feasible, use GPS time to correct system clock error.</i> ✓ <i>Implement procedures to ensure that alignment and calibration of sensors is maintained and tested in all potential weather conditions.</i> ✓ <i>Various guidelines, standards and checklists exist for assessing the ergonomic quality of the human-machine interface for ICT systems</i> ✓ <i>Be aware that some system components may become corrupted over time with continuous use (e.g. flash memory cards).</i> ✓ <i>Revisit the installation manual for all platforms.</i>
--

- ✓ Consider the need to obtain waivers/special licences from regulatory authorities for equipment that is non-compliant (e.g. radars that operate outside legal bandwidths).
 - ✓ Standard testing of vehicle modifications by a competent authority may be necessary with respect to safety features (e.g. proper deployment of airbags following modification to vehicle interiors).
 - ✓ *Be aware that some systems (e.g. displays) that are not OEM-installed may fail in automotive environments.*
 - ✓ *Where appropriate, test for radio frequency (RF) interference effects (e.g. from overhead tram wires), which may adversely affect system operation. Also ensure that normal vehicle systems (e.g. FM radio and remote locking) are not affected by installed equipment.*
 - ✓ Ensure that the computers powering the data collection system and support systems are powerful enough to ensure that the data sampling rate is consistent and at the rate specified.
 - ✓ Don't assume that OEM systems that are already installed in test vehicles have been ergonomically assessed against appropriate standards and guidelines. Ergonomic assessment of systems prior to system deployment can be useful in identifying ergonomic problems that may explain or confound treatment effects.
 - ✓ *Resolving any technical, usability, ergonomic, and certification issues may require several iterations. Do not underestimate the time required for this process.*
- General Advice**
- ✓ This activity is *not* about pilot testing – it is about testing the performance, security and reliability of systems – to ensure that all technologies to be deployed perform in accordance with the functional requirements and performance specifications developed for them in previous activities.
 - ✓ An Acceptance testing Protocol is a test protocol for testing that all systems to be used in the study (FOT systems, data collection systems and support systems) meet the functional requirements and performance specifications developed for them by the FOT project team, under all foreseeable operating conditions.
 - ✓ The term “usability” can mean different things to different people. The test plan should use a standard definition of usability (e.g. ISO 9241).
 - ✓ Be aware that the frequency used by some radar-based systems may interfere with the operation of other systems used by police, emergency services or other operators (or vice versa) when used in other countries or jurisdictions. This must be investigated where the FOT is conducted across State and international boundaries.

Activity 15: Develop FOT recruitment strategy and materials		
Tasks and Sub-Tasks	Person/Team/Organisation Responsible for Activity	Done
15.1 Develop a recruitment strategy, including user entry and exit requirements and procedures	Project Manager, Research Team, Legal and Ethical Advisors	•
15.2 Develop recruitment materials and procedures	Project Manager, Research Team, Public Relations and Communications Advisor	•

<p>15.3 Sign off on recruitment strategy, materials and procedures</p>	<p>Project Manager, Research Team, Project Management Team</p>	<p>•</p>
---	--	----------

<p>Critical Considerations (the “dos” and “don’ts”)</p> <ul style="list-style-type: none"> ✓ Consider whether participants should be representative of the relevant population to ensure generalisability of results. ✓ Assume that there will be an attrition rate of about 10 to 15% when using company employees, who come and go, and retire. ✓ Be aware that, when company employees change jobs within their companies, this may have a dramatic effect on their annual travel. ✓ If fleet drivers are recruited via a fleet owner or manager, it is also necessary to get buy-in from individual drivers. ✓ With respect to safety, select drivers who do not pose a risk to themselves, others or the project. Be aware of the potential for bias in the results. ✓ Do not underestimate the complexities involved in recruiting company employees. ✓ Be aware that some commercial operations may have employee turnover rates approaching 100% per annum. <p>General Advice</p> <ul style="list-style-type: none"> ✓ <i>The ethical requirements for recruitment of users may be difficult to adhere to when recruiting company employees.</i> ✓ Ideal companies to approach to recruit fleet vehicle drivers have the following characteristics: many vehicles; drivers have high mileage rates; drivers drive primarily in the geographical areas of interest in the FOT; and management has a commitment to the aims and objectives of the FOT. ✓ It is not possible in many countries to obtain personal information about drivers that can be used to screen them for inclusion in the study (e.g. has a drunk driving record). ✓ It may not be possible in some countries to obtain directly from car dealers the names of drivers of particular makes and models of vehicles. ✓ In some countries (e.g. France), potential participants must be screened by a registered doctor. ✓ <i>The recruitment materials and procedures will need to have been incorporated and approved as part of the FOT ethics and legal approval processes.</i>

Activity 16: Develop training and briefing materials		
Tasks and Sub-Tasks	Person/Team/Organisation Responsible for Activity	Done
16.1 Conduct training needs analysis (TNA) to identify training requirements of participants and other relevant actors	Project Manager, Research Team with Consultant (if appropriate)	•
16.2 Design and develop briefing and training materials, based on outputs of the TNA	Project Manager, Research Team with Consultant (if appropriate)	•
16.3 Design and develop an FOT system(s) user manual (if appropriate)	Project Manager, Research Team, Legal and Ethical Advisors	•
16.4 Design and document the procedures for delivery of the briefing and training to the FOT participants	Project Manager, Research Team	•
16.5 Sign off on training and driver (and company) briefing materials and delivery processes	Project Manager, Research Team, Project Management Team	•

Critical Considerations (the “dos” and “don’ts”)

- ✓ *Ensure that training programmes and briefing materials are designed in a way that does not confound experimental treatment effects.*
- ✓ Ensure all users understand all existing systems and functions to be used (including test systems).
- ✓ Don't underestimate the time required for the development of briefing and training materials – it is a time-consuming activity.
- ✓ When pre-testing the user-friendliness of a function a self-learning approach may be used.
- ✓ Be aware that an excess of training might affect the possibility to understand the short-term unintended effects of the system.
- ✓ Provide drivers with a mini operating manual to keep in the vehicle and prepare written materials (brochures, DVDs & CDs) that can be taken away after briefing sessions.
- ✓ Provide a written statement for the participants to keep (in the vehicle) which confirms their participation in the FOT and the nature of vehicle modifications – in case they are challenged by police or other authorities.

General Advice

- ✓ Look for examples of training and briefing materials used in a previous FOTs.
- ✓ *Refresher training may be required if FOT systems are not activated for several weeks or months into the FOT.*

✓ The training and briefing materials and procedures will need to have been incorporated and approved as part of the FOT ethics and legal approval processes.

Activity 17: Pilot test FOT equipment, methods and procedures		
Tasks and Sub-Tasks	Person/Team/Organisation Responsible for Activity	Done
17.1 Develop a protocol for pilot testing FOT equipment, methods, procedures, evaluation tools and materials (including training, briefing materials and data collection, downloading and analysis procedures)	Project Manager, Research Team, Technical Support Team, Legal and Ethical Advisors	•
17.2 Recruit, brief and train pilot participants	Project Manager, Research Team	•
17.3 Deploy a small sample of FOT platforms under a representative range of external conditions that will be experienced in the FOT, as per the pilot testing protocol	Project Manager, Research Team, Technical Support Team	•
17.4 Fine tune FOT platforms and technologies, systems, procedures, evaluation tools and protocols, as required, on the basis of the pilot data yielded	Project Manager, Research Team, Technical Support Team, Project Management Team	•
17.5 Sign off on pilot testing (green light for starting the main tests)	Project Manager, Research Team, Technical Support Team, Project Management Team	•

Critical Considerations (the “dos” and “don’ts”)

- ✓ *Do not truncate your pilot test plan, and do not underestimate the time required for comprehensive pilot testing. The importance of pilot testing cannot be overstated.*
- ✓ Undertake a ‘full dress rehearsal’ with participant involvement and a duration that is representative of the duration that will occur in the FOT.
- ✓ Use pilot testing also as a means of estimating the amount of time required to complete activities, as this will enable more accurate budgeting during the remainder of the project.
- ✓ Pre-test all data analysis procedures to ensure appropriate data is collected – particularly data related to event recording triggers.
- ✓ Ensure that the routes used in pilot studies maximise the likelihood of critical situations of relevance to the FOT. Consider using a test track to verify the logging of critical situations.

<ul style="list-style-type: none"> ✓ Add independent monitoring systems to pilot platforms to ensure the validity of data derived from sensors. ✓ <i>In the pilot phase listen to the users and, when involved, owners and managers of the vehicle fleet – their ideas are likely to be different.</i>
<p>General Advice</p> <ul style="list-style-type: none"> ✓ For data collection systems, ensure that data is being recorded, determine the accuracy of data recorded, test downloading procedures and equipment, test reader software and analyse samples of pilot data.

Activity 18: Run the FOT		
Tasks and Sub-Tasks	Person/Team/Organisation Responsible for Activity	Done
18.1 Ensure that all sign-offs have occurred for previous activities	Project Manager, Project Management Team	•
18.2 Manage the FOT: <ul style="list-style-type: none"> ○ Monitor project activities, timelines, budgets and resources ○ Prepare regular progress and financial reports for sponsor ○ Convene and attend regular meetings with research and support teams ○ Maintain communication with sponsor and key stakeholders 	Project Manager, Research Team, Technical Support Team, Administrative Support Team, Project Management Team	•
18.3 Recruit participants	Project Manager, Research Team	•
18.4 Organise training session times/materials	Project Manager, Research Team	•
18.5 Brief and train participants	Project Manager, Research Team	•
18.6 Brief fleet managers (if appropriate)	Project Manager, Research Team	•
18.7 Deploy FOT platforms	Project Manager, Research Team, Technical Support Team	•
18.8 Regularly monitor participant progress, including kilometres travelled	Project Manager, Research Team	•

18.9 Administer questionnaires and implement other data collection methods at pre-determined intervals	Project Manager, Research Team	•
18.10 Collect, enter into database (unless automated) and store subjective data	Project Manager, Research Team	•
18.11 Record, download and store objective (i.e. logged) data	Project Manager, Research Team, Technical Support Team	•
18.12 Collect special data (e.g. fuel docket) needed to analyse surrogate performance indicators	Project Manager, Research Team	•
18.13 Monitor for, collect and document data on technical problems and user feedback	Project Manager, Research Team, Technical Support Team	•
18.14 Commence preliminary evaluation of data, to identify instances of dangerous driving and any other findings of interest/relevance to FOT outcomes	Project Manager, Research Team, Technical Support Team	•
18.15 Repair and re-deploy platforms (as required)	Project Manager, Technical Support Team	•
18.16 Routinely ensure all platforms are properly maintained and legal in other ways (e.g. registered, licensed, tyres properly inflated)	Project Manager, Research Team, Technical Support Team	•
18.17 Report dangerous driving behaviours (if legally required)	Project Manager, Research Team, Technical Support Team	•
18.18 Conduct exit interviews with users and the other relevant actors	Project Manager, Research Team	•
18.19 Remove systems and equipment from private vehicles (if used)	Project Manager, Technical Support Team	•
18.20 Sign off on completion of this activity of the FOT	Project Manager, Research Team, Technical Support Team, Administrative Support Team, Project Steering Committee, Project Management Team, Accounting/Auditing Advisor, Sub-Contractors	•

Critical Considerations (the “dos” and “don’ts”)

✓ Anticipate, and plan for, participant ‘dropout’ throughout the FOT – over-sample. It is rarely possible to replace participants who drop out after more than a few days without affecting the timing plan.

- ✓ *Develop protocols for responding to drivers with technical and other problems (e.g. provide drivers with a dedicated cell phone to report problems; ensure at least two people have pagers to receive problem calls; etc.) Timely responses will keep drivers happy.*
- ✓ Anticipate problems that may increase the drop-out rate (e.g. higher fuel consumption in the FOT vehicle than in the driver's own vehicle) and take steps to prevent or mitigate these problems.
- ✓ Monitor closely system usage for drivers who may be tempted to 'demonstrate' novel systems to friends and neighbours.
- ✓ Adhere to quality control mechanisms to ensure that data is being properly recorded and downloaded.
- ✓ Adhere to calibration procedures to ensure accuracy of measurements/sensors over time and help prevent data drift issues.
- ✓ Find a suitable location for training drivers where you can also assess transfer of training to the test vehicles in a safe environment
- ✓ *If the number of kilometres driven by drivers is being controlled for, conduct regular calibration checks of cumulative distance travelled.*
- ✓ Assume that it will take you 50% longer than you think to recruit participants if recruiting company drivers.
- ✓ Check logged data as soon as you receive it to verify accuracy and completeness of data and verify kilometres travelled.
- ✓ *Monitor and record critical factors that could have an impact on the measured outcomes/dependent variables (e.g. changes in police enforcement strategies, unseasonal weather conditions). If these are not controlled for in the experimental design, or accounted for in the analyses, they could confound the measured effects of the systems being tested.*
- ✓ Where company fleet vehicles are involved in the study, advise fleet managers not to "demonstrate" their vehicles, as this may compromise the aims of the study.
- ✓ *Give sponsors early warning of potential problems that could compromise the integrity of the study, or increase the budget.*
- ✓ Encourage participants to report technical problems as soon as possible.
- ✓ *Don't assume that all systems in the test vehicles are functioning as required. Develop systems to check, at appropriate times, that they are operating properly.*
- ✓ Don't assume that drivers will do what you ask them to do (e.g. fill out questionnaires; maintain vehicles). They need regular reminding and follow-up.
- ✓ Where data downloading is manual, don't forget to replace flash memory cards, or other storage devices, with new (empty) ones on a regular basis.
- ✓ Do not always assume that drivers will clock up their kilometres evenly over the trial. Contact them on a regular basis to check cumulative distance logged.
- ✓ If legally required, don't forget to report to the appropriate authorities (e.g. company fleet managers) recorded instances of dangerous driving by test drivers.
- ✓ *Don't assume that drivers will drive the vehicles without trailers, bike racks and other accessories. These may affect the operation of some FOT systems (e.g. reverse collision warning devices).*
- ✓ Minimise interference to commercial operations during FOTS, especially trucking operations. Problems that compromise commercial productivity may result in companies withdrawing trucks from the FOT.
- ✓ *Make sure fleet managers are, and remain, motivated. Their support is critical.*
- ✓ Be careful about the feedback given to drivers. They may be concerned about the possibilities of 'unintended consequences', e.g. their managers learning how and when they take rest breaks etc.

<ul style="list-style-type: none"> ✓ Participants are more likely to comply with what is asked of them if they engage with the project. Ongoing communication and even small incentives can enhance perceived engagement and improve compliance. However, the level of engagement must not compromise the outcomes of the study. ✓ Remember that long-term involvement in a research study can be onerous for a participant. At all times treat them as participants in the study process, not simply subjects of a study. ✓ Allow sufficient time for any data entry which has to be done manually (e.g. responses from pencil and paper questionnaires, focus groups). As far as is possible, manual data entry should be carried out routinely during the course of the data collection phase and not all left to the end. ✓ A system for basic inventory management is recommended for FOTs with more than a few vehicles in use. For such a system to be efficient, sensors, data acquisition system units, vehicles and all other equipment need to be included, as well as relevant supporting procedures developed.
<p>General Advice</p> <ul style="list-style-type: none"> ✓ Ongoing communication with key stakeholders is important during the FOT to ensure that the aims and objectives of the FOT are clear, that stakeholders stay committed to the project, and that the aims and objectives of the FOT are not misquoted, misrepresented or misunderstood.

Activity 19: Analyse FOT data		
Tasks and Sub-Tasks	Person/Team/Organisation Responsible for Activity	Done
19.1 Develop a data analysis plan	Project Manager, Research Team, Technical Support Team	•
19.2 Analyse the objective (i.e. logged and recorded data)	Project Manager, Research Team	•
19.3 Analyse subjective data (i.e. data obtained from interviews, questionnaires, focus groups, hotlines, etc.)	Project Manager, Research Team	•
19.4 Draw conclusions with respect to the hypotheses generated for the FOT	Project Manager, Research Team	•
19.5 Sign off on completion of all required analyses	Project Manager, Research Team, Project Management Team, Project Sponsor(s)	•

<p>Critical Considerations (the “dos” and “don’ts”)</p> <ul style="list-style-type: none"> ✓ Plan for the fact that there will be constant demand for study findings, such as general trends in the data, early in the project, even though the data may not be statistically reliable enough to report with any confidence.
--

<ul style="list-style-type: none"> ✓ In a well-powered study, null findings (i.e. where no effect is found and the hypotheses refuted) are potentially as interesting as when the hypotheses are supported. ✓ <i>Anticipate the requirement to have to perform supplementary analyses for the funding organisation, which may be expensive and not originally budgeted for. This will require negotiation with the sponsor if these analyses are expected to be carried out within the original budget.</i> ✓ Anticipate that, unless distance travelled is controlled for in the FOT, the distance travelled by different drivers will vary significantly. Take this into account in the analysis to ensure results are not skewed. ✓ Don't forget to run "reality checks" on the data, to be sure that the data are "clean". This is essential. ✓ <i>If data is reduced/aggregated, always keep a copy of un-aggregated data.</i> ✓ Ensure that all data analysts have used the test vehicles and understand the circumstances in which data was/is collected. ✓ All team members who handle participant data should receive appropriate training regarding data privacy. ✓ Work out how best to filter logged data and deal with missing data.
<p>General Advice</p> <ul style="list-style-type: none"> ✓ There may be a requirement to conduct ongoing analysis, such as ongoing identification of dangerous drivers, determining whether adaptation to systems is occurring early enough to warrant a shorter FOT duration (e.g. to save money and time), and to identify early trends in the data. These checks should be built into the analysis plan at the start of the project. ✓ Some FOTs have developed novel ways of turning ADAS technologies on and off to control precisely the amount of exposure to the technologies that are being evaluated. ✓ Sponsors need to be aware of the relative costs of running FOTs. For example, the cost of running simulation models at the end of the FOT to estimate safety and other benefits of ICT technologies is a fraction of the cost of preparing and deploying the FOT vehicles.

Activity 20: Write minutes and reports		
Tasks and Sub-Tasks	Person/Team/Organisation Responsible for Activity	Done
20.1 Write minutes of regular project management team meetings	Project Manager	•
20.2 Write regular minutes of Project Steering Committee meetings	Project Manager	•
20.3 Write quarterly progress reports for the sponsor(s)	Project Manager	•
20.4 Write the draft FOT report	Project Manager, Research Team, Technical Support Team	•

20.5 Send the draft FOT report to relevant stakeholders and peers for peer review	Project Manager	•
20.6 Convene 1 or 2 meetings to discuss feedback with sponsor/peers	Project Manager	•
20.7 Incorporate feedback and write the final report	Project Manager, Research Team, Technical Support Team	•
20.8 Deliver the final report to sponsor(s)	Project Manager	•
20.9 Sign off on completion of all required reports	Project Manager, Research Team, Technical Support Team, Project Management Team, Project Sponsor(s)	•

<p>Critical Considerations (the “dos” and “don’ts”)</p> <ul style="list-style-type: none"> ✓ <i>Use regular progress reports to document problems, solutions and lessons learned.</i> ✓ <i>Allow sufficient time for sponsor review of draft and final reports, but not so long that the review process drags out unduly. Six to 8 weeks is recommended.</i> ✓ <i>Consider peer review of major outputs; this will improve their quality but delay their release.</i> ✓ <i>Document all lessons learnt in the final FOT report.</i> ✓ <i>Ensure that the final report contains practical recommendations for wider-scale deployment of those systems found to be effective, and for fine-tuning of those with potential to be more effective.</i> ✓ <i>Develop, in consultation with the Project Steering Committee, a suggested plan for implementing the recommendations deriving from the FOT. Document the implementation plan in the FOT final report.</i> 		
<p>General Advice</p> <ul style="list-style-type: none"> ✓ <i>The FOT lifecycle is long. Hence, it is advisable to write separate reports on each critical stage of the FOT, particularly the lessons learned, to ensure that nothing important that should be documented is forgotten.</i> ✓ <i>Formal meeting minutes are a critical resource for the project in confirming departures from the project plan.</i> 		

Activity 21: Disseminate the FOT findings		
Tasks and Sub-Tasks	Person/Team/Organisation Responsible for Activity	Done
21.1 Send regular project reports to the sponsor	Project Manager	•

<p>21.2 Disseminate preliminary and final findings at seminars, conferences and special events</p>	<p>Project Manager, Research Team, Technical Support Team</p>	<p>•</p>
<p>21.3 Prepare reports on preliminary findings for the sponsor</p>	<p>Project Manager, Research Team, Technical Support Team</p>	<p>•</p>
<p>21.4 Send sponsor draft and final FOT reports</p>	<p>Project Manager</p>	<p>•</p>
<p>21.5 Provide other stakeholders with access to the FOT final report(s) and, if allowed, raw or filtered data from the FOT</p>	<p>Project Manager, Research Team, Technical Support Team</p>	<p>•</p>
<p>21.6 Showcase the vehicles at relevant events during the FOT (e.g. Smart Demos, motor shows) to promote awareness and wider deployment of systems</p>	<p>Project Manager, Technical Support Team, Project Steering Committee, Project Management Team, Project Sponsor(s)</p>	<p>•</p>

Critical Considerations (the “dos” and “don’ts”)

- ✓ Disseminate the findings in accordance with the previously agreed communications plan.
- ✓ *Agree on what can and cannot be disseminated and said at different points in the study.*
- ✓ Seek necessary permissions prior to divulging FOT findings to any third party.
- ✓ FOT reports are large and expensive to print. Allocate a sufficient budget at the beginning of the project for printing.
- ✓ FOT reports are large and hard to read. It is desirable to produce conference papers along the way that document the outputs of the study at different phases. Prepare a concise 1 or 2 page synopsis of the study outcomes that can be read and easily digested by politicians, chief executives and relevant others in positions of authority.
- ✓ Agree in advance who is empowered to release and comment on results.

General Advice

- ✓ Where private industry is a participant in the FOT, it may be necessary to seek permission from the manufacturer before divulging certain information deriving from the FOT. This must be established.
- ✓ Maintain at least one vehicle for demonstrations; preferably at a location that is convenient to politicians, officials and the press.
- ✓ A demonstration and briefing to an influential politician is likely to be far more effective than sending them a report.

Activity 22: Decommission the FOT		
Tasks and Sub-Tasks	Person/Team/Organisation Responsible for Activity	Done
22.1 Conduct de-briefing interviews with participants to elicit feedback on the FOT that can be used to improve future FOTs	Project Manager, Research Team	•
22.2 Dispose of test vehicles which are no longer needed (if vehicles are not privately owned)	Project Manager, Research Team, Technical Support Team, Administrative Support Team, Project Management Team, Accounting/Auditing Advisor, Project Sponsor(s)	•
22.3 Retrieve installed data logging equipment (if vehicles are privately owned)	Project Manager, Research Team, Technical Support Team, Administrative Support Team, Project Management Team, Accounting/Auditing Advisor, Project Sponsor(s)	•

Critical Considerations (the “dos” and “don’ts”)

- ✓ Ensure that participants return relevant items at the end of the study (e.g. flash memory cards, i-buttons) and perform other required activities to decommission the FOT vehicles (e.g. disconnect power to support systems).
- ✓ *Keep one vehicle until all data analyses are complete.*
- ✓ Consider providing public access to FOT databases, where ethically allowed, that enables others to use the data for other research purposes after the FOT has been de-commissioned (but remember to fully explore and address anonymity issues). The data collected and stored after the FOT is de-commissioned should be regarded as “living data”.
- ✓ *Don’t lose momentum at the end of the FOT. Lobby stakeholders to ensure that there is commitment to implementing the recommendations of the FOT.*

General Advice

- ✓ *Consider keeping one or two vehicles as showcasing vehicles after the study, to allow stakeholders in positions of authority to experience the look and feel of the vehicles.*
- ✓ It may be necessary to consider legal issues of decommissioning the FOT as far as the de-installation of data logging equipment is concerned (in a contract with participants).
- ✓ In some FOTs, it may be possible for the test users to continue using the systems as part of a commercial service being offered near the end of the test period. It is important to consider the contractual arrangements at the end of the project and handover to commercial activities.

Annex B Informed Consent Checklist

This checklist is intended to support user test organizers when preparing legal agreements, together with lawyers, that specify conditions that the participants must agree to, in order to join the tests. In many cases the format used to obtain informed consent includes providing the prospective participant an information sheet giving details about the study, how their data will be treated, and their rights to withdraw consent or request that their personal data not be used. The information sheet is accompanied by a consent form in which the participant acknowledges they have read and understood the information sheet, and agrees to participate in the study. In other cases, particularly where the FOT involves long term testing on public roads, a more comprehensive form of agreement will be required informed by legal expertise from the jurisdiction in which the FOT will take place.

Informed consent is about ensuring that the test participants understand the details of the legal agreement as well as practicalities of the test itself. The validity of the agreement is high when both sides have truly read and understood it. The information sheet/consent or agreement should be drafted clearly, in non-technical language. It is important that enough time must be given (e.g. agreement text made available beforehand, if it is lengthy) and preferably, participants should be explained the key points in an introductory session. To ensure that the test participants pick up key points from the agreement, it is recommended to use checkboxes, following key clauses. Please see the [Data Sharing Framework](#) (FOT-Net Data and CARTRE, 2019), Chapter 4.3, for examples on requesting e.g. use of short video clips from tests for public project presentations.

Information sheet and participant agreement

The participant information sheet and participant agreement must cover:

- Main aspects of the study and what the data will be used for (purpose)
- Funding sources
- Whether personal data will be used, specifying what data and for what purpose
- If the data collection includes video (which is personal data), this should be specified
- Data sharing and re-use topics/areas in high detail – preferably still allowing future work and collaboration with third parties on analysis
- Participant's responsibility for the adherence to related laws and traffic rules, obligation in case of administrative fines
- Outline potential risks associated with participation
- Any potential benefits of participation
- Parameters of insurance coverage provided within the FOT
- If the participant suffers from health impairments which may create additional risk during test activities, options may include:
 - Where the participant does not disclose the health condition, a recommendation it is not recommended they participate
 - Where the participant discloses the health condition, an additional risk assessment may be required.

- That the participant holds valid licenses (considering obligations to inform Test Management if the license, e.g. a driving license, is revoked)
- Explicitly mention in the form and ask for approval if there are plans to collect and process health data, even e.g. motion sickness
- Instructions on safe system use (point out participant's responsibilities for safe use and interaction with the system, warn against overreliance)
- Instructions covering breakdown and accident (e.g. who should be informed / asked for assistance)
- Permissibility of system use by others, e.g. family members
- Clarify rewards and incentives for participating in the study
- Approximate amount of time required of the participant
- Explain the participant's right to withdraw from the study and how this may be done to
- Contact details for the study.

The participant should be given a copy of his/her signed consent form. Keeping a separate and confidential document is recommended, where the agreement is recorded by an alias identifier linked with a test user ID (or another pseudonym) that is visible in test data.

Consent Form

The consent form should include

- A statement that the participant has either read or had the opportunity to discuss the contents of the information sheet and has understood its contents
- A statement indicating a willingness to participate in the study
- Confirmation of the right to withdraw
- Contact details for the study.

Annex C Checklist for Participant Introduction

As a best practice, FESTA recommends organizing an introductory session for persons participating in user tests. This is especially relevant when the participants will be, for example, driving on public roads with a vehicle that has prototype features – or generally, when participants would be operating a system that raises safety concerns.

The main purpose of the introductory session is to keep testing safe: the participants should understand the basic behaviour of tested systems, so that no accidents would be caused because of basic misunderstandings or surprises. Further, the session should be used to explain key parts of the informed consent forms (legal agreements for entering the tests), ensuring that the participation and test conditions are well understood. The test participants must have enough time to study the consent form before being asked to sign it.

The introductory session should cover:

- Background of the study, the research project
- Information regarding the tested system, minimally safety-related aspects
- Information regarding the setup of the test
- When testing is arranged as a single session, an overview of the timeline: introduction, questionnaires, use of the system and e.g. interviews afterwards. Could participants be contacted again and for what purpose
- If the experiment last for a longer period: how/where participants get the introduction and tested systems, how long the data collection lasts, what testing periods such as baseline data collection may be involved, how often participants will be e-mailed about questionnaires or for other purposes
- Participant's tasks and responsibilities
- Organiser's tasks and responsibilities
- Insurances and liabilities
- Hotline and contacts in case of trouble, behaviour in an emergency
- How to withdraw from the study
- How to return test equipment
- How research data will be managed according to good scientific principles and that they have some options to review and ask for deletion of collected data
- How and when the results of the study will be published. For example, results would be based on group averages calculated from data, or that if video showing a test participant would be published, a separate permission for that would be asked in the consent forms.