C-MOBILE

Accelerating C-ITS Mobility Innovation and deployment in Europe

D6.1: Validation and impact assessment methodology

Status	Final
Main authors (alphabetical order)	Hüseyin Avsar (DLR) Michael Böhm (DLR) Yanja Dajsuren (TU/e) Raul Ferrandez (Applus+ IDIADA) Manuel Fünfrocken (HTW) Florian Koller (DLR) Areti Kotsi (CERTH) Kerry Malone (TNO) Tudor Mocanu (DLR) Katharina Preuk (DLR) Serguei Roubtsov (TU/e)
Work Package	WP6 – Validation and impact assessment
Task	T6.1 – Validation and impact assessment methodology
Dissemination level	Public
Issue date	30/06/2018

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 723311

Project coordination

Mr. Marcos Pillado Quintas
Applus+ IDIADA
PO Box 20 Santa Oliva
43710, L'Albornar, Tarragona
Spain

Telephone: +34 977 18 93 60
E-mail: marcos.pillado@idiada.com

Revision and history sheet

Version	Date	Authors	Summary of changes	Status
0.1	28/07/2017	Raúl Ferrandez (Applus+ IDIADA)	Initial ToC	Prepared
0.2	03/08/2017	Raúl Ferrandez (Applus+ IDIADA)	Added content for section 1. Introduction	Prepared
0.3	10/08/2017	Raúl Ferrandez (Applus+ IDIADA)	Added content for field operational test concept and scope	Prepared
0.4	17/08/2017	Raúl Ferrandez (Applus+ IDIADA)	Added initial function descriptions	Prepared
0.5	08/09/2017	Katharina Preuk (DLR)	Added initial function descriptions and reviewed ToC	Prepared
0.6	19/09/2017	Raúl Ferrandez (Applus+ IDIADA)	Added Ethical and legal issues	Prepared
0.7	26/09/2017	Raúl Ferrandez (Applus+ IDIADA)	Finalization of the evaluation criteria, performance indicators and measures	Prepared
0.8	3/10/2017	Raúl Ferrandez (Applus+ IDIADA)	Modifications from comments provided by CERTH	Prepared
0.9	9/10/2017	Raúl Ferrandez (Applus+ IDIADA)	Added information on subjective/objective evaluation per pilot site	Prepared
0.10	30/10/2017	Kerry Malone (TNO)	Added content for EWD	Prepared
0.11	31/10/2017	Areti Kotsi (CERTH)	Added content for section 4.9.1 Ex-post cost-benefit analysis	Prepared
0.12	1/11/2017	Katharina Preuk/ Tudor Mocanu/ Florian Koller (DLR)	Added content for section 4.9.1 and 4.9.4	Prepared
0.13	6/11/2017	Moisés Rial Martínez (CTAG)	Added content for section 4.8	Prepared
0.14	14/11/2017	Manuel Fünfrocken (HTW)	Added content for section 2	Prepared
0.15	15/11/2017	Serguei Roubtsov/ Yanja Dajsuren (TU/e)	Added content for section 2	Prepared
1.0	20/11/2017	Katharina Preuk (DLR)	Added final comments	Prepared
1.1	30/11/2017	Marcos Pillado (Applus+ IDIADA)	Review	Approved
2.0	25/05/2018	Areti Kotsi (CERTH)	Updated content for section 4.10 Ex-post Cost-Benefit Assessment	Prepared
2.1	20/06/2018	Michael Böhm/ Hüseyin Avsar/ Tudor Mocanu/ Florian Koller (DLR)	Updated Content for section 4.9.3 and 4.9.4	Prepared
2.2	25/06/2018	Raúl Ferrandez (Applus+ IDIADA)/ Florian Koller (DLR)	Reviewed	Prepared
2.2	26/06/2018	Areti Kotsi (CERTH)	Reviewed	Prepared
2.2	27/06/2018	Michael Böhm (DLR)	Reviewed	Prepared
2.2	13/07/2018	Graeme Hill (UNEW), Charlotte Fléchon (PTV), Kerry Malone (TNO), Patrick Hofman (MapTM), Moisés Rial Martínez (CTAG)	Reviewed	Prepared
2.2	16/07/2018	Marcos Pillado (Applus+ IDIADA)	Review	Approved
2.3	10/07/2019	Florian Koller (DLR)	References corrected	Prepared
2.3	30/07/2019	Alex Vallejo (IDIADA)	Review	Approved

Table of Contents

Figures

Tables

Abbreviations

Abbreviation	Definition
3G	3rd generation of mobile telecommunications technology
ACC	Adaptive Cruise Control
AIDE	Adaptive Integrated driver-vehicle interface
CAM	Cooperative Awareness Message
CC	Communication Consortium
CM	Critical Mass
C2C	Car to Car
CACC	Cooperative Adaptive Cruise Control
CBA	Cost-Benefit Analysis
CCC	Conventional Cruise Control
CVRIA	Connected Vehicle Reference Implementation Architecture
D	Deliverable
DB	Database
DEA	Data Envelopment Analysis
EBA	Event Based Analysis
EEIS	Energy Efficient Intersection Service
ETSI	European Telecommunications Standards Institute
EV	Electric Vehicle
EWD	Evaluation While Doing
FCD	Floating Car Data
FESTA	European project, developed methodology for conducting FOTs
FG	Focus Group
FOT	Field Operational Test
GLOSA	Green Optimal Speed Advisory
GPS	Global Positioning System
HGV	Heavy Goods Vehicle
HMI	Human Machine Interface
ICT	Information & Communication Technology
IRR	Investment Return Rate
ITS	Intelligent Transport Systems
KPI	Key Performance Indicator
M	Month
MT	Management Team
OM	Optimal Mass
OBU	On Board Unit
PI	Performance Indicator
PS	Pilot site
RHW	Road Hazard Warning
RLVW	Red Light Violation Warning
RSU	Road Side Unit
RQ	Research Questions
RWW	Road Works Warning
SUS	System Usability Scale
SVW	Signal Violation Warning
TAM	Technology Acceptance Model
TJAW	Traffic Jam Ahead Warning
TMC	Traffic Management Centre
TRA	Theory of Reasoned Action
TTC	Time To Collision
VRU	Vulnerable Road User
WP	Work Package

Executive Summary

In the past years, there has been tremendous progress in the field of intelligent transport systems; several successful cooperative mobility initiatives have proven potential benefits of cooperative systems in increasing both energy efficiency and safety for specific transport modes. However, the large variety of cooperative applications have been designed for different goals, stakeholders or specific settings / environments and have been developed on a silo-based approach and deployed independently from each other, serving however, at higher level, similar goals and functionalities for the end-user. Scalability, IT-security, decentralization and operator openness are some of the most important properties that a technically and commercially successful solution must provide.

C-MobILE aims to stimulate / push existing and new pilot sites towards large-scale, real-life C-ITS deployments that are interoperable across Europe. Well-defined operational procedures will lead to decentralized and dynamic coupling of systems, services and stakeholders across national and organizational borders in an open, but secure C-ITS ecosystem, based on different access technologies, the usage of which is transparent for service providers and seamless and continuous for the end-users across different transport modes, environments and countries.

The main scope of this document is to present the evaluation framework that is followed in the C-MobILE project. The evaluation framework takes into account the experimental procedure, evaluation criteria and performance targets, defined at project level, in order to provide the evaluation requirements for the implementation of the C-ITS services. The assessment focuses on the performance of the systems at service and at bundle level.

In C-MobILE two different architecture approaches are followed. To assess the quality of the implementation of the non-functional requirements, as well as of the more general criteria, such as (perceived) usability, (perceived) performance, or stability, the Architecture Trade-off Analysis Method (ATAM) and FESTA are used. To assess the quality of the architecture description with respect to the functional requirements, a combination of the Cost Utility Analysis (CUA) and the Scenario-based Architecture Analysis Method (SAAM) are used.

The results of both methods are integrated into a common *validation tree*, which is based on CUA. This tree not only shows the validation results of the different approaches, but also gives an aggregated assessment result, based on the weighted results of the different approaches.

Key Performance Indicators (KPIs) have been defined and will be used for the evaluation of the Field Operational Tests (FOTs) and their categorization. The evaluation criteria have been defined taking into account the user requirements defined in Task 2.1, since the different services require different data to be assessed.

Research questions and hypotheses, serving the evaluation purposes, are aligned to impact areas and technical categories. Impact areas comprised of; personal, environment, efficiency, safety and user experience, while technical categories comprised of; performance, security, resilience, replicability, sustainability, interoperability, availability and reliability. Each research question and hypothesis will be evaluated using KPIs.

The data collected can come in different forms, i.e. qualitative and quantitative. Quantitative data will be collected by sensors. This data will be collected at deployment site level and it will be merged regularly with a central database. Qualitative data will be gathered by other means, such as questionnaires.

Finally, this document describes the assessment of the added value and economic viability of C-ITS services bundling for key stakeholders and end-users by means of a comprehensive cost-benefit analysis (CBA).

This deliverable will be complemented by D6.2 “Technical validation report”, which will update in some parts its contents based on the final implementations and data availability.

1. Introduction

1.1. C-MobILE at a glance

The C-MobILE (Accelerating C-ITS Mobility Innovation and depLoyment in Europe) vision is a fully safe & efficient road transport without casualties and serious injuries on European roads, in particular in complex urban areas and for Vulnerable Road Users (VRU). We envision a congestion-free, sustainable and economically viable mobility, minimizing the environmental impact of road transport. C-MobILE will set the basis for large scale deployment in Europe, elevating research pilot sites to deployment locations of sustainable services that are supported by local authorities, using a common approach that ensures interoperability and seamless availability of services towards acceptable end user cost and positive business case for parties in the supply chain.

1.2. Objective

C-MobILE aims to stimulate / push existing and new pilot sites towards large-scale, real-life C-ITS deployments interoperable across Europe. Well-defined operational procedures will lead to decentralized and dynamic coupling of systems, services and stakeholders across national and organizational borders in an open, but secure C-ITS ecosystem, based on different access technologies, the usage of which is transparent for service providers and seamless and continuous for the end-users across different transport modes, environments and countries.

The main scope of this document is to present the validation and impact assessment methodology that is followed in the C-MobILE project. The framework follows the guidelines of the FESTA methodology for FOTs in Europe and draws also upon methodologies of similar projects, such as DRIVE C2X, FREILOT, Compass4D, CO-GISTICS, C-ROADS, and C-The-Difference.

1.3. Intended audience

The audience of this deliverable are deployment site leaders, service providers and public authorities involved in the implementation of the C-MobILE services in each of the deployment sites. Besides, the deliverable is written for all stakeholders interested in learning about the validation and assessment of C-ITS services.

1.4. Expected impacts

C-MobILE will demonstrate the way for C-ITS deployment in Europe, serving as the reference for all stakeholders and new cities/ regions, interested in investing in C-ITS. Furthermore, C-MobILE results will: significantly reduce fragmentation of C-ITS developments across the EU, accelerate deployment and market uptake of C-ITS, make C-ITS benefits concrete and interpretable by relevant decision makers and investors, and stimulate the competitiveness of related EU industries.

- Expected Impact 1: Improved level of performance for the entire surface transport system, including more efficient and sustainable traffic management, improved safety and contribution to overall socio-economic development.
- Expected Impact 2: Testing and demonstration of fully integrated C-ITS concepts in real-life, complex environments.
- Expected Impact 3: Greater collaboration (and partnerships) between multiple stakeholders to deploy applications and facilitate the interoperable interactions across all elements of the road transport system, including the use of data from multiple sources.
- Expected Impact 4: A comprehensive Cost-Benefit Analysis (CBA) demonstrating the value added and economic viability of C-ITS services and solutions for users and other stakeholders.
- Expected Impact 5: Validated results and proven impact on user acceptance, safety, resilience and security with respect to transport demand and the environment.
- Expected Impact 6: Development of validated guidelines for the large-scale deployment of operational and sustainable C-ITS services in Europe.

WP6 (validation and impact assessment) addresses the C-MobILE objectives 3, 7 and 8 listed below. Their successful achievement will be measured by different means such as the publication of corresponding deliverables, end-user acceptance at deployment sites, successful use of produced guidelines/ operational procedures, Key Performance Indicators (KPIs) and public events.

- Objective 3: Assess the cumulative, real-life benefits of bundling C-ITS applications and integrating multiple transport modes in the C-ITS ecosystem.
  - Evaluate the cumulative, real-life benefits (including Cost-Benefit Analysis) of C-ITS cross-modal applications in terms of end-user acceptance, personal mobility as well as the impact of C-ITS bundles on the surface transport system in terms of efficient traffic management, safety and environment. Extend C-ITS info-structure to allow end-users easy access to new applications deployed in C-MobILE. Present business models, mapping costs and benefits to all C-ITS parties based on the CONVERGE institutional role-model to assess the economic viability of C-ITS services and solutions to support deployment strategies.
- Objective 7: Release testing methodologies to evaluate C-ITS architectures and the effectiveness of applications.
  - Establish harmonised testing methodologies to validate C-ITS architectures and applications in order to guarantee pan-European interoperability of C-ITS architectures. The testing methodology will be based on FESTA and adapted to assess overall performance, security, resilience, replicability and sustainability. Eventually, it will be provided to the EC (e.g. C-ITS Platform), UNECE (e.g. GRRF WP.29 ITS/AD), main standardisation bodies (e.g. ETSI, C2C-CC) and assessment platforms (e.g. Euro NCAP).
- Objective 8: Demonstrate the added value and economic viability by means of a comprehensive Cost-Benefit Analysis and impact assessment.
  - Define an impact assessment methodology via the automation of the testing cycle for evaluation while testing that will quickly provide answers and early adaptations of C-ITS services. An ex-ante cost CBA will analyse the impact of each application and the bundling of applications. The appraisal results obtained will be compared with an ex-post CBA. The CBA findings will be used for the development of business plans for the large-scale deployment.

1.5. Document structure

This deliverable presents the C-MobILE overall validation and impact assessment methodology. In order to facilitate the use of this deliverable by project partners, as well as to ensure that its content is taken into account by the related project activities, the concrete sections and content, to be taken into account by each activity and deployment site leader, are listed below:

- Section 1 “Introduction” provides the rationale and context of this project.
- Section 2 “C-ITS Architecture Validation Methodology” presents the methodology for the validation of the C-MobILE architecture. The assessment is divided into the validation of functional and non-functional requirements. Requirements highlighted in WP2 are mostly functional, therefore the assessment of non-functional requirements is difficult to conduct in detail at this stage of the project. The functional assessment will be performed using CUA (Cost Utility Analysis) and SAAM (Scenario-based Architecture Analysis), while the non-functional assessment will be performed using ATAM (Architecture Trade-Off Analysis) and FESTA.
- Section 3 “Field Operational Test concept and scope” provides details on the requirements set for FOTs and defines the basic concepts relevant to the project work.
- Section 4 “C-MobILE FOT methodology” describes the FOT methodology. It is noteworthy that all the details needed in order to make the description at hand final have not yet been in place during the writing phase. This applies, for example, to the precise definition of testing procedures and affects the analysis methods. The deliveries from WP2 determine to a large extent the details of the services. However, the framework of the methodology can be presented already at this stage. Thus, some aspects concerning the methodology are still subject to revision and change depending on the results of WP2 and others.

There are still a number of open issues/ details, concerning various methods, necessary for the set-up of a complete testing system for numerous functions listed later in this document. This applies to precise test design and methods needed to be tailored for different functions. It is obvious that the same testing procedures cannot apply to all functions. Furthermore, only after the needs and requirements for implementation have been carried out, can the design of testing procedures be completed and the methodology finalised.

2. C-ITS Architecture Technical Validation Methodology

In C-MobILE two different architecture evaluation approaches are followed.
To assess the quality of the implementation of the non-functional requirements, as well as the more general criteria, such as (perceived) usability, (perceived) performance, or stability, the Architecture Trade-off Analysis Method (ATAM) and FESTA are used.

To assess the quality of the architecture description with respect to the functional requirements, a combination of the Cost Utility Analysis (CUA) and the Scenario-based Architecture Analysis Method (SAAM) are used.

The results of both methods are integrated into a common *validation tree*, which is based on CUA. This tree not only shows the validation results of the different approaches, but also gives an aggregated assessment result, based on the weighted results of the different approaches, as shown in Figure 1: Overall assessment structure.

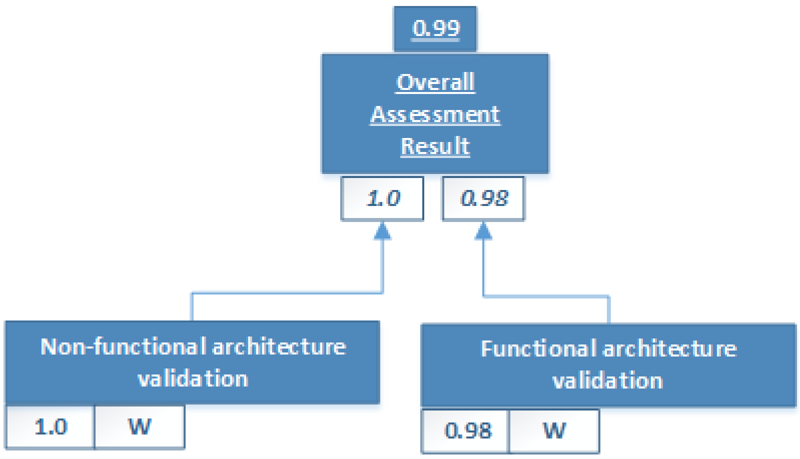

Figure 1: Overall assessment structure

2.1. C-ITS Functional Architecture Validation Methodology

Similar to any large, software heavy system, the behaviour of the C-MobILE system is defined by use-cases and functional requirements. Functional requirements are requirements, which define functions of the systems, e.g. calculations, data manipulation, technical details, and others. It is apparent, that compliance to the requirements is an indicator of the overall quality of the system. However, it is simultaneously hard to compare architectures based on requirements fulfilment, as the exact way in which requirements are implemented differs and is hard to quantify. Furthermore, in the context of C-ITS, there are no architectures implementing the same set of requirements, thus making comparison even harder. To account for this fact, C-MobILE is not trying to assess its architecture against other architectures, but to perform an assessment with respect to general software quality criteria and requirement fulfilment, as described below.

The functional architecture validation process combines the results of two sub-processes, the *functional requirement fulfilment process,* and the *scenario-based architecture assessment process*, into a single result, the *functional assessment result* as indicated by Figure 2: Functional assessment process. The functional requirement fulfilment process is based on CUA, whereas the scenario-based architecture assessment process is based on SAAM.

Both processes are described in sections below.

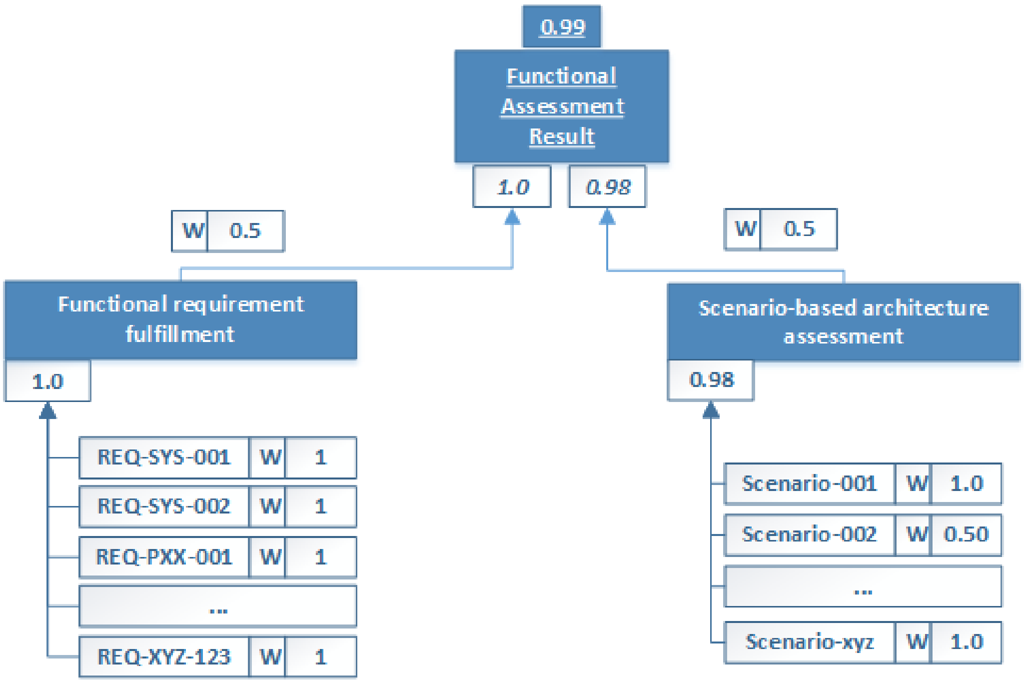

Figure 2: Functional assessment process

With those two sub-processes, the overall assessment can be calculated as the weighted sum of overall requirement fulfilment and the overall scenario fulfilment. This is formally defined as:

$$F_A = W_{A_R} * F_R + W_{A_S} * F_S \quad (1)$$

With F_A the functional assessment, W_{A_R} the weight of the requirements. F_R the degree of fulfilment of the requirements, W_{A_S} the weight of the scenario assessment, and F_S is the combined degree of fulfilment of the scenario branch. With

$$0.0 \leq F_R \leq 1.0, F_R \in \mathbb{Q} \quad (2)$$

$$0.0 \leq F_S \leq 1.0, F_S \in \mathbb{Q} \quad (3)$$

Whereas the sum of the weights is equal to one:

$$W_{A_R} + W_{S_R} = 1, with\ W_{A_R}, W_{S_R} \in \mathbb{Q} \quad (4)$$

Taken alone, those values are hardly useful without further explanation. However, they may be used by future initiatives, aiming to compare their architectures against the C-MobILE architecture. The detailed calculation of the various degrees of fulfilment is described in the following sub sections.

2.1.1. Requirement Fulfilment Process

Requirements for the C-MobILE architecture have been obtained in work package 2 and documented in D2.3. From this process, a list of requirements was developed, which shall be implemented by the architecture. These requirements are individually accessed by checking against the architecture description, created in work package 3, i.e. if the requirement is fulfilled by the architecture.

For each requirement, the following information produces an individual requirement fulfilment assessment:

Req.-No.:		Req.-Name:	
Req.-Description:			
Assessment-Rational:			
Degree of fulfillment (DoF):		Requirement Weight (RW)	

Table 1: Exemplary requirement assessment table

Table 1: Exemplary requirement assessment table shows, how the information could be structured in an assessment template, which is then completed by an assessor. The information therein is the following:

Field	Type	Description
Req.-No	String (letters and figures)	The Requirement-Number. This is a string, uniquely identifying each requirement. It has been assigned by WP 2 to each requirement and is used to identify requirement through the project.
Req.-Name	String	The Requirement Name. A comparatively short name of the requirement.
Req.-Description	Text	The Requirement description. A free text, describing the requirement. This has been defined in WP2 and is given here as a reference.
Assessment Rational	Text	This is free text, describing why the assessor has decided to rate this requirement with the DoF below. It shall also give a clear indication to the parts of the architecture description, which are relevant for this requirement.
Degree of fulfilment (DoF)	Floating number	The fulfilment degree indicates the degree to which this requirement has been fulfilled by the architecture, with 0.0 = not fulfilled and 1.0 = fulfilled.
Requirement Weight (RW)	Floating number	This is the relative weight of the requirement in relation to all other requirements. The cumulated weight of all requirements equals 1.0.

Table 2: Requirement assessment fields

Once the assessment of each requirement will be done, the assessment result will be integrated into a single *overall degree of fulfilment*, as shown in Figure 3: Requirement assessment structure.

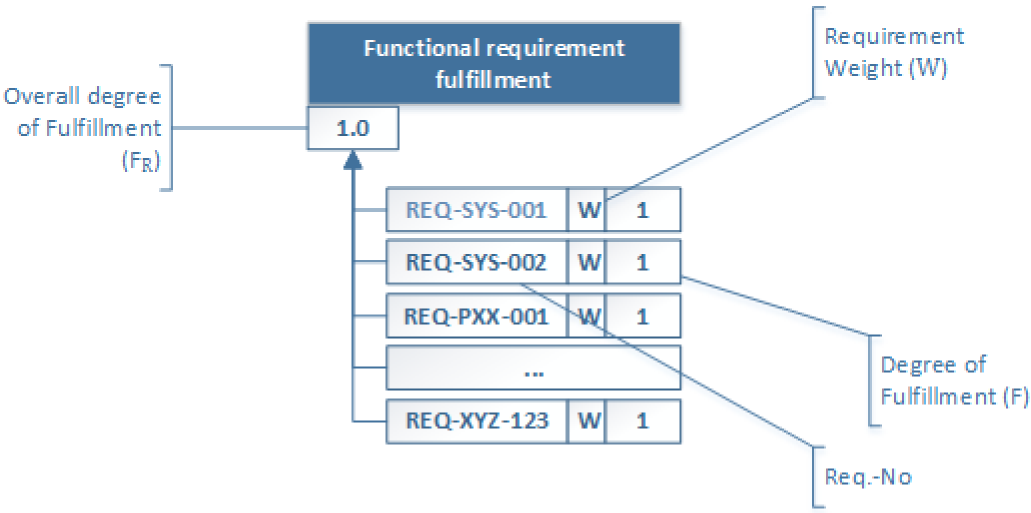

Figure 3: Requirement assessment structure

The whole process can be more formally described as shown below.

The overall degree of fulfilment for the requirements F_R can be calculated as

$$F_R = \sum_{k=0}^{n} W_{R_k} * F_{R_k} \quad (5)$$

Where W_{R_k} is the weight of requirement k and F_{R_k} is the degree of fulfillment of requirement k.

F_R will be between zero and one, or more formally,

$$0.0 \leq F_R \leq 1.0, F_R \in \mathbb{Q} \quad (6)$$

Additionally, the sum of all weights needs to equal one:

$$\sum_{k=0}^{n} W_{R_k} = W_{R_G} = 1, 0.0 \leq W_{R_k} \leq 1.0, W_{R_k} \in \mathbb{Q} \quad (7)$$

In addition, the fulfilment itself is measured as a binary value:

$$F_{R_k} \in \{0,1\}, F_{R_k} \in \mathbb{N}, 0 = not\ fulfilled, 1 = fulfilled \quad (8)$$

2.1.2. Scenario-Based Assessment

As the requirement assessment process only allows evaluating the architecture with respect to the requirements, and not with respect to the more general *software architecture quality criteria*, a second evaluation step has been introduced. This scenario-based assessment evaluates the C-MobILE architecture in the context of several hypothetical scenarios. These scenarios are distinct from the use-cases considered by work package 2, in order to identify the requirements. However, these scenarios are derived from the C-ITS context. They constitute possible future uses of the architecture. Each of the individual scenario assessments could be interpreted as a question, asking, "Can the C-MobILE architecture be used in this context?"

Each scenario is assessed by an expert evaluating the scenario description. The expert will then give a rational, describing how the scenario could be implemented. If the scenario cannot be implemented, the rational shall describe, what is missing. In both cases, the rationale will refer to the relevant sections of the architecture description. An exemplary template for the scenario-based assessment is given in Table 3: Exemplary scenario assessment template below.

Scenario-ID:		Scenario-Name:	
Scenario-Description:			

Assessment-Rational:			
Degree of fulfillment (DoF):		Relative Weight (W)	

Table 3: Exemplary scenario assessment template

The fields of this assessment are described in Table 4: Scenario assessment fields:

Field	Type	Description
Scenario-ID	String (letters and figures)	A string, uniquely identifying each scenario.
Scenario Name	String	A comparatively short name of the scenario.
Scenario Description	Text	A free text, describing the scenario. This is the description of the scenario, on which the assessment is based.
Assessment Rational	Text	This is free text, describing why the assessor has decided to rate this scenario with the DoF below. It shall also give a clear indication to the parts of the architecture description, which are relevant for this assessment.
Degree of fulfilment (DoF)	Floating number	The fulfilment degree indicates the degree to which this scenario has been fulfilled by the architecture. See below for details on this.
Relative Weight (RW)	Floating number	This is the relative weight of the scenario in relation to all other scenarios. The cumulated weight of all scenarios equals 1.0.

Table 4: Scenario assessment fields

In contrast to the requirement assessment described above, the scenario-based assessment uses fixed intervals for the *degree of fulfilment* (DoF), described in Table 5 below:

Level	Description	Value
Feasible	The scenario is already fully supported without any adjustment of the architecture.	1.0
Conditionally feasible	The scenario is conditionally feasible, if an interface and its protocol have to be assimilated.	0.75
Adaptably feasible	The scenario is adaptably feasible, if a component needs to be extended with an additional interface to itself or another component.	0.5
Hardly feasible	The scenario is hardly feasible, if a new component is needed to realize it.	0.25
Not feasible	The scenario is not feasible at all without heavy changes like an insertion of a new layer.	0.0

Table 5: Scenario DoF-Levels

Once each scenario has been assessed individually, the various DoF are combined to an *Overall Degree of Fulfilment*, as shown in Figure 4: Scenario-based assessment structure below:

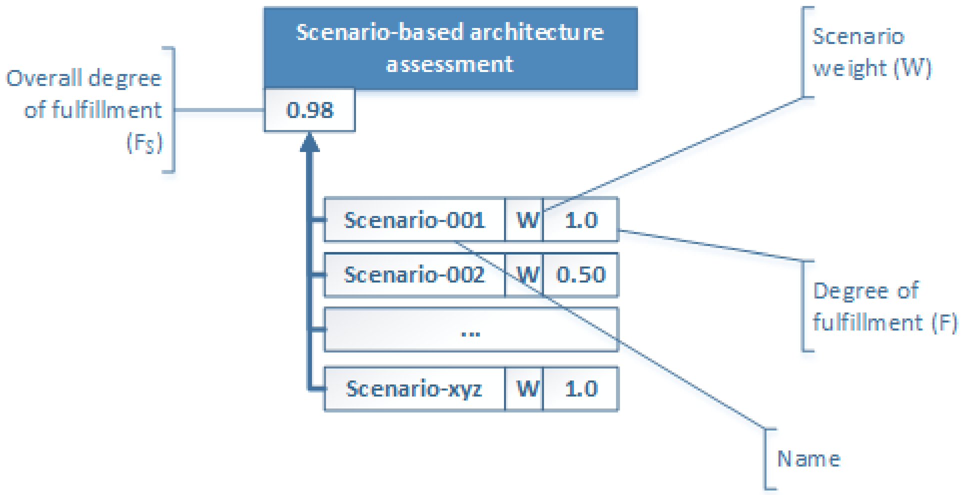

Figure 4: Scenario-based assessment structure

The whole process can be more formally described as shown below.

The *Overall Degree of Fulfilment* for the scenario assessment F_S can be calculated as

$$F_S = \sum_{k=0}^{n} W_{S_k} * F_{S_k} \quad (9)$$

Where W_{S_k} is the relative weight of the scenario *k* and F_{S_k} is the DoF of requirement *k*. The following two conditions apply:

First, the sum of all weights W_{S_G} is equal to one

$$\sum_{k=0}^{n} W_{S_k} = W_{S_G} = 1, 0.0 \leq W_{S_k} \leq 1.0, W_{S_k} \in \mathbb{Q} \quad (10)$$

Second, the fulfilment degree of each scenario F_{S_k} is express as percent value

$$0.0 \leq F_{S_k} \leq 1.0, F_{S_k} \in \mathbb{Q} \quad (11)$$

2.2. C-ITS Non-Functional Architecture Validation Methodology: Quality Attributes

In this section the approach to validate the C-ITS architecture from the perspective of the non-functional requirements or quality attributes is described.

2.2.1. Validation Methodology Objectives

2.2.1.1. Goals of the task

The C-Mobile Project Agreement states the task as follows [4]:

"Definition of a methodology for the validation of the C-ITS implementation architecture to be developed in Task 3.4".

The methodology requires the definition of approaches, methods, techniques and tools applied, in our case, to the validation of non-functional or quality requirements for the implementation of the C-ITS architecture. This also implies that both architecture and requirements to validate against must be known as prerequisites to the study.

2.2.1.2. Relation to other Work Packages

The common requirements to the C-ITS architecture are developed in WP2, the implementation architecture itself is the final task for WP3. Figure 5 shows the relation between these work packages. In the package descriptions, the project delivery months are shown in brackets.

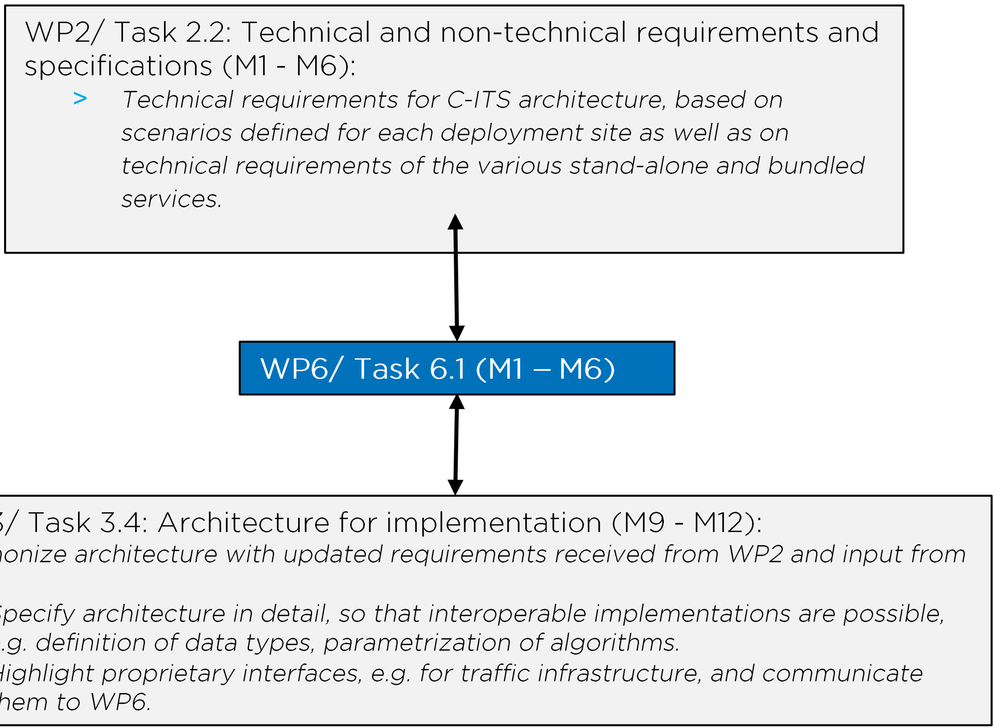

Figure 5: Relation of WP6 to other C-MobILE work packages

2.2.1.3. Quality attributes to validate

The definition of the groups of non-functional requirements, or quality attributes, is part of the WP3 effort. This effort is led by TU/e. Currently, the following set of quality attributes are considered by WP3 partners as important for validation purposes:

- Performance,
- Security,
- Reliability,
- Usability,
- Availability,
- Adaptability (modifiability),
- Interoperability.

Terms and meanings of the quality attributes and are provided as specified in the ISO/IEC 25010 international standard, also known as Systems and software Quality Requirements and Evaluation (SQuaRE) quality model [19].

2.2.1.4. Risk to the validity of the results

As Figure 5 shows, the delivery month of the technical requirements is M6, which is in the same time as the current deliverable. The delivery month of the implementation architecture is M12. That is, the methodology has to be developed at the same time than the requirements to validate against and well before the subject of the validation, the implementation architecture. Besides, there is an understanding between the partners in

WP2 that the requirements are going to be refined later in the project, when more specifics are added to the use cases in WP2.

Obviously, this situation imposes the *risk to the validity of the developed methodology*. It is our understanding that the methodology must be quite generic at the current stage of the project in order to be refined in the Task 6.2 later on, when, both, the technical requirements and the implementation architecture are delivered.

2.2.2. Architecture Validation

2.2.2.1. Validation and Verification Process

The current IEEE Standard for Software Validation and Verification [10] defines validation as "the process of evaluating a system or component during or at the end of the development process to determine whether a system or component satisfies specified requirements," and verification as "the process of evaluating a system or component to determine whether a system of a given development phase satisfies the conditions imposed at the start of that phase."

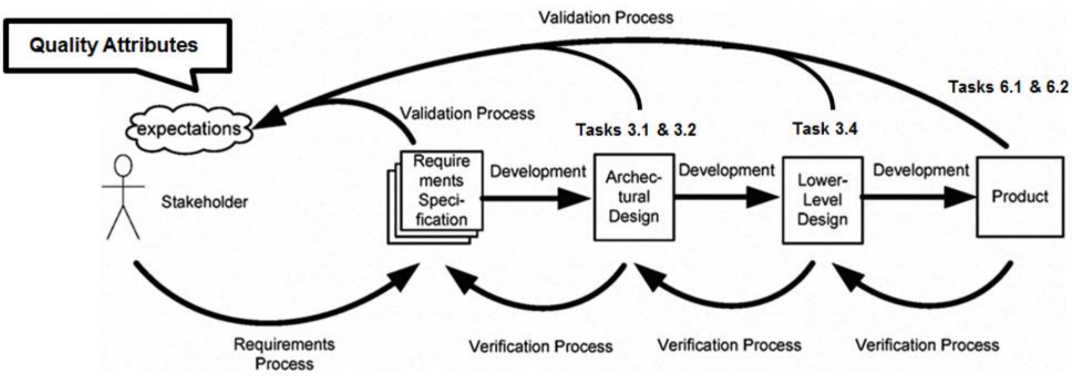

Figure 6: Validation and Verification process

Figure 6 shows the validation and verification process (adopted from [20]) where we put the references to the corresponding C-MobILE tasks. The task 2.2 creates the requirements to validate against; the tasks 3.1 and 3.2 define reference and concrete architectures, and task 3.4 produces the implementation architecture to validate. Finally, tasks 6.1 and 6.2 of the current work package constitute the validation process. The stakeholder expectations, in our case, are the quality attributes defined in 2.2.1.3.

From the architectural level perspective, the architectural design and lower-level design correspond to the reference architecture of the task 3.1, concrete architecture of the task 3.2 and the implementation architecture from Task 3.4. Figure 7 (adopted from [20]) shows the levels of architectural abstraction in relation to the requirements satisfied and models specified.

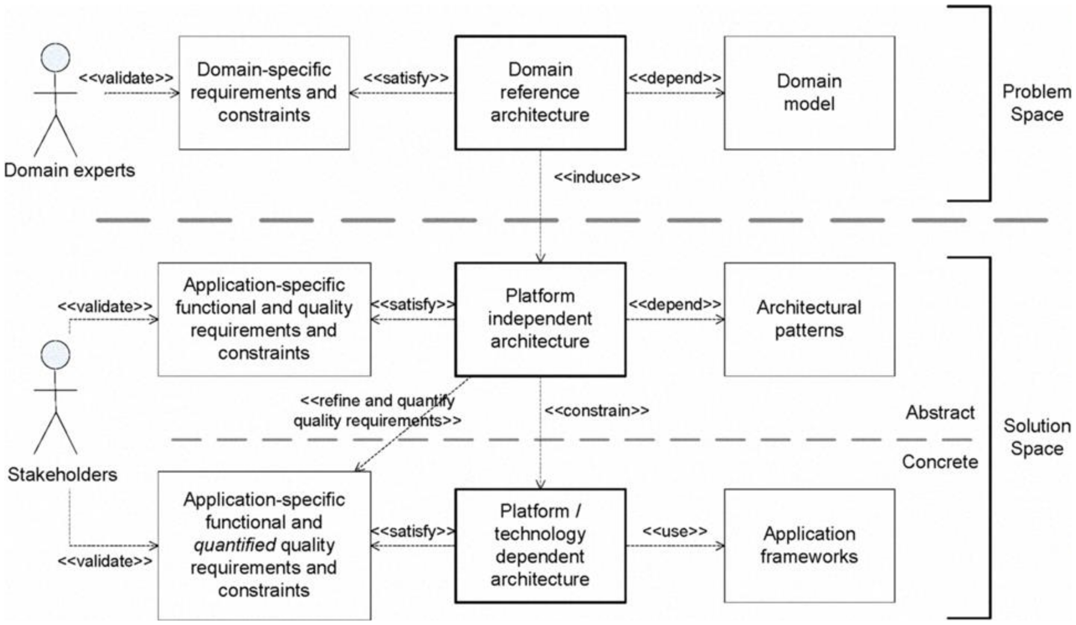

Figure 7: Levels of Architectural Abstractions

2.2.2.2. Architectural Description Language

The validation approach must take into account the architectural description language (ADL) in which the validation target architecture is described. On the basis of information provided by WP3, SysML is used as an architecture description language for C-MobILE architectures. SysML is a general-purpose architecture modelling language for systems engineering, supporting the specification, analysis, design, verification and validation of a broad range of systems and systems-of-systems [21]. SysML is adopted by the Object Management Group (OMG) and *OMG SysML v1.5* is used by WP3 for modelling the C-MobILE architecture. The relation between quality attributes and artefacts of the architecture description is shown in Figure 8.

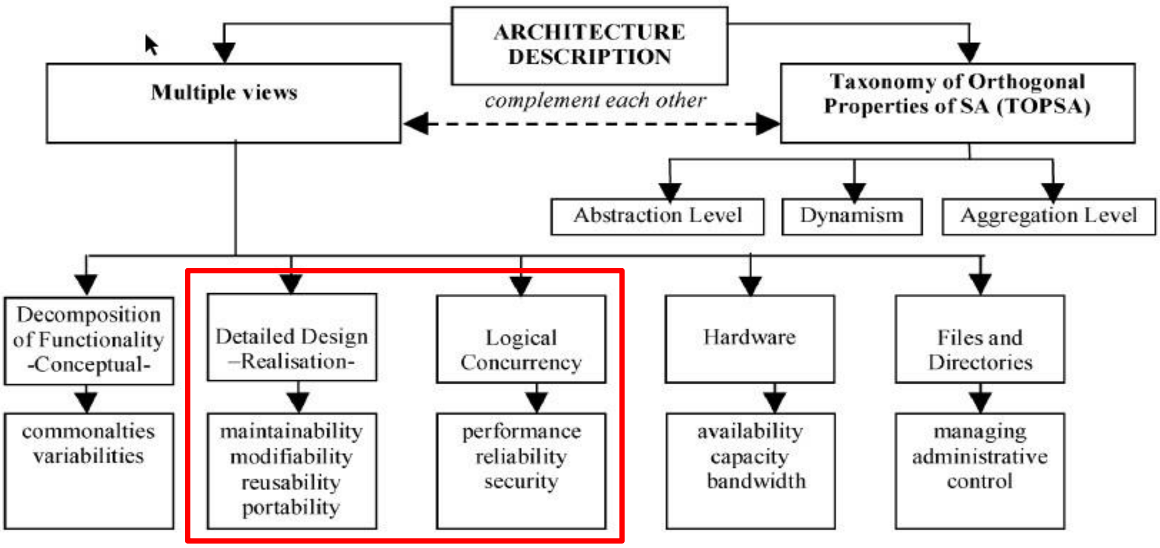

Figure 8: Architecture description and the relevance to analysis of quality attributes

In this figure adopted from [22], we highlighted with a red square the relevant quality attributes to our study and their corresponding design artefacts. SysML diagram types are briefly described in the resulting deliverable from task 3.1 and shown in Figure 9 [21].

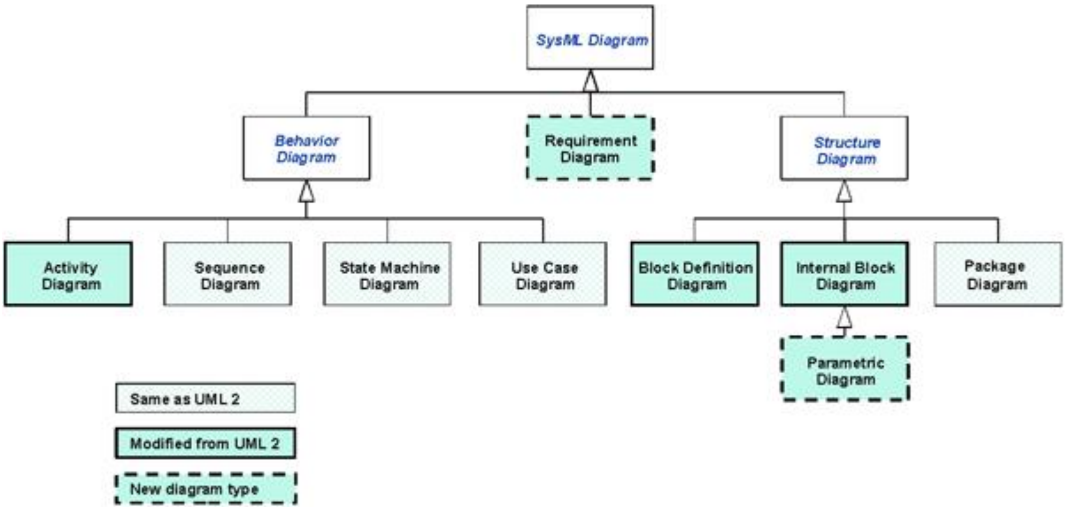

Figure 9: SysML Diagram Types

2.2.2.3. SysML in relation to quality attributes

SysML diagram types are describing the detailed design perspective and, thus, the tasks relevant to the non-functional validation:

/ Block Definition Diagram

/ Internal Block Diagram

/ Package Diagram

The diagram types relevant to the logical concurrency are:

/ Activity Diagram

/ Sequence Diagram

/ State machine Diagram

/ The special SysML type of diagram, the requirements diagram, can be useful in the non-functional requirements validation, since it can give an overview of the requirements in their interrelation.

/ All these types of diagrams can show in a human friendly form the SysML model representing the implementation architecture to validate. Regarding the way to transform that model form the XMI format to all those diagrams, still remains a research question, as at the moment the only high level structural models are defined in the scope of the reference architecture, while the SysML models for the implementation architectures have not been defined due to the fact that the task itself has not started yet (for the risks to validity see 2.2.1.4).

2.2.3. Architecture Validation Approaches

2.2.3.1. Overview and justification of choice

To validate the architectural decisions in meeting the non-functional or quality requirements against the stakeholder's expectations, there are four major categories of techniques. These techniques address the assessment of the quality of software architectures [20]:

/ Scenario-based,

/ Simulation-based,

/ Mathematical-/logical-based,

/ Metric-based.

Scenario-based approaches are useful in validating non-quantifiable non-functional requirements (e.g., security, safety, fault-tolerances, adaptability, reusability, performance, etc.). One can use the scenario-based method to develop a set of scenarios that elaborate the meaning and implications of a non-functional requirement and then show the stakeholder the scenarios to discuss the adequateness of the software architecture in addressing quality [20].

Simulation-based approaches make use of executable architectural models, which can substitute the real product at the design stage of the development. This way, the requirements can be validated early and possible design flaws corrected at much lower cost. These approaches require sophisticated models and tools usually specific to every quality attribute and, as a result, demand significant effort of highly skilled developers. That's why those approaches are usually applied in critical areas such as security.

Mathematical-/logical-based approaches are the realm of so-called formal methods. In these methods, the investigated quality attribute is described as a required property in a formal way allowing to prove that the system possesses that property. Naturally, different properties yield different types of models. Formal methods are very complex and time and effort consuming. Their complexity grows rapidly with the complexity of the system. They are usually applied to precisely targeted tasks in where the validity of the results is of crucial importance, such as safety.

Metric-based approaches are trying to assign a value – a 'metric' - to some system artefacts or a feature and, then, to relate that metric to particular quality attributes. The system of such relations is called a 'quality model' [19]. Each metric in the quality model has its own threshold or scale, which allows one to reason whether and how the system is 'good' or 'bad' from the point of view of that quality attribute. The basis for the quality model and those thresholds can be scientific research, industrial studies ('best practices') or just expert opinions. Metric-based approaches are relatively low-cost. Once established, they do not require much effort and experience to apply. However, any of them is as good as the quality model (metrics) it uses. The metric-based approach has proven to be quite useful in software maintainability assessment, to which our *adaptability* attribute is related [19]. As for the other architecture related quality attributes it has not proven to be so successful. The problem lies in the quantifiability of certain quality attributes and the justification of the corresponding metrics and thresholds.

The brief overview above illustrates *our rationale for choosing scenario-based approaches* as they are:

- potentially applicable to different quality attributes
- relatively low effort and resource consuming
- not requiring sophisticated models and tools
- relatively easy to learn

However, we do not completely rule out using ***simulation-based*** approaches for critical areas should the corresponding tools exist and be used by the project partners. Neither do we completely reject ***metric-based approaches*** taking into account that FESTA [1] is one of them. Below in this report, we describe the relation of the methodology we propose to FESTA.

2.2.3.2. Overview of existing methods and choice criteria

2.2.3.2.1. Architecture Trade-off Analysis Method (ATAM)

2.2.3.2.1.1. Description of ATAM

The Architecture Trade-off Analysis Method (ATAM) is one of the most well-known methods in the area of software architecture evaluation. The ATAM evaluates software architectures relative to quality attribute goals and its evaluations expose architectural risks that potentially inhibit the achievement of an organization's business goals [23]. Furthermore, the ATAM evaluation used in the early stage helps in revealing the degree to which the architecture satisfies particular quality goals and how they trade off against each other. This assists in refining the architecture under evaluation, as well as to avoid costly errors occurring in later development stages.

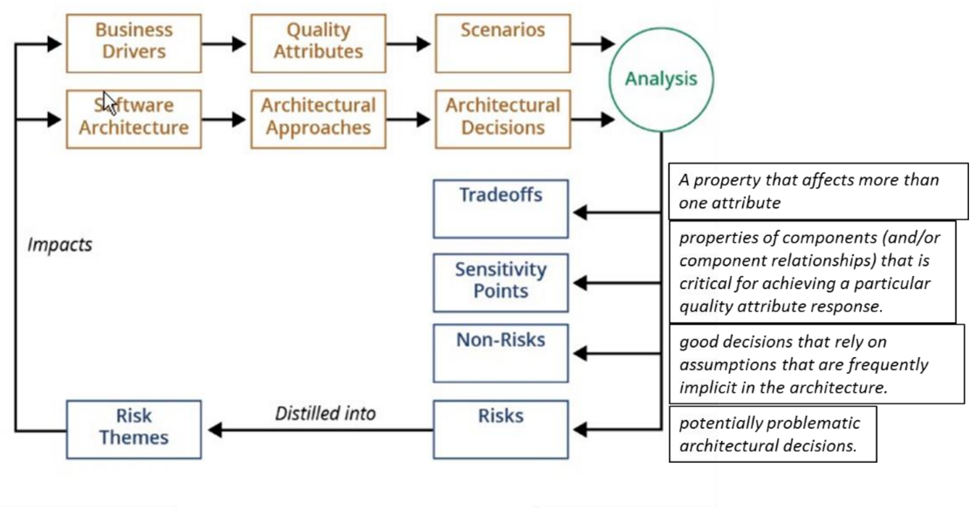

Figure 10: ATAM conceptual flow with descriptions

Using ATAM at the early stage of the development, i.e. architecture and design phase, helps selecting a suitable architecture for a software system by identifying trade-offs and sensitivity points. In Figure 10, the descriptions for the conceptual flow of ATAM are provided in a box (in italics). The reason that ATAM evaluation is useful for validating the C-MobILE implementation architecture is that there are many design options at implementation architecture level, since it covers not only architecture for a single system but for a set of C-ITS systems.

ATAM consists of the following nine steps which are separated in two phases [23]:

Presentation phase
1. **Present the ATAM.** The evaluation leader describes the evaluation method to the participants. 2. **Present business drivers.** A project spokesperson describes what business goals are motivating the development effort and hence what will be the primary architectural drivers (e.g., high availability or time to market or high security). 3. **Present architecture.** The architect will describe the architecture, focusing on how it addresses the business drivers. 4. **Identify architectural approaches.** Architectural approaches are identified by the architect but are not analysed. 5. **Generate quality attribute utility tree.** The quality factors that comprise system "utility" (in task 3.1 it is referred as "architectural perspective") (performance, availability, security, modifiability, usability, etc.) are elicited, specified down to the level of scenarios, annotated with stimuli and responses, and prioritized. 6. **Analyze architectural approaches.** Based on the high-priority factors identified in step 5, the architectural approaches that address those factors are elicited and analysed (for example, an architectural approach aimed at meeting performance goals will be subjected to a performance analysis). During this step, architectural risks, sensitivity points, and trade-off points are identified.
Evaluation phase
7. **Brainstorm and prioritize scenarios.** A larger set of scenarios is elicited from the entire group of stakeholders and voted/ ranked by the group members. 8. **Analyze architectural approaches.** This step reiterates the activities of step 6 but using the highly ranked scenarios from step 7. Those scenarios are considered to be test cases to confirm the analysis performed so far. This analysis may uncover additional architectural approaches, risks, sensitivity points, and trade-off points, which are then documented. 9. **Present results.** Based on the information collected in the ATAM (approaches, scenarios, attribute-specific questions, the utility tree, risks, non-risks, sensitivity points, trade-offs), the ATAM team presents the findings to the assembled stakeholders.

Table 6: ATAM steps separated in two phases

After phase 1, the state and context of the project, the driving architectural requirements and the state of the architectural documentation are known and phase 2 finishes the evaluation.

Making decisions for platform dependent design choices are almost impossible without proper architecture evaluation method. Therefore, we propose ATAM to help evaluate and refine the implementation architecture which later can be adapted to different deployment site architectures. During the adaptation phase, ATAM evaluation can be carried out for the deployment site architectures to ensure the specific quality goals and business drivers relevant for each deployment site.

2.2.3.2.1.2. Use case example: C-ITS architecture and security concerns

As described in step 5 of the ATAM process, the quality factors comprise system "utility" (in task 3.1 it is referred as "architectural perspective") e.g. performance, availability, security, modifiability, and usability. A utility tree consists of quality factors or quality attributes which are specified down to the level of scenarios, annotated with stimuli and responses, and prioritized.

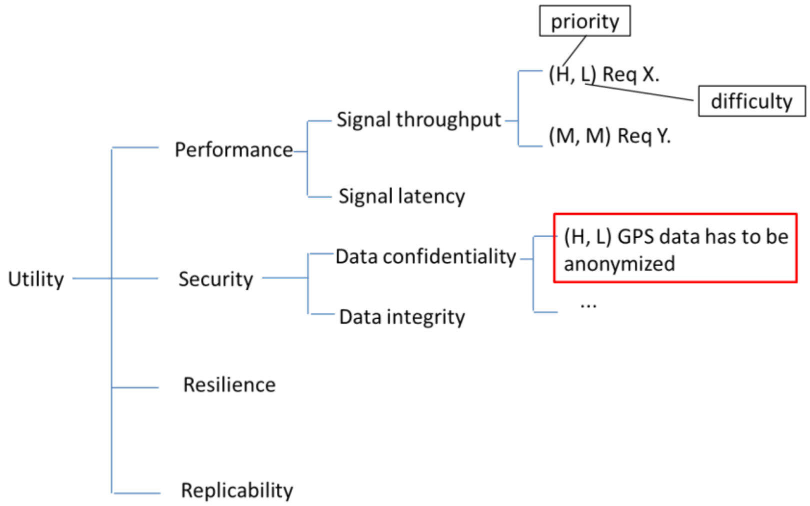

Figure 11: ATAM Utility Tree Example

Table 7 illustrates an example of an ATAM utility tree. Architectural tactics for security are described in deliverable 3.1 which can be applied when ensuring security concerns. An example security scenario and trade-off analysis are described in the tables below.

Scenario	GPS-data leak in transport domain
Attribute	Security – data confidentiality
Stimulus	Network transmission is interrupted by an intruder
Response	No private data is leaked

Architectural decision	Sensitivity	Trade-off	Risk	Non-Risk
GPS +Asset info data is sent via Internet	Data transmission is sensitive to outside intruders	Usability (cost?) vs Security	Private data is leaked to public domain	
Asset info is stripped from personal/car identification, IP address data	Data preprocessing is sensitive to overall network response time	Security vs Performance	Data delay/latency more than 200 ms	

Table 7: An example security scenario and trade-off analysis

2.2.3.2.2. ATAM and FESTA

In 2.2.3.1 it is mentioned that metric-based approaches can be used for the task of architecture validation. FESTA is essentially a metric (measurement)-based approach to *product validation against* (in many cases) *non-functional impact areas,* such as safety, efficiency, environmental impact, etc. In C-MobILE, the quality attributes of the C-ITS services (2.2.3.1) have been included as target *technical impact areas*. In this context, the question whether or not we could apply FESTA to the validation of the C-ITS implementation architecture needs to be addressed.

Within the C-MobILE framework, the FESTA approach can be applied to the architecture validation provided that:

- the validated [part of] architecture has a working prototype which implements it and
- the corresponding performance indicator (PI) justifies it.

First of all, this means that the corresponding PIs have to be developed for such quality attributes (a.k.a. technical impact areas) as performance, security, reliability and so on should the developer wish to use the FESTA approach to validate it.

On the other hand, using the FESTA approach to the C-MobILE implementation architecture has limitations and risks:

1. Not all or at least most of the developed C-ITS services have full-fledged working prototypes.
2. Current FESTA does not have PIs related to technical impact areas. The question here lies in the possibility of developing such PIs.
3. Impacts are assessed in isolation; some of the technical impacts can influence each other (architectural & SW implementation trade-offs are well established fact in computer science, e.g., performance/security; security/usability; performance/adaptability (reuse), etc.).
4. If the C-MobILE architecture comprises architectural decisions from different prototype services, they can have negative effect on each other from a technical impacts point of view.

As mentioned in the previous sections, ATAM validates the technical quality attributes of the product, which do not yet exist, at the earlier stages of the development process, when errors are less costly to fix. The method's strength lies in the fact that it points out at the trade-offs between different attributes; indicates risks and sensitivity points.

Therefore, we argue that in cases where FESTA can be applied to justify certain design decisions from the point of view of certain technical impacts, ATAM can complement it by the relevant trade-off analysis.

2.2.3.3. Tooling for architecture validation

2.2.3.3.1. Support from commercial tools: Enterprise Architect

Although the application of ATAM requires sophisticated manual analysis by human experts, automation can help at some steps of the analysis.

First of all, the target XMI format of the C-ITS implementation model is not very user-friendly. Tools can help to transform it into readable format, preferably into the set of SysML diagrams relevant to the validation approach (2.2.2.3). Enterprise Architect v11 [24] has been chosen by our partners from WP3. This tool has a strong support for SysML and, thus, can be used for the purpose of ATAM architecture analysis.

2.2.3.3.2. ATAM-Specific Tool Support

Except anecdotal references to academic tools, no tool support for scenario-based architecture validation exists. Probably, various specialized templates and tables need to be developed within Task 6.2 to facilitate making utility trees and designing ATAM scenarios. This way, the consistency of the analysis can be supported across the C-ITS services.

3. Field Operational Test concept and scope

3.1. The Field opErational teSt supporT Action (FESTA) methodology

The original FESTA handbook was produced in 2008 by the FESTA consortium (Field opErational teSt supporT Action, 2007–2008). The FOT-Net and FOT-Net 2 consortia updated this handbook several times in order to take into account the lessons-learned from the many FOTs that have been conducted since, and the insights and ideas shared between experts in workshops, international workshops, seminars and stakeholder meetings. The latest version, version 6, was produced in the end of 2016 by the FOT-Net Data consortium and is available at www.fot-net.eu [1].

A FOT is defined by the FESTA Handbook as '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', while the EC officials define FOT as 'large-scale testing programmes aiming at a comprehensive assessment of the efficiency, quality, robustness and acceptance of ICT solutions used for smarter, safer, cleaner and more comfortable transport solutions, such as navigation and traffic information, advanced driver assistance and cooperative systems'.

The FESTA methodology is summarized below and demonstrated in Figure 12. There are several steps, which although described in a linear way, are performed in iteration. The V-shape shows the dependencies between the different steps in the left- and right-hand side of the V. The steps can be summarized as:

- Defining the study: Defining functions, use cases, research questions and hypotheses
- Preparing the study: Determining performance indicators, study design, measures and sensors, and recruiting participants
- Conducting the study: Collecting data
- Analysing the data: Storing and processing the data, analysing the data, testing hypotheses, answering research questions
- Determining the impact: Impact assessment and deployment scenarios, socio-economic cost benefits analysis

There are several steps that are of importance for setting up the evaluation framework for C-MobILE. It contains the preparation activities for setting up the test. The bottom part represents the data acquisition during the use of the systems. And the right side represents the data analyses and the interpretation of results.

The preparation phase follows a research-oriented approach. First, the functions to be tested are defined, the use cases are described and the related research questions listed. The use cases should describe daily situations where the system is expected to respond according to the specific functions, while the research questions should be statistically testable and evaluate the performance of the systems within the use cases. Second, hypotheses, performance qualitative or quantitative indicators and measures and sensors should be defined. The hypotheses should answer the research questions through direct measures or indirect estimations/ calculations of the related indicators.

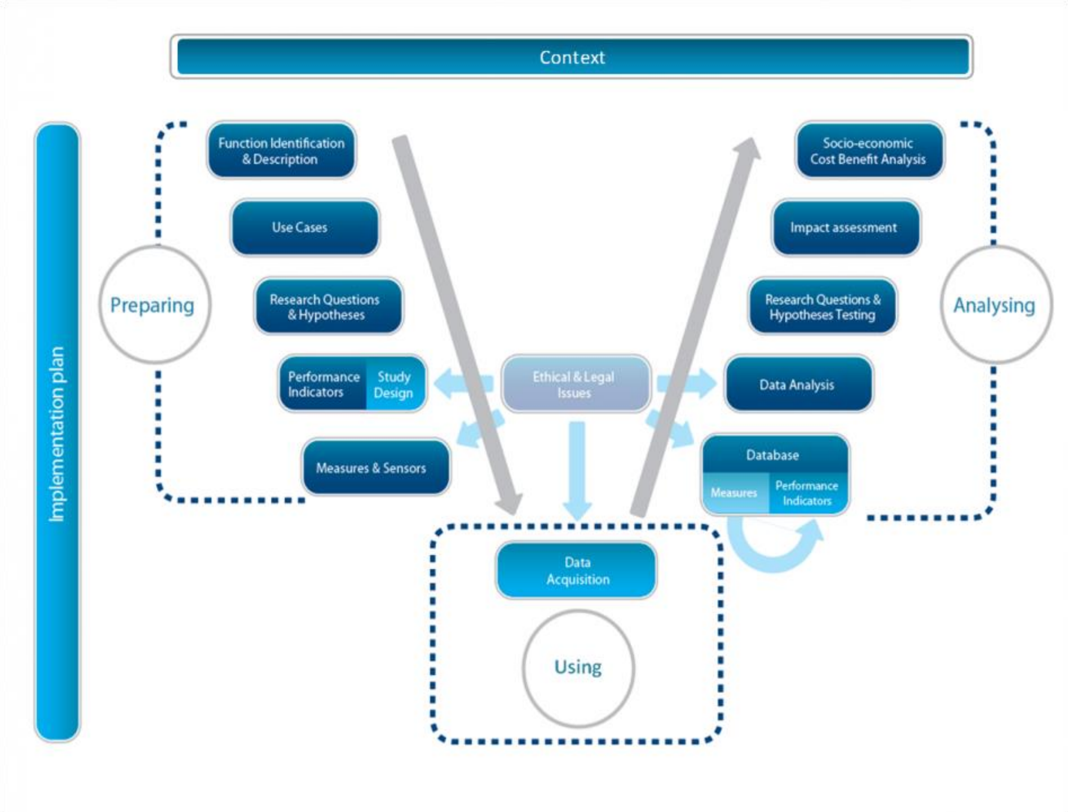

Figure 12: FESTA Steps for carrying a FOT

3.1.1. Basic Concepts

3.1.1.1. System

A system is a combination of hardware and software enabling one or more functions [1]. In a system, there are a set of elements (e.g. sensors, a controller, and an actuator) related to each other according to a design. An element of a system can be another system at the same time. Then this is called a subsystem according to FESTA-methodology.

3.1.1.2. Function

From an engineering point of view, a function is an action, activity or task that must be accomplished by a system to achieve a desired outcome. It indicates the outcome of the system from the driver's point of view. The outcome of a function is directly related with driver behaviour and expected impacts on driver behaviour (EAST-EAA in [47]).

3.1.1.3. Use case

A use case defines a subset of functionalities of a system. It is a description of how a function is intended to interact with the driver in a particular target scenario. Use cases treat the system as a black box, and the interactions with the system, including system responses, are perceived as from outside the system. Each use case captures: 1) the actor (driver); 2) the interaction, how does the system react to the driver's input; 3) driver's goal [39].

3.1.1.4. Scenario and target scenario

A scenario is a synthetic description of an event or series of actions and events. A target scenario is a precise formulation of the problem situation to be addressed by a function or use case developed.
The target scenario describes the scenario where an envisioned system can be operational and proactive in order to prevent accidents or other undesired outcomes. A target scenario thus describes the problem to be

solved, the accident scenario, but not the solution (the interaction between the system and the driver). The target scenario is defined in terms of environmental-, vehicle- as well as driver parameters.
Use cases are generally derived from target scenarios. Thus, the target scenario defines the problem and the use case defines, on a suitable level of abstraction, how it should be solved by the function in interaction with the user. The next step is therefore to define how the intended functions should prevent/mitigate the undesired outcomes defined by the target scenarios. This is the role of the use cases.

3.1.1.5. Test scenario

A test scenario is a use case in a specific situation according to FESTA methodology. A situation is a combination of certain characteristics of circumstances and features of a system.
Test scenarios are test cases, and the sequence in which they are executed. Test scenarios are test cases that ensure that everything is tested from beginning to end. Test scenarios can either be independent tests or a sequence of tests where each is dependent upon the output of the previous one. Test scenarios are prepared by reviewing functional requirements and preparing logical groups of functions that can be further broken down into test procedures. Test scenarios are designed to represent both typical and unusual situations that may occur in/with the system. Specification of a test scenario defines which features of the system and which circumstances will be covered.
The relation of the basic concepts described above is illustrated with simplified exemplary cases in Table 8. Test scenarios will be documented in D6.2 “Technical validation report” and D6.4 “Report on impact of C-ITS on surface transport system”.

Terminology	Example
System	C-MobILE bundle 1
Function	In-Vehicle Signage
Use Case	Driver is approaching a school zone sign and gets information about it.
Scenario	> Driver approaching a school zone in the morning/ in the evening > Information is given about 60s/30s before the traffic lights > Etc.

Table 8: Basic concepts in FOT

3.1.1.6. Research Questions

A research question (RQ) expresses a relevant research objective that is targeted within C-MobILE. A RQ can be translated into a hypothesis, which in turn results in certain Performance Indicators (PIs) that can be studied in order to be able to answer the hypothesis and research question.

3.1.1.7. Hypothesis

A hypothesis is formulated as an answer to a RQ and an assumption of an outcome of an event. A hypothesis is a specific statement which can be tested with statistical means by analysing measures and performance indicators(PIs) [1]. In an experimental design a hypothesis is a testable postulate of that outcome. In the experimental language a hypothesis is the predicted outcome of an experimental manipulation stating the relationship between two variables. In case of C-MobILE, a hypothesis addresses the relationship between the use of a system and its benefit such as the use of In-Vehicle Signage and the number of speeding’s.
This can be translated into the statistical and thus testable null (H0) and alternative hypotheses (H1) which are statements about the conditional probability of data under a certain distribution within the population.
For example, the null hypothesis states that there is no relationship between two variables, whereas the alternative hypothesis states that there is a relationship between theses variables and can also state the direction of the relationship as presented in the example in the following Table 9.

Hypothesis	Example
H0	The use of In-Vehicle signage does not influence the number of speeding’s
H1 (directed)	The use of In-Vehicle Signage is reducing the number of speeding’s
H1 (undirected)	The use of In-Vehicle signage influences the number of

	speeding's

Table 9: Example for hypotheses

3.1.1.8. Event

An event is a specific driving situation relevant to the function. Time-wise, logging an event begins when the driver reaches the distance at which the function is activated or would be activated and ends when the driver encounters the location/ situation that triggered the activation of the function or deviates to an alternative route, or the influence of the function ends. In case of the example of In-Vehicle-Signage presented above, the event can be reaching a school zone sign within a certain frame of meters.

3.1.1.9. Performance indicator and success criteria

According to FESTA, a PI is a quantitative or qualitative measurement agreed on beforehand expressed as a percentage, index, rate or other value which is monitored at regular/irregular intervals. It can be obtained directly from measures or derived from CAN-bus of the vehicle, from external sensors, simulation procedures, questionnaires or events. PIs are chosen to test and allow the indication of support for hypotheses. The PIs can be compared with one or more success criteria that are quantitative or qualitative thresholds for performance indicators.

3.1.1.10. Measures and sensors

A measure can be logged directly from a sensor, a simulation, or derived from other measures. Sensors indicate how measures will be collected and can be independent elements or part of system hardware or also an internal procedure within simulation software. Measures can have different categories as stated within the project eCoMove (see Table 10).

Measure category	Description
Direct (raw) measures	A direct measure is logged directly by a sensor, without any processing before saving the data to the log file, coming from the CAN bus of the vehicle, for instance.
Derived measurements	A pre-processed measurement is not directly logged by a sensor, but is either a variable that has been filtered or is a combination of two or more direct or derived measurements, such as CO_2 emissions derived measurement from fuel consumption measurements.
Self-reported measurements	Measurements that are obtained from questionnaires, interviews, rating scales, etc. used, for instance, when measuring user acceptance

Table 10: Measure categories and descriptions

3.1.1.11. Baseline and treatment

The situation without the use of the systems (system non-present, off or on without providing information/warning) is defined as the baseline, whereas the situation with the use of the systems (system on and providing information/warning) is defined as treatment period. The baseline is the basis for the benefit assessment of the system and the comparison between systems.

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 of equal length so that there is the opportunity for the same variations to occur both in the baseline period and in the treatment period (such as seasonal effects).

4. C-MobILE FOT Methodology

C-MobILE sets the basis for large scale deployment in Europe, elevating research pilot sites to deployment locations of sustainable services that are supported by local authorities, using a common approach that ensures interoperability and seamless availability of services towards acceptable end user cost and positive business case for parties in the supply chain.

Large-scale deployment covers three dimensions:

- Number of C-ITS services: C-MobILE realizes and deploys a large number of C-ITS services. Services are be combined into bundles. Bundling of applications provides additional benefits to the transport system, as it creates a flexible and intelligent suite of C-ITS applications, to be deployed and used by a large number of end users. Furthermore, applications bundling eases the widespread introduction of C-ITS, as it makes several C-ITS services relevant and accessible to all end users, as well as due to the variety of policy related objectives of stakeholders that can be met through their deployment. The application bundling was revised early in the project, to ensure that it still matches the deployment site environments and reflects any advancement in the current state-of-the art in C-ITS technologies.
- Number of users: C-MobILE targets a higher number of users. Moving from the basis of consortium members and initial associated partners, the partnership in each deployment site will be extended all along the project duration with local, international and global actors from the private sector (including telecom, infrastructure and service providers, transport service operators, and original equipment manufacturers), public authorities, emergency services, and other stakeholder groups, including professional and private end-users. The stakeholder's forum will serve to liaise/ exchange also with other cities, projects on innovative solutions for C-ITS deployment at pan European and international level.
- Size of geographic area: C-MobILE targets a higher geographic area. Activities are conducted with a wide range of external cities/ regions experienced or interested in deploying innovative C-ITS solutions to address mobility challenges in their territory. The project targets key personnel in public authorities (i.e., representatives, TMC and technical staff, and management staff), in order to raise knowledge, understanding and create necessary skills, and ultimately increase the replication possibilities in cities external to C-MobILE.

C-MobILE elevates the evaluation:

- Improved level of performance for the entire surface transport system, including more efficient and sustainable traffic management, improved safety and contribution to overall socio-economic development. C-MobILE will upgrade and expand the deployment of C-ITS services at a wide level, utilizing and leveraging knowledge and experience gained from past and ongoing C-ITS pilots and corridors. The project demonstrates, validates and assesses the wide-scale C-ITS deployment in European cities, through the provision of reliable figures pertaining to – benefits in traffic safety (including VRUs), efficiency, sustainability, economic viability, and user acceptance. Building on the findings of the C-ITS Platform, the C-MobILE bundling approach is expected to leverage individual C-ITS applications impacts and improve safety, mobility and sustainability levels.
- *Testing and demonstration of fully integrated C-ITS concepts in real-life, complex environments.* The C-MobILE innovative applications bundles will be demonstrated at 8 deployment sites. They are going to be deployed in complex urban settings with diverse, yet representative environments across Europe, involving large and different user groups (i.e. drivers, pedestrians, cyclists). C-MobILE will exploit, extend and integrate current applications across Europe, developing and demonstrating viable C-ITS solutions that will ensure sustaining these activities well beyond the project duration, due to the increased efficiency of bundled C-ITS applications, provided the support of local authorities and engagement of other stakeholders, as well as the vision towards next generation of automated vehicles and intelligent road infrastructures.
- Greater collaboration (and partnerships) between multiple stakeholders to deploy applications and facilitate the interoperable interactions across all elements of the transport system, including the use of data from multiple sources. C-MobILE brings together the entire ecosystem of C-ITS stakeholders and creates communities comprised by all key stakeholders in the business value chain: city authorities, policy makers, service-, technology- and telecom providers, vehicle fleets, user associations, innovation agencies, and various categories of end-users (both professional and private citizens). Such stakeholders contribute to create the C-MobILE C-ITS Framework paradigm and to the achievement of project objectives, by committing actively to the extended, viable and interoperable deployment of C-ITS across Europe. C-MobILE actively supports existing and new collaborations and formulation of formal and informal partnerships (depending on country and city specific arrangements), so that deployment and

uptake of C-ITS are widely supported. Furthermore, C-MobILE is adopting an innovative C-ITS architecture that utilises multi-source datasets to address key safety and efficiency issues. The architecture promotes interoperability and also allows for improved use of and accessibility to data in the transport domain. Thus, the project is expected to actively contribute to further innovation for the formation of sustainable collaborative schemes between all involved stakeholders. C-MobILE finally strongly promotes an active and mutually beneficial dialogue with related experts and entities who are at the forefront of C-ITS in the USA and internationally, fostering a structured and harmonized dialogue for continuous knowledge and experience exchange, also in relation to the development of harmonised validation methodologies.

- A comprehensive Cost-Benefit Analysis (CBA) based on the deployment sites demonstrations to prove the added value and economic viability of C-ITS services and solutions for users and other stakeholders. C-MobILE assesses mobility improvements and C-ITS infrastructure requirements for similar C-ITS services at a variety of deployment locations. This will lead to an overview of C-ITS deployment efforts and effectiveness in relation to road infrastructure, traffic demands, and other effects, which is the foundation of a comprehensive CBA. C-MobILE has adopted a comprehensive dual approach for increasing the effectiveness and use of the CBA results, by performing such analyses both ex-ante (in advance, using estimated data), as well as ex-post demonstration (afterwards, using measured data), in order to ensure that bundled C-ITS applications deployed within the project are financially viable and that they can be sustained also after the end of the project from a business perspective. Furthermore, the CBA model developed by C-MobILE will significantly contribute to decreased uncertainties and risks of relevant stakeholders.
- Validated results and proven impact on user acceptance, safety, resilience and security with respect to transport demand and the environment. C-MobILE is developping and implementing a commonly accepted methodology framework for the evaluation of the performance of the proposed C-ITS application bundles based on real-life testing. C-MobILE contributes to an improved understanding of C-ITS impacts by conducting thorough impact assessment, based on the FESTA methodology, for all deployed C-ITS applications addressing both users' acceptance and traffic safety. The variety of deployment locationsallows for achieving a high-level understanding of the performance of the C-ITS application bundles under different conditions by linking the results of the assessment to external conditions, such as transport demand, traffic conditions as well as geographic and functional characteristics of the network.
- *Development of validated guidelines for the large-scale deployment of operational and sustainable C-ITS services in Europe.* C-MobILE develops validated, transferable and interoperable C-ITS implementation and operational guidelines for cities beyond those participating in the consortium, to foster the wide adoption of the proposed services at EU level. Moreover, the methodology used in C-MobILE is being disseminated into guidelines for identification of prospect deployment in complex urban and extra-urban environments across Europe. These guidelines support the decision process on deployment of C-ITS by European city authorities and local communities.

C-MobILE follows the guidelines of the FESTA methodology for FOTs in Europe. In this sense, the way to evaluate the studied services is based on the interaction of various users and stakeholders with the services. The first step to follow a FOT is to identify the functions and use cases aiming at solving traffic related problems.
The evaluation process itself is based on the study of the behaviour of the users in different use cases against their behaviour during the baseline operation.
To do that effectively, research questions and hypotheses are defined which will be studied using performance indicators. These indicators influence which data need to be collected and what measurements have to be recorded.

Therefore, the following sections describe for all C-MobILE services the:

- Function identification and description of services and bundles (mainly defined by WP2)
- Use cases and scenarios (mainly defined by WP2)
- Research questions and hypotheses (WP6 with support from WP2, WP3, WP4 and WP5)
- Performance indicators (WP6)

4.1. Objectives

This WP is dedicated to the validation of C-ITS services and architecture and to the post-demonstration assessment of their impact and value. It aims to prepare evidence for decision-makers on the pros and cons of bundled C-ITS services by addressing the following objectives:

- Define a methodology for the validation and impact assessment of C-ITS services
- Validate and evaluate the functionality, reliability and security of C-ITS services
- Assess the impact of C-ITS services on the end-users and the surface transport system
- Assess the acceptance of end-users to ensure C-ITS services uptake
- Demonstrate the economic viability of C-ITS services, and
- Provide guidelines for CBA for the large-scale deployment of C-ITS

4.2. Services identification and description

In order to develop an evaluation methodology, it is important to understand the functions of the services in detail. The functions are supposed to contain all relevant specifications of the system and to describe the limitations and operational description of each service. Partially, they are described in D2.2 "Analysis and determination of use cases".

A short description of all services is provided next:

Rest-Time Management (RTM)

Rest time management supports managing the working hours of drivers engaged in the carriage of goods and passengers by road. The process is regulated by policies, laws or regulations (e.g. [6]) that lay down the rules on driving times, breaks and rest periods for the drivers.

Motorway Parking Availability (MPA)

Motorway parking availability provides motorway parking availability information and guidance for truck drivers to make informed choices about available parking places. Existing solutions provide information about the location of truck parks, capacity, available equipment, facilities on site, security equipment and information about dangerous goods parking.

Urban Parking Availability (UPA)

Urban parking availability provides parking availability information and guidance for drivers to make informed choices about available parking places. This service aims to reduce congestion, time loss, pollution, and stress caused by cruising for parking.

Road Works Warning (RWW)

Road works warning aims to inform the drivers in a timely manner about road works, restrictions, and instructions. This allows them to be better prepared for potential works downstream on the road, therefore reducing the probability of collisions.

Road Hazard Warning (incl. traffic jams) (RHW)

The road hazard warning service aims to inform the drivers in a timely manner of upcoming and possibly dangerous events and locations. This allows drivers to be better prepared for the upcoming hazards and make necessary adjustments and manoeuvers in advance. (This is also known as "Hazardous location notification" [5] or 'Road hazard signalling').

Emergency Vehicle Warning (EVW)

Emergency vehicle warning uses information provided by the emergency vehicle to inform a driver of another vehicle about an approaching emergency vehicle even when the siren and light bar of the emergency vehicle may not yet be audible or visible. This is also known as "Emergency Vehicle Alert (EVA)", which alerts the driver about the location and the movement of public safety vehicles responding to an incident so the driver does not interfere with the emergency response. The service is enabled by receiving information about the location and status of nearby emergency vehicles responding to an incident [31].

Signal Violation Warning (SVW)

Signal Violation Warning aims to reduce the number and severity of collisions at signalised intersections by warning drivers who are likely -due to high speed- to violate a red light. Also known as the "Signal violation / Intersection Safety" or "Red Light Violation Warning".

Warning Systems for Pedestrian (not limited to crossings) (WSP)

Warning system for pedestrian aims to detect risky situations (e.g. road crossing) involving pedestrians, allowing the possibility to warn vehicle drivers. Hence, the warning is based on pedestrian detection. The scope of the service can be extended to cover other VRUs (e.g. cyclists). The service is particularly valuable when the driver is distracted or visibility is poor. (Also known as "Vulnerable road user Warning" [5].

Green Priority (GP)

Green priority aims to change the traffic signals status in the path of an emergency or high priority vehicle (e.g., public transportation vehicles), halting conflicting traffic and allowing the vehicle right-of-way, to help reduce response times and enhance traffic safety. This service is also known as "Traffic signal priority request by designated vehicles" [7] or "Priority Request" [32]. Different levels of priority can be applied, e.g. extension or termination of current phase to switch to the required phase. The appropriate level of green priority depends on vehicle characteristics, such as type (e.g. HGV or emergency vehicle) or status (e.g., public transport vehicle on-time or behind schedule). The vehicles request priority for an intersection, and the traffic light controller determines in what way it can and will respond the request.

Green Light Optimal Speed Advisory (GLOSA) / Dynamic eco-driving

GLOSA provides drivers an optimal speed advice when they approach to a signalized intersection. This advice may involve maintaining actual speed, slowing down, or adapting a specific speed. If a green traffic light cannot be reached in time, GLOSA may also provide time-to-green information when the vehicle is stopped in the stop bar. Application of GLOSA takes advantage of real-time traffic sensing and infrastructure information, which can then be communicated to a vehicle aiming to reduce fuel consumption and emissions.

Cooperative Traffic Light for Pedestrian (CTLP)

Cooperative traffic light for VRUs aims to increase the safety of pedestrians through warranting priority or additional crossing time (i.e., extending the green light phase or lessening the red phase) based on pedestrian characteristics (or on special conditions, such as weather). The service can also be extended to cover other VRUs, such as cyclists.. The service is also known as "Pedestrian Mobility" [31] or "Traffic light prioritisation for designated VRUs".

Flexible Infrastructure (HOV, peak-hour lanes) (FI)

Flexible infrastructure aims to interchange information about the lanes provided to the traffic users according to the time of the day. It includes solutions such as reserved lane.

In-Vehicle Signage (e.g. Dynamic speed limit) (IVS)

In-vehicle signage aims to provide information to the driver about the road signs (and dynamic information, e.g., local conditions warnings identified by environmental sensors [31]). The purpose of this service is to increase the likelihood of drivers being aware of potentially dangerous conditions in case a roadside traffic sign is not noticed.

Mode and Trip Time Advice (MTTA)

Mode & trip time advice (e.g. by incentives) aims to provide a traveller with an itinerary for a multimodal passenger transport journey, taking into account real-time and/ or static multimodal journey information.

Probe Vehicle Data (PVD)

Probe Vehicle Data is data generated by vehicles. The collected traffic data can be used as input for operational traffic management (e.g., to determine the traffic speed, manage traffic flows by - for instance- alerting users in hot spots, where the danger of accidents accumulates), long term tactical/strategic purposes (e.g. road maintenance planning) and for traveller information services. Also known as Floating Car Data (FCD).

Emergency Brake Light (EBL)

Emergency Brake Light aims to avoid (fatal) rear end collisions, which can occur if a vehicle ahead suddenly brakes, especially in dense driving situations or in situations with decreased visibility. The driver is warned before s/he is able to realize that the vehicle ahead is braking hard, especially if s/he does not see the vehicle directly (vehicles in between).

Cooperative (Adaptive) Cruise Control (CACC)

Cooperative Adaptive Cruise Control represents an evolutionary advancement of conventional cruise control (CCC) and adaptive cruise control (ACC) by utilizing V2V communications to automatically synchronize the motion of many vehicles. While ACC uses Radar or LIDAR measurements to derive the range to the vehicle in front, CACC also takes the preceding vehicle's acceleration into account.

Slow or Stationary Vehicle Warning (SSVW)

Slow or stationary vehicle warning aims to inform/ alert approaching vehicles of (dangerously) immobilized, stationary or slow vehicles that impose significant risk.

Motorcycling approaching Indicator (including other VRUs) (MAI)

Motorcycle approaching indication informs the driver of a vehicle that a motorcycle is approaching/passing. The scope can be extended to cover other VRUs, such as cyclists and other Powered Two Wheelers (PTW). The motorcycle could be approaching from behind or crossing at an intersection.

Blind spot detection / warning (VRUs) (BSP)

Blind spot detection aims to detect and warn the drivers about other vehicles of any type located out of sight.

C-MobILE will realize and deploy C-ITS services bundles.

> A bundle is a coherent set of services that are deployed to address a particular scenario or need. Ideally, regarding evaluation, the effect of controlled dynamic traffic management operations, including traditional and smart mobility services, should be compared to regular dynamic traffic management operations, including traditional services. As controlled operations is only necessary in case of specific incidents/situations, it is not possible to compare this situation with other situations and therefore impossible to evaluate properly. The added value for operators-managers is to understand the effect of multiple simultaneously active services.

> A bundle is an open, modular and extendable wrap application that brings together a complete suite of C-ITS services under one common user environment, able to operate either in an automated mode or in a user-selected mode, with the ultimate goal of covering multi-parameter needs as well as easing the widespread introduction of C-ITS.

Type of evaluation and analyses

In principle the evaluation will be carried out as is described in this document, but depending on the frequency of events and warnings, a shift in focus might be possible. We want to put our resources on the services and bundles for which we expect the most impact. That means that possibly some research questions and/or hypotheses will be adjusted later during the deployment.

General technical and non-technical requirements for evaluation

In task 2.2 “Technical and non-technical requirements”, there have been some activities in order to identify general requirements related to the evaluation. Annex 1 includes the general requirements identified that are directly linked with the evaluation. Those requirements are quite general due to the early stage of the project and have been validated with deployment site leaders.

4.3. Use Cases

The use cases definitions are elaborated within Task 2.1 and describe how the systems work, identifying system requirements and providing a general description of the intended functionality of the systems, as a basis for the more detailed specification of the technical and non-technical requirements. D2.2 will contain a detailed description of use cases within C-MobILE.

4.4. Research Questions & Hypotheses

After the definition of use cases, research questions, that influence a relevant impact area, are defined and formulated. The impact areas foreseen within C-MobILE are: personal mobility, user experience/ acceptance, efficient traffic management, safety and environment.

C-MobILE objectives can be formulated as one or more questions that should be answered by evaluation of the deployment sites.

Research questions can be formulated generally or in more details.

- What are the benefits of bundling C-ITS services?
- Does bundling help to widespread C-ITS services?
- What is the suitability of C-MobILE bundles?
- Will C-MobILE bundles improve safety compared to isolated C-ITS services?
- Will C-MobILE bundles improve traffic congestion compared to isolated C-ITS services?

The research questions mentioned above need to be refined or decomposed to more concrete questions that can be answered more precisely.

The formulation of research questions is an elaborate and iterative process. Research questions can be further specified in the form of sub-research questions. From those sub-research questions, 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.

Research questions and hypotheses ideally should be valid for each deployment site and should be agreed upon by all stakeholders for consistent evaluation later on.

Research questions and hypotheses are collected and maintained in a spreadsheet on C-MobILE web repository ProjectPlace. The spreadsheet is a living document that will be adapted and extended throughout the project for services and bundles, evaluation methodologies, and refinement during evaluations.

Final research questions and hypotheses will be delivered in D6.2 “Technical validation report”, D6.3 “Report on impact of C-ITS on stakeholders and end-users” and D6.4 “Report on impact of C-ITS on surface transport system”.

4.5. Performance indicators (PIs)

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

PIs are very diverse in nature. There are key performance indicators as well as basic performance indicators assessed on individual level, observed and self-reported (subjective), calculated from continuous and from discrete data, and so on. An example of a key PI based on continuous log data would be the mean of mean speeds on motorways, whereas an example of a basic PI based on discrete, self-reported data would be the level of perceived usability of a function reported by one single user. 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 can be normalised in order to allow comparisons.

In C-MobILE, performance indicators at service level and/ or bundle level are being defined. Considering that bundles are open, modular and extendable wrap applications, which might contain a different set of C-ITS services in each deployment site, we can expect PIs for bundle evaluation to be a combination of discrete and self-reported data such as ratings. PIs should also allow to analyse changes on measurements defined for single C-ITS services and observe variations while deploying them as part of a bundle.

4.6. Measures

Measurement defines how the indicator is measured in the FOT. Four different types of measures are identified, namely direct measures, indirect/ derived measures, self-reported measures, and situational/ control variables, which are described in more detail below. A measure does not have a "denominator". Therefore, it is not comparable to other instances of the same measure or to external criteria. The measure itself, however, can very well be a fraction. 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.

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 must 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.

A *derived measure* is not directly logged from a sensor, but, for example, 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 speed difference between a vehicle and another vehicle or object is another derived measure, based on the speed of the vehicle as read from the CAN bus, for example, and the calculated speed of the other vehicle or object.

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 at least every 0.1 second (10 Hz).

A number of PIs are based on *self-reported measures*, which are gleaned from questionnaires, rating scales, interviews, focus groups, or other methods requiring introspection from the participant. These subjective measures are typically not logged continuously, but rather only once or a few times during the course of a 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 bundle, 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.

Situational variables are properties of the traffic systems the vehicles passed by while driving. They can be logged like direct measures or computed like derived measures. They can also be self-reported and 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 included in the PI matrix in the measures table, but they 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.

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 lane, potentially 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.

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

C-MobILE uses the PI-Measures-Sensors matrix, a spreadsheet in Excel format containing three tables: "Performance indicators", "Measures" and "Sensors", that are utilised later to create a relational database. 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 PI-Measures-Sensors matrix will be delivered in D6.2 "Technical validation report".

4.7. Data Quality Tools

After archiving and before uploading objective data from a vehicle or a vulnerable road user's (VRU) device (e.g. smartphone), a well-defined algorithm should be applied to all the data in order to verify data consistency and validity and identify possible sources of errors.

Being able to query a 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.

It should be kept in mind that, very often, less data with higher quality is more useful than many data, since the complete data set can often not be fully analysed due to:

- Delays
- Missing data
- Poor data quality
- Budget restrictions
- Limited time
- Restricted access to gathered data in the database

Methods for automation of the analysis are needed in order to increase work efficiency, especially when processing the data. A set of scripts will check automatically a set of quality parameters once the raw data is received in the central server and before introducing this data in the central database. In case of failure detection, the data quality tools will generate a report indicating the list of errors found in the raw data received. This report will be sent to the deployment sites in order to detect possible malfunctions. This process is especially important in the first phase of the data collection activities.

A general data quality analysis is not only important for the project itself but in general since the data is to be re-used after the project and as the researchers might not have been involved in the data collection process and are not aware of the data quality issues. Data quality analysis starts from the FOT database and 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 the following four sub-steps reported below (and shown in Figure 13):

- Assessing and quantifying missing data (e.g. percentage of data actually collected compared to the potential total amount of data which it was possible to collect).
- Ensuring that data values are reasonable 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 dynamic 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 of standard deviation of lane offset, the lane offset measure should be at

least 10s long; additionally, this time length may depend on the sampling rate— see AIDE Deliverable 2.2.5, section 3.2.4 (Östlund et al., 2005)) [3].

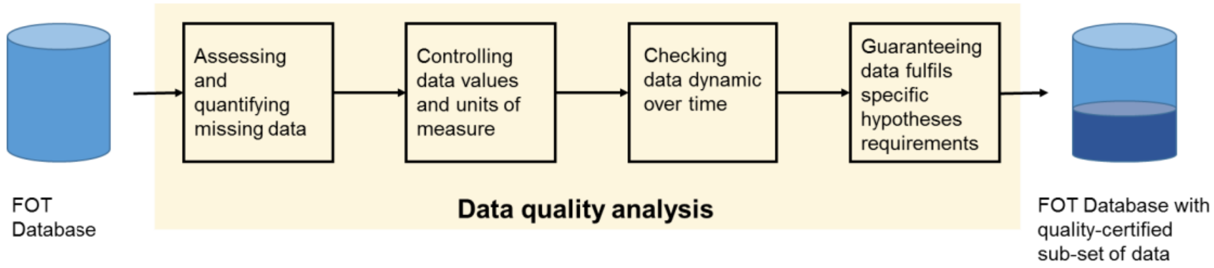

Figure 13: Block diagram of data quality analysis

4.8. Monitoring Tools

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.

Data logging systems from deployment sites should offer continuous monitoring capabilities to ensure the validity of data derived from sensors.

Monitoring tools display 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.

The central database from CTAG will host monitoring tools and provide a web-based interface to the monitoring teams of each deployment site. These monitoring tools will focus in a list of pre-defined monitoring indicators and will be running in the server hosting the central database. Once the data from the different deployment sites is sent to the central, and after performing the data quality analysis, these tools will include the data in a monitoring database that can be accessible from a web interface or through summary reports that can be sent using a “mail sending” tool.

The functionalities of the monitoring tools will be:

- User login
- Selection of time period:
  - Pre-sets: week, day
- Free selection of start / end
- Selection of indicators to monitor
- Export function (e.g. csv, pdf, image)

To avoid that deployment site monitoring personnel actively has to log in on the web-based tool, an automatic email system is also advised. An email would be periodically sent (e.g. daily or weekly) with the deployment site monitoring information of the previous time period.

4.9. Impact Assessment

4.9.1. Impact assessment and validation on the user and stakeholder level

4.9.1.1. User Behavior

Individual user behaviour is based on data collected during the FOT. It can be analysed in terms of safety, personal mobility and performance of private and professional drivers and vulnerable road users. These users' behaviour data can be used as input to traffic modelling in order to aggregate the individual user behaviour on traffic efficiency and environmental effects (see section 4.9.2.1). In terms of safety, personal mobility and performance on the individual users' level the following can be addressed:

- Safety: Some C-ITS services such as road hazard warning or blind spot detection warning aim to increase safety in traffic. Effects on traffic safety can be derived from analysing the frequency of crashes. Crashes describe any contact that one subject has with another moving or fixed object [13]. However, as crashes rarely occur, surrogate safety measures such as time to collision and evasive driving manoeuvres (e.g., deceleration, steering) can be suitable indicators to evaluate the safety criticality of driving situations [14].
- Personal mobility: Users 'mobility can also be assessed as some services such as mode & trip time advice aim to provide travellers with itinerary for multimodal passenger transport journeys. Mobility can be assessed from the individual travellers' or network perspective. The latter is described in section 4.9.3. From the individual travellers' perspective, mobility in terms of, for example, travel time can be analysed.
- Performance: Users' performance can also be assessed and reflects how well users adapted their behaviour to the recommendations and advice of C-ITS services. For example, for GLOSA; drivers' performance while approaching traffic lights can be assessed in terms of speed choice and acceleration/ deceleration patterns. This data can be used as input for upscaling the effect on the whole traffic flow (see section 4.9.3).
- Acceptance: Alongside an attitudinal component gathered by questionnaires, acceptance becomes manifest in user behaviour of a system. A simple way of operationalisation is to log frequency and duration of system usage and compares that for certain periods of time (e.g. by week) in order to assess changes. A constantly high or increasing usage indicates high acceptance, whereas a decrease suggests the opposite. A more sophisticated way of analysing the usage of an app could be to utilise the display-side camera of a smartphone for observing gaze behaviour of the user.

In addition to the users' behaviour data, a series of additional situational data (e.g., traffic and driving conditions) and users characteristics (e.g., demographic characteristics) will be identified that needs to be measured in order to provide key background information needed to derive and complement behaviour data.

4.9.2. User Acceptance Evaluation

User acceptance is the degree to which an individual intends to use a system and, when available, incorporates the system in her or his driving/ travelling (cf. [40]) . It can be assessed objectively via analysing the number of interactions with a system or service or subjectively with questionnaires. In C-MobILE, we are going to apply questionnaires to collect subjective data.

4.9.2.1. User Acceptance Models

In terms of assessing user acceptance subjectively an acceptance model is required to understand how user acceptance is formed. Acceptance models consist of factors (in other sources – constructs or variables) which may affect user acceptance. Most frequently used acceptance models are (followed by factors);

- Theory of Reasoned Action (TRA) → (ATB, SN)
- Technology Acceptance Model (TAM) → (EV, PU, PEOU, BI)
- Theory of Planned Behaviour (TPB) → (ATB, SN, PBC)
- Unified Theory of Acceptance and Use of Technology (UTAUT) → (PE, EE, SI, FC, EV)

In each of these models, technology acceptance is affected by several factors;

- Attitude Toward Behaviour (ATB): An individual's positive or negative feelings (evaluative affect) about performing the target behaviour [17]
- Subjective Norm (SN): The person's perception that most people who are important to him think he should or should not perform the behaviour in question [17]
- Perceived Usefulness (PU):The degree to which a person believes that using a particular system would enhance his or her job performance" [15]
- Perceived Ease of Use (PEOU): The degree to which a person believes that using a particular system would be free of effort" [15].
- Behavioural Intention (BI): The degree to which a person intents to use a system.
- Perceived Behavioural Control (PBC): The perceived ease or difficulty of performing the behaviour [17]

- Performance Expectancy (PE): The degree to which an individual believes that using the system will help him or her to attain gains in job performance" [41]
- Effort Expectancy (EE): The degree of ease associated with the use of the system" [41]
- Social Influence (SI): The degree to which an individual perceives that important others believe he or she should use the new system [41]
- Facilitating Conditions (FC): The degree to which an individual believes that an organizational and technical infrastructure exists to support use of the system" [41]
- External Variables (EV): EV are factors like age, gender, experiences and voluntariness which influence the relationship of two variables[41]

The comparison of the original definitions of the factors indicates that user acceptance models are based on previously developed theories of technology acceptance. For example, TAM is an adaptation of the TRA specifically tailored to model user acceptance of information systems [15] or UTAUT is a consolidation of the factors of eight models that earlier research had employed to explain information systems usage behaviour (including TRA, TAM and TBP) [41]. In studies about driver assistance systems (or intelligent transportation systems) researchers proposed additional factors to model driver acceptance. These are for example:

- Trust (T): The belief of drivers that the system would perform its intended task with high effectiveness [42]
- Perceived Safety (PS): The degree to which an individual believes that using a system will affect his or her well-being. [43]
- Affordability (A): Affordability is the monetary amount that drivers are willing to pay to purchase, install and maintain the system [45]

A review of empirical studies on user acceptance of driver assistance systems revealed that TAM is the most frequently used model [46]. The TAM has been applied successfully in multiple studies for evaluating the user acceptance for various systems and has proven to be a cost-efficient tool [15] [16]. However, to better understand the user acceptance toward driver assistance systems, additional factors need to be included to complement the existing construct in TAM.

Other EU funded projects like eCoMove and Drive C2X, which had comparable aims and objectives with C-MobILE used TAM as theoretical framework to assess user acceptance. Depending on deployed ITS services they added additional factors to the original TAM model. The factors of the theoretical models are measured by standard survey questions which are usually available on the original research paper. In the context of driver acceptance research, the questions are slightly modified to match the task and to avoid misinterpretation.

It is desirable to include as many factors as possible into the C-MobILE user acceptance framework. However, this is practically very difficult. It is envisioned that some questionnaires (for example ITS service and ITS bundle evaluation) will be filled by participants voluntarily in their free time in an uncontrolled environment. There is a typically low response rate in such questionnaires. Participants are tending to quit questionnaires if they have too many items. The preliminary analyses revealed that researchers used at least 4 (in average 5-6) items per factor. In order to get the full data set from each participant consistently, it is very important to select the number of factors and items carefully.

Similar to previous EU funded projects, C-MobILE user acceptance framework (see Figure 14) will be based on TAM. The factor perceived safety and trust was added to the model as proposed by [43]. In addition, the factor behavioural intention was enhanced by the factor affordability. Based on this theoretical framework, the user acceptance assessment will focus on the following dimensions; External Variables, Perceived Usefulness, Perceived Ease of Use, Perceived Safety and Trust and Behavioural Intention. This model will serve as a theoretical framework for the development of questionnaires.

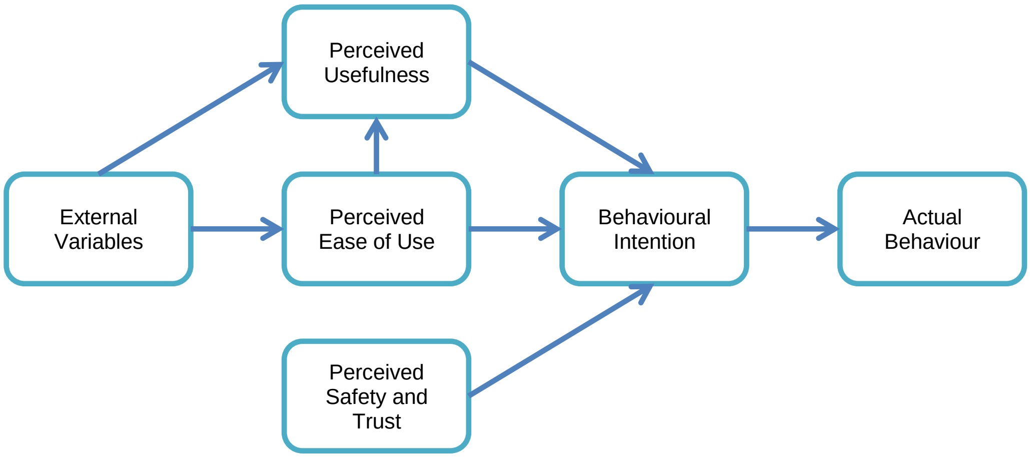

Figure 14. Estimating C-MobILE user acceptance framework

4.9.2.2. Questionnaires

Generally, the performance of a system is evaluated using qualitative KPIs, based on self-reported measurements. Self-reported measurements require that the participants express their opinion on the implemented services and bundles. Typically, these measurements are acquired from questionnaires, rating scales, interviews, focus groups, or other methods. In C-MobILE, qualitative KPIs are focused on the assessment of the user acceptance, the effects on the driving behaviour and the effects of C-ITS bundles. Questionnaires are designed to gather subjective data and obtain information and opinions of the different stakeholders and answer the subjective KPIs. Results can be analysed using tools like Excel, SPSS and/ or R.

In D2.2 a preliminary survey was conducted to get an insight into the requirements and expectation of various stakeholders. Results revealed three major stakeholder profiles. These were: end-users (drivers and vulnerable road users (pedestrians, cyclists), public authorities (cities, municipalities, traffic managers, road operators), and private companies (private industry consisting of C-ITS technology, service, or solution providers). In order to keep the questions relevant for each respondent three different questionnaires were designed to evaluate user acceptance;

- End-User Questionnaire: to be filled by drivers and VRUs who are participating in the C-MobILE project.
- Private Company Questionnaire: to be filled by the person of the company who makes purchasing decisions
- Public Authorities Questionnaire: to be filled by the person who makes purchasing decisions on the road side units (RSUs)
  - All these questionnaires will be provided in English and the local deployment sites will be in charge of translating them into the local language. The deployment sites will have paper hardcopies, an online version or by email at their disposal. As an alternative to physical paper copies, it is possible to convert the questionnaire template into an online equivalent using one of the many survey websites available (e.g. LimeSurvey, SurveyMonkey).
  - It is preferable to adopt online based systems for the electronic compilation in order to reduce costs and the overall manual work. It facilitates a wider circulation of the questionnaire and make the responses easier to retrieve, possibly in a suitable file format (.xls, .csv, etc.). However, an online version could potentially exclude some participants who may not have access to a (workplace) e-mail address, or not be willing to complete surveys online. Therefore, it is also recommended that hardcopies are also available as a back-up solution.

4.9.2.3. Execution of the Questionnaires

In this section the execution procedure of the questionnaires will be presented. First the content of the questionnaires will be listed then it will be described how and when the parts of the questionnaires will be conducted.

The End-User Questionnaire will consist of basic data collection, ITS service evaluation and ITS bundle evaluation. It is preferable that participants fill in the basic data questionnaire before they use the new ITS services which will be deployed within the C-MobILE project. However, this section is optional and respondents do not need to fill it in if they feel uncomfortable sharing their personal information. Towards the end of the basic data collection, participants will be asked if they have past experience with ITS services. Most of the deployment sites have already ITS services activated. Participants will indicate whether they have practical experience with ITS services. Participants who already have experience with activated services will fill the "ITS service evaluation" for each used service. In case that two or more services are categorised under a specific bundle, participants are going to fill the "ITS bundle evaluation". The first step will provide us a picture of the current situation at the deployment site.

It is likely that ITS services in C-MobILE will be released one by one. First "ITS service evaluation" will be conducted after service release, and if participants had at least one interaction with the specific ITS interface. Data will be collected for one month. Results of the first impression will be analysed. Based on the results, recommendations to optimise the interface will be sent to service providers. Providers will have a certain time to implement the HMI recommendations. After the implementation (or update) participants are going to use the service for another period of time. If participants had at least one interaction with the updated interface they are going to repeat the "ITS Service evaluation" and conduct if applicable the "ITS bundle evaluation" questionnaire. If there is no HMI improvement required participants are going to repeat the "ITS service evaluation" and fill the "ITS bundle evaluation" at least three months before the project ends.

4.9.2.4. Basic Data Collection (External Variables)

The subjective evaluation will start with the 'basic data collection'. The first four questions are about demographics, then five questions about user profile/role and driving habits. After that, if applicable, two questions will be about past experience with implemented ITS services. If participants indicate that they already have experience with implemented ITS services, they are going to rate the services as described in the "ITS service evaluation". The final question will be about use expectation from new ITS services. Items in the questionnaire are as followed:

- Age
- Gender
- Educational Level
- Occupation
- Blue Badge Holder
- Driving License / since
- Frequency of Mode Usage (per trip category)
- Average Mileage
- Knowledge about available ITS services
- Experience with ITS services
- Expectation from new ITS services
- Driving behaviour scale

4.9.2.5. ITS Service Evaluation

Participants will rate their experience on newly implemented ITS specific questionnaire, consisting of 14 closed questions (5-point Likert Scale). As described before the constructs of the questionnaire are perceived usefulness (PU), perceived ease of use (PEOU), perceived safety & trust (PST) and behavioural intention (BI). In the following section development of the ITS service questionnaire will be described. The intention was to find a good balance between general question and specific questions.

Researchers who created the user acceptance and behaviour models provided questionnaire items to test the constructs. A significant number of researchers used these standardized questionnaires in their studies, which has a crucial advantage of comparability with other studies. Questionnaire items for PEOU, PST and BI (10 out of 14 closed questions) are based on standardized questionnaires.

1. Perceived Usefulness (PU)

Users need to perceive the system as being useful or they will not attempt to use it regardless of how easy or difficult it is to use. PU questions evaluated usefulness with terms like productivity, performance, quality and efficiency. These are broad terms which can be interpreted by each participant in a different way. To avoid this, four specific question (awareness, task accomplishment, productivity and efficiency) were developed for each service. Use case descriptions in D2.2 and the research questions/hypothesis table supported the development of these questions. Summary and objective sections in the use cases are used to develop the awareness and task accomplishment questions. Expected benefits and expected impact descriptions were used to create the productivity and efficiency questions.

- Awareness

Most of ITS services which will be implemented during the C-MobILE project are designed to improve the awareness of road users (drivers and VRUs). They inform road users about the current state of the traffic situation and/or warn them about potential hazards. The first question will be about improved awareness. Here are some examples:

/ Road Hazard Warning enabled me to spot the location of the hazards quickly

/ Emergency Vehicle Warning enabled me to identify the position of the Emergency Vehicle quickly

/ GLOSA enabled me to identify the location of the signalized intersection

/ Flexible Infrastructure enabled me to identify which lane is activated quickly

- Task Accomplishment

Once the ITS service informed or warned road users about the current traffic situation or hazards respectively, usually it provides advice (primarily to improve safety or efficiency) to accomplish the task. The second question will be about how the service advice made it easier for the road user to accomplish the task. Here are some examples:

/ Urban parking availability made it easier to me to find a parking space that met my requirement

/ Signal Violation Warning aided me to stop safely before I made a red light violation

/ Green Priority made it easier to drive through congested intersections

/ MTTA made it easier to plan a journey

- Productivity

Each service is designed primarily to enhance safety, time efficiency or driving efficiency. Participants are going to rate the primary purpose of the service in this question. Here are some examples:

/ Road Works Warning made driving through a work zone safer

/ Warning system for pedestrians increased my safety and comfort in urban areas

/ Green Priority improved my punctuality

/ Urban Parking Availability reduced search/reservation time for a parking space.

- Efficiency

Primary objectives are usually not the only impact area of the service. In the last question the efficiency of the service will be considered in relation to another important variable. In example, Green Priority primary objective is to increase the punctuality of public transport vehicles. However, it is also important that drivers perceive enhanced fuel efficiency. Here are some additional examples:

/ I found the parking space always empty after reservation (UPA or MPA)

/ Emergency Vehicle Warning reduced my stress/hassle while giving way to the EV

/ GLOSA enhanced my fuel efficiency by avoiding unnecessary stops

/ CTLP reduced the number of dangerous situations with other road users

In order to keep the questionnaire as short as possible PU will be assessed by only two items of which one is always accounting for awareness and the other refers to one of the three concepts left in this factor.

2. Perceived Ease of Use (PEOU)

While driving it is very important to quickly perform tasks without lasting periods of trial and error. Interaction with C-ITS interface can be considered as a secondary task, which affects driving performance and safety. It is essential that a system is easy to use and the system interaction is clearly understandable. If intuitive design is neglected, user might reject the ITS service immediately. Individuals are likely to use a system if they believe it is easy to use and will increase their performance productivity. Systems that are difficult to understand are more likely to be turned off, which reduces the assistance effect. The following 4 items of the questionnaire will be more about the interface design of the ITS service.

An ITS interface should be intuitive and user should understand the purpose and the way of interaction immediately without much technical support. Therefore, this question will evaluate whether the interface of the particular service was easy to learn.

- Learning to use … was easy (5 point Likert Scale)

The majority of HMI of C-MobILE services will be presented on touchscreen interfaces. There is a known problem with recognition of touchable areas on such an interface. Since the interaction will occur in a non-stationary environment usually touchable areas are larger than interactive areas on a smartphone. This has the consequence that font and icon sizes are smaller than required, which may have an impact on readability. There are also other ways of interaction, such as light or voice signals which should be clear and understandable, as well.

- Interaction with … was clear and understandable (5-point Likert Scale)

C-Mobile ITS services will provide “push messages” to inform/warn road users about the current traffic situation. It is essential that users receive messages timely. Messages which are received too late would reduce the chance to adapt user behaviour properly. Messages which are received too early can be supressed by other service messages and users may have problems to adapt the desired driving behaviour.

- Information from … was received timely (Bipolar scale: -4 too early ←ideal→ +4 too late)

The amount of displayed information on the display is a significant factor for fast and precise interaction. Clutter of information on the display should be avoided. To achieve a better indication both item use bipolar rating scales.

- The amount of displayed information was … (Bipolar scale: -4 too less ←ideal→+4 too much)

Single services have limited features therefore it is easier to learn how to use one service than a bundle of services. That’s the reason why it is more appropriate to ask the question on bundle level. The last three items in this construct for ITS services are therefore suggested.

3. Perceived safety and trust (PST)

This part of the questionnaire will evaluate perceived safety and trust of users towards ITS services. The first two questions are about perceived safety while interacting with the interface and the last two questions are about trust in the technology.

As mentioned before interacting with ITS services has a secondary priority while driving. Especially, drivers may feel themselves in a dangerous situation if they have to look at the interface for a prolonged period of time to understand the messages, instruction or/and warnings.

- I can use … without looking at the display too much (5-point Likert scale)

The majority of C-Mobile services will send warning signals and instructions. Various feedback modalities can be used to pull the attention of the driver towards a new message/signal. Applying a feedback mode that is likely to be overseen by the users should be avoided. On the other hand, push messages should be not disruptive to the drivers.

- The signals generated by the … were (Bipolar scale: -4 not recognisable ←ideal→ very disruptive +4)

The primary aim of ITS services is to improve user safety and comfort. It would be controversial if users may feel that they are likely to be involved in a risky situation while interacting with the system while driving. In addition, drivers should not have the feeling that a misconduct with the interface will lead them to dangerous situations.

- I feel safe/confident while using …

Faulty or missing information would reduce the trust of users to the ITS service. Faulty message in this context is if a system sends a message or a signal indicating a potential threat, but there is actually nothing. Missing information would be the opposite situation where a potential threat is present, but the system is not warning the user. Reliability can also be influenced by other services. In example, the time to green time in GLOSA can be influenced by GP or CTLP.

- … works reliable and faultlessly (5-point Likert scale)

In order to keep the questionnaire items as short as possible we propose to apply “The signals generated by the … were” and “… works reliable and faultlessly” questions in the questionnaire.

4. Behavioural Intention (BI)

All the constructs or items of the questionnaire presented above may influence the behavioural intention of the end-user to buy and use the services or bundles. All user acceptance and behaviour models mentioned in the introduction presume that behavioural intention have a significant influence on actual use of the technology. Behavioural intention will be evaluated with two questions;

eCoMove project revealed that most of the participants would not like to spend more money on the system than they could save using the system. A more detailed questionnaire, like the Van Westendorp Price Sensitivity Meter (VWPSM), is not relevant for end-users because some mobile apps, such as google maps, provide different sort of traffic information for free.

- For an affordable price I would purchase/subscribe … (5-point Likert Scale)
- I would pay … Euros for this service

This is the last closed question where users will be asked whether they are going to use the service or not.

- I intend to use … regularly when I am driving (5-point Likert Scale)

It is more appropriate to ask this construct on bundle level since services cannot be purchased/subscribed separately nor customized by the user.

4.9.2.6. ITS Bundle Evaluation

In C-MobILE, ITS services will be consolidated to bundles, which will create a more comprehensive safety or efficiency system. There are two widely used system usability and acceptance scales in this domain; System Usability Scale and Van der Laan Acceptance Scale.

The System Usability Scale (SUS, citation count ca.6000) is a ten-item (5-point Likert Scale from strongly disagree to strongly agree) questionnaire administered to users for measuring the perceived ease of use of software and systems. This is a general applicable scale which is widely recognised as an industry standard scale, to return a usability value between 0-100. A 'good' SUS score is 68 or higher; the most usable systems will obtain a SUS score of 80.3. These are the SUS questionnaire items

- I think that I would like to use this application frequently.
- I found the system unnecessarily complex.
- I thought the system was easy to use.
- I think that I would need the support of a technical person to be able to use this system.
- I found the various functions in this system were well integrated.
- I thought there was too much inconsistency in this system.
- I would imagine that most people would learn to use this system very quickly.
- I found the system very cumbersome to use.
- I felt very confident using the system.
- I needed to learn a lot of things before I could get going with this system.

To calculate the SUS score:

- For odd numbered items: subtract one from the rating.
- For even-numbered items: subtract the rating from 5
- This will convert all values from 0 to 4 (with four being the most positive response).
- SUS score is calculated by multiplying the sum of the item values by 2.5

1. Open Questions

Closed ended questions are helpful to verify information and open ended questions provide valuable information, greater insights, and more understanding. End-users are able to provide insights about the HMI. Therefore, two open questions will be about HMI Design. The last question is reserved for final comments;

- Is there anything that you feel is missing on the application?
- If you were to make alterations to the app, what would be the first thing you would do to improve?
- Which of the following services did you like the most in the bundle?
- Do you have any other final comments about the … app?

4.9.2.7. Questions for Legal Authorities

Questionnaires for legal authorities were adopted from Compass4D. There will be 16 closed questions about the importance of the policy goals for purchasing decisions. Participants will rate the importance on a 3-point scale (low, medium and high). There will be 3 open questions regarding legal, organizational and technical/operational issues which could prevent system integration of C-Mobile services. The last question will be about whether they could imagine adopting C-MobILE services in its current form.

Impacts on the environment and fuel consumption:

C-MobILE services/bundles are expected to reduce the environmental impact of vehicles in urban areas with a reduction in fuel consumption, CO_2 emissions and other pollutants. The questions are; C-MobILE services…

- have a positive effect on the fuel consumption of fleet vehicles.
- reduce CO_2 emissions from vehicles
- have a positive influence on other pollutants (NOx, SOx).
- reduce noise levels from road transport.

These performance indicators focus on the effect of C-MobILE services on driver behaviour such as eco-friendly driving style. The questions are; C-MobILE services…

- promote eco- driving behaviour in terms of driving speed, acceleration and braking behaviour.
- are expected to improve the traffic efficiency and network performance.
- promote a safer driving behaviour in terms of their driving speed, acceleration and braking behaviour
- improve the network journey time
- improve journey time reliability on the network
- reduce delays on the network
- improve the network speed
- influence the traffic flow

Main indicators of traffic safety are the number of injuries and fatalities that have occurred in traffic accidents. The questions are; C-MobILE services…

- improve time headways
- reduce variations in speed
- reduce the number of hard braking events
- reduce the number of speeding events

Open questions regarding legal, organizational technical/operational and willingness to purchase questions are as followed:

- If the C-MobILE services/bundles were developed for the market in their current form, do you know of any, legal, organizational technical/operational barriers which would prevent the system from being adopted in your organisation/sector?
- Overall, if the C-MobILE system was developed for the market in its current form, would your organisation/sector consider adopting the system?

4.9.2.8. Objective User Acceptance Evaluation

As already mentioned before, user acceptance can not only be assessed by questionnaires but also by analysing user behaviour. In the simplest case described above this is done by logging system usage frequency and duration. Comparing these for certain periods of time (e.g. by week) allows to assess changes in user behaviour. A constantly high or increasing usage indicates high acceptance, whereas a decrease suggests the opposite. Unfortunately, this logic only applies for smartphone applications or similar services, where users have the choice of running it or not. If a service is offered on an OBU and/or cannot be switched off the method is obviously not telling anything about user behaviour, and thus acceptance.

Another way of evaluating user acceptance by objective measures is to look at compliance. If a user follows recommendations which she/ he is receiving it can be assumed that acceptance is high. This applies for instance if a driver adjusts speed according to GLOSA advice. On the other, hand measuring compliance is not always that straightforward. For example, a driver reducing speed when approaching a construction site might be doing that because of a road works warning message. However, alternative explanations such as obeying to traffic rules indicated by road signs or being forced to do so by dense traffic are plausible as well. Thus, the driver's behaviour might be compliant but is not necessarily linked to the service provided.

The aforementioned possibilities and constraints of assessing system acceptance from objective data suggest that both ways described should be deployed. While data about usage behaviour on mobile devices should be easily collectable through the app itself, compliance must be evaluated from data gathered for the impact assessment on the transport system level, which appears to be mostly GPS trajectories. The following table shows possible compliance criteria for most services, if applicable. For some services—especially the ones mainly offering convenience instead of safety advice (such as parking availability)—the concept of compliance doesn't fit. In every case a unique user ID assigned to each participant and dedicated to each dataset would be worthwhile in order to check for correlations between subjective and objective acceptance indicators.

Service	Compliance Criteria
RTM	Parked at the nearest occasion when rest time is reached
MPA	n.a.
UPA	n.a.
RWW	Deceleration Behaviour
RHW	Deceleration Behaviour
EVW	Deceleration/Acceleration Behaviour
SVW	Stop on request
WSP	Deceleration Behaviour / Stop on request
GP	Deceleration/Acceleration Behaviour
GLOSA	Deceleration/Acceleration Behaviour
CTLP	n.a.
FI	Higher mean cruise speed / smaller speed variation / Deceleration/Acceleration Behaviour
IVS	Speed deviation from limit
MTTA	n.a.
PVD	n.a.
EBL	Deceleration Behaviour
CACC	Smaller speed variation / Deceleration/Acceleration Behaviour
SSVW	Deceleration Behaviour
MAI	Deceleration Behaviour
BSD	Deceleration Behaviour

Table 11. Compliance criteria for services

4.9.3. Impact assessment and validation on the transport system level

One of the aims of C-Mobile is to determine the services'/ bundles' impact on the level of a greater driving population or area, thus, across users and non-users in the presence of the services' usage, in order to:

- provide evidence to decision makers on large-scale (larger number of services bundled, larger number of users, larger areas, longer usage time lapses) implementation effects of different bundles;
- provide expected disadvantages of the implementation that cannot be derived from individual measurements;
- provide information that can be used by municipalities to advertise the implementation;
- provide information for service/ infrastructure suppliers to advertise the implementation (in accordance with the ex-post CBA).

The validation on the transport system level will add upon the validation on the user level by generalising over users/ non-users and especially over areas or time frames and conditions, and moreover, by comparing these results with the targeted results on the system level. Different approaches to generalise the impact on individual users, locations and time frames will be employed, dependent on their availability for the deployment sites:

- baseline and treatment comparison of data collected and population effects at user (group) level before and during the service deployment (e.g. traffic count data/ volumes, travel times and average speeds on specific road sections, accident statistics etc.);
- demographic and spatial expansion of average effects for area of interests on a higher level (deployment site or city) by statistic extrapolation [26][33], transport models and traffic simulations in order to examine the interactions between the individual users' and non-users' behaviour

The main focus in C-MobILE lies upon the first and the real world impact the services unfold during the C-MobILE deployment phase.

Data on the individual level is assessed in a so-called *area of interest*. This is defined as the service and deployment site specific corridor/ route or region in which a specific service is active and receives and sends messages due to an incident, e.g. when passing a traffic light, a hazard occurring etc. These areas of interest can be dynamic and are further detailed in the technical reports related to each impact assessment and validation level. Within an area of interest, a specific incident point is situated, e.g. a stop line (a *control node*) or a hazard location (on a link or in a *control segment*). In contrast, the area of interest is different from an entire control segment as the latter constitutes an area where a vehicle is able to potentially receive messages related to the services, e.g. a whole district where a hazard warning could be given since a hazard appears.

Thus, the analysis in C-MobILE is event-based, meaning it relies on data collection that is triggered by an incident in the incident related area of interest. Mainly, data from system users is gathered no matter if the system actively provides the service's information or not (i.e. baseline vs. treatment conditions). The latter may serve as data a non-user would have produced. Collecting this proxy data is restricted to areas of interest in that both user types (information given vs. not given) are present at the same time. Dependent on local capabilities data on actual non-users can be observed and taken into account, e.g. roadside cameras. Alternatively, microscopic traffic simulation can show basic interactions between users (with information given) and non-users.

The evaluation on the system level requires large time lapses of collecting data. Ideally, the time period chosen should be large enough to enable the collection of a statistically relevant number of observations and to prevent seasonal and day-to-day variability from distorting the observed impacts. Further, it is imperatively necessary to obtain data that serves as a baseline to the bundle implementation. This data could be already assessed by the deployment sites or already simulated or has to be collected during a baseline study. All the real-life treatment data will be gathered during the deployment in C-MobILE and will constitute, on top of existent simulation data, the input for new estimations/ simulations. Baseline and treatment data will include:

- mobility related data
- safety related data
- efficiency related data
- environment related data

During C-MobILE the following research questions will be answered on this evaluation level:

- What impact do the services and bundles of services unfold in terms of safety, efficiency, sustainability and mobility behaviour?
- Does the usage of services and bundles of services take the intended and expected effect? In particular, do they increase safety, increase efficiency, decrease negative environmental impact and change mobility behaviour (in a specific amount)?

Based on the service descriptions in C-MobILE deliverable D2.2 and the stated expected impact, universal hypotheses were derived accordingly in order to answer these research questions.

In the following, a general overview of the impact assessment methodology is given. The C-MobILE deployment sites may adapt these approaches in order to better suit their local conditions, particularities and requirements when assessing the impact of the services.

4.9.3.1. Impact on mobility behaviour and accessibility

Following Litman (Last version 2011) [25], mobility and accessibility are two interconnected terms used to assess the quality of a transport system. Mobility measures the movement of persons and freight, using indicators such as the numbers of (car) trips per day, the distances travelled, the times spent in travel etc., whereas accessibility describes the ability of people and companies to reach the desired goods, services and activities.

The use of the service bundles that are implemented during C-MobILE will possibly have an impact on the mobility of both users and non-users, as well as on the overall accessibility. On the one hand, short-term (microscopic and local in terms of the Compass4D evaluation framework) positive effects for individuals (e.g. lower travel times, less time spent in traffic) may lead to rebound long-term effects on the system level, i.e. increased attractiveness (due to safer, more comfortable or more reliable experiences, or to lower costs) for car travel elicits more car traffic and therefore an increase in congestion. On the other hand, the use of C-ITS services has the potential to encourage multimodal travel behaviour, as users become more aware of alternatives to car travel.

The starting point for the impact assessment in C-MobILE will be the data on mobility patterns collected (subjectively) on an individual level before and during the FOT in the area of interest, pertaining to:

- frequency of car usage / number of car trips per day;
- average or total distances driven and time spent in traffic;
- usage of other modes of transport, including walking and cycling.

The related derived hypotheses are:

Hypotheses Index	Hypothesis	KPIs	PIs observed on individual level	KPI Success Criteria
HY-1	The use of C-MobILE service X changes car usage	Mean car usage percentage change across users	Frequency of car usage, Frequency of trip, Total / average number of trips	Significantly different from zero
		Total/ Average distance driven by car across users	Average/ Total trip length	Significantly lower
HY-2	The use of C-MobILE service X increases bicycle usage	Mean bicycle usage percentage change across users	Frequency of bicycle usage, Frequency of trip; Total or average number of trips	Significantly higher than zero

Table 12. Hypotheses addressing mobility behaviour

The results on the individual level will pertain to a changed share in the mode usage on all of the defined trips one user undertakes. The trips can be defined as one specific recurring route, e.g. a commute, or all trips.

These changes are firstly averaged to estimate the average user (group) population effect of the service including various situation/ control variables, e.g. user group, day time etc. Service activation and compliance rates will be noticed.

Changes in an appropriate number of individual behaviours could then be used to upscale the results and to determine the expected impact on the overall population of the specific area of interest in the deployment site including various situational/ control variables, e.g. activation rate, compliance rate, trip categories, day time etc. By integrating secondary data on mobility behaviour possible carry-over effects and effect weights will be noticed.

Level of upscaling the results	Specific trip	All Trips
Individual	Percentage difference for (frequency of car usage/ frequency of specific trip) → Individual effect for service on trip category mobility behaviour	Percentage difference for (frequency of car usage/ frequency of all trips) including carry-over effects → Individual effect for service on mobility behaviour
Users	Mean percentage difference across users for specific trip → Average effect for service on trip category mobility behaviour among users	Mean percentage difference across users for all trips including possible carry-over effects → Average effect for service on mobility behaviour among users
Users + Non-Users	Mean percentage difference across travellers for specific trip	Mean percentage difference across travellers for all trips including possible carry-over effects

Table 13. Example for upscaling levels (grey means not of interest)

Example for Mode and Trip Time Advice:

One of the hypotheses used to test the impact of MTTA is that this service will change the users' car usage patterns. For each user, the PIs mentioned in Table 13 (hypotheses addressing mobility behaviour) will be used to determine if due to the availability of the MTTA service the user has increased or decreased her/ his number of trips per day, average trip length, car usage frequency, etc. Aggregating the effects over all MTTA users will result in the mean percentage difference of all MTTA users. An additional step of the upscaling procedure could involve transferring the results to the entire population of the city/ region in a fictional scenario where everyone is a MTTA user, in order to assess the maximum benefits which can be achieved by this service. In the case of MTTA, estimating the changes in car usage for the entire population could easily be achieved by multiplying the mean percentage difference across users with the total population size, assuming that current MTTA users are not a special group prone to reacting differently to the service due to external conditions (e.g. users with very good public transport accessibility, young and technology-oriented persons, etc.). If the demographic attributes (age, employment status, etc.) of MTTA users are known, then separate mean percentage differences for separate user groups could be estimated and scaled up to match the overall population and thus provide a more realistic result of the upscaling procedure. Additionally, usage attributes (activation rate, compliance rate) are used to upscale the results according to real life usage of the service.

In case either the travel conditions (travel times, costs, etc.) or the travel patterns exhibited by the test persons significantly change after the introduction of the C-ITS services, running a transport demand model might be necessary to evaluate the impact on traffic and accessibility in the entire study area even more elaborately.

A preliminary table with all hypotheses derived from C-MobILE Deliverable D2.2 "Analysis and Determination of Use Cases" allocated to services and cities can be found in the Annex 2 Table 21. An individual description of every test scenario and available factors for scaling up the results in a specific deployment site are documented in the reports D6.3 "Report on impact of C-ITS on stakeholders and end-users " and D6.4 "Report on impact of C-ITS on surface transport system".

4.9.3.2. Impact on safety

Safety refers to the users of service bundles implemented during C-MobILE, to affected non-users, i.e. other drivers and other Vulnerable Road Users, as well as to other entities, e.g. road work units. Generally, it could be objectively assessed by examining the number of accidents, collisions, crashes, fatalities or safety related incidents per subject or per area and time unit. As accidents, collisions or crashes are very rare events and the time available for monitoring during deployment is limited, other indicators are necessary to evaluate the safety level. For example, the occurrence and frequencies of dangerous or last moment manoeuvres could be assessed, i.e. too fast approaching, hard braking or steering. Alertness related behaviour can be explored, i.e. the distance in time or pathway of safety behaviour like braking to an incident point [33]. Further, the frequency of rule violations could serve as mirror for safety, i.e. violated speed limits, signals or lane keeping. Furthermore, speed data can be used to assess the risk of accidents, and their severity respectively. The risk and seriousness of an accident increase with higher speeds driven [34][35] and may increase with higher differences between involved actors' speeds driven. Averaging the subjectively assessed perceived safety indeed does not validly reflect safety on the system level as the perception of risk in traffic is highly distorted but could be a factor when it comes to changes in the overall mobility behaviour (see 4.9.3.1).

Every C-MobILE service (integrated in C-MobILE bundles) is intended to be evaluated in terms of safety impacts.

The related hypotheses pertain to the following and every performance indicator is meant to be observed per driver, per time and **within the area of interest**, e.g. approaching and passing an intersection or hazard, key performance indicators refer to a group of users within the area of interest:

Hypotheses Index	Hypothesis	KPIs	PIs observed on individual level	KPI Success Criteria
HY-3	The use of C-Mobile service X increases drivers compliance with driving/resting times	Total number of driving/resting time violations	Resting time violation	Significantly lower
HY-4	The use of C-Mobile service X reduces the risk of accidents	(Weighted) Mean of mean speeds	Mean speed, spot speed	Significantly lower (alternatively: remains the same if variance is lower)
		Pooled speed variance	Speed variance, Mean speed, spot speed	Significantly lower
HY-5	The use of C-Mobile service X reduces the number of collisions/ accidents (with sb./sth.)	Total number of collisions/ accidents		Significantly lower
HY-6	The use of C-Mobile service X reduces the severity of accidents	Total number of severe accidents		Significantly lower
		Mean of mean speeds	Mean speed, spot speed	Significantly lower
HY-7	The use of C-Mobile service X reduces red light violations	Total number of red light violations	Red light violation per traffic light	Significantly lower
HY-8	The use of C-Mobile service X does not change the number of red light violations	Total number of red light violations	Red light violation per equipped user or traffic light	Remains the same
HY-9	The use of C-Mobile service X increases the distance (time and path) of braking events to an incident point	Average distance between braking event and incident point	Distance between braking event and incident point, braking event, time	Significantly higher
HY-10	The use of C-Mobile service X reduces hard braking events	Total number of hard braking events	Hard braking event	Significantly lower
		Mean (maximum) deceleration	Maximum deceleration	Significantly lower
HY-11	The use of C-Mobile service X reduces speed limit violations	Mean percentage of speed violators	Speed violation	Significantly lower

		Total number of speed violations	Speed violation	Significantly lower
		Average time/ distance of speed violations	Spot speed, speed violation, time, number of speed violation measurement points	Significantly lower
HY-12	The use of C-Mobile service X does not elicit lane violations	Total number of lane violations	Lane violation per network element	Equals zero

Table 14. Hypotheses addressing safety (opposing hypotheses address different services)

The impacts on safety will be assessed by quantifying and averaging the increase or decrease of safety related indicators at individual level.

Depending on the services being assessed and the specific areas they cover (e.g. intersections, traffic signals, etc.), the aggregation on population level will occur either by scaling up the effects per number of persons, per network element, per incident category or a combination of them including situational/ control variables, e.g. activation and compliance rates etc. Simulations and accident prediction models, if available, could also be employed to estimate the relationship between driving patterns (e.g. speeding, signal violations, etc.) and the frequency of accidents.

Example for Road Hazard Warning:

This service informs users about certain abnormal traffic situations downstream. As a result of this information, the users are expected to reduce their speed and drive more carefully. This should reduce the number of sudden braking manoeuvres and also the average speed and speed variance between the point where the driver receives the warning and the hazardous location. These PIs can be observed for each user at an individual level. After comparing the behaviour of users/ non-users, the KPIs mean reduction of hard braking events, mean reduction of the average speed and speed variance can be derived. These KPIs describe the impact of RHW on the safety of the users. By means of statistical models [35] these safety improvements can be converted into the number of avoided injury accidents.

A preliminary table with all hypotheses derived from C-MobILE Deliverable D2.2 “Analysis and Determination of Use Cases” allocated to services and cities can be found in the Annex 2 Table 21. An individual description of every test scenario and available factors for scaling up the results in a specific deployment site are documented in the reports D6.3 “Report on impact of C-ITS on stakeholders and end-users ” and D6.4 “Report on impact of C-ITS on surface transport system”.

4.9.3.3. Impact on efficiency

Efficiency refers to the performance of individuals (C-ITS users and affected non-users within a wider area or during a longer time period), but also to the performance of the traffic network at a deployment site. For individuals, the energy (fuel) efficiency is an important indicator as well as the time and distance it takes them to reach a certain destination, e.g. the right parking lot. The infrastructure efficiency can be measured e.g. in throughput, uptime (congestion-free time), alternatively also on an individual level, e.g. speeds driven, travel time etc.

In order to assess C-ITS induced improvements in fuel efficiency, comprehensive data on the test vehicles will be required. Ideally, C-MobILE users will keep a record of all refuelling activities during the test period, thus enabling accurate fuel consumption rates (in l per km) to be calculated. Alternatively, fuel consumption can be modelled as function of speed and acceleration data derived from GPS or on-board units [37]. As the fuel economy is influenced by types of routes driven (urban, rural, with or without intersections, etc.) one of the main challenges will be to ensure comparability of results from the baseline and the treatment scenario. If available, a microsimulation of a well-defined situation (e.g. crossing the same intersection), calibrated with real in-vehicle data (speed, acceleration, braking, etc.) from the deployment, might enhance the validity of the results. The resulting impact on energy efficiency will be measured in the decrease of average fuel consumption per km.

The flow of the vehicles is assessed by road site units (cameras or induction loops) that can provide data like traffic counts or congestion detected. Proxy indicators for flow are also average speeds, acceleration and especially speed variances. Pooled variances from users on a link, for example, decrease with a better flow, as average acceleration does.

The related hypotheses pertain to the following and every performance indicator is meant to be observed per driver, per time and **within the area of interest**, e.g. approaching and passing an intersection, key performance indicators refer to a group of users within the area of interest:

Hypotheses Index	Hypothesis	KPIs	PIs observed on individual level	KPI Success Criteria
HY-13	The use of C-Mobile service X reduces parking space search time	Average time spent for parking		Significantly lower
HY-14	The use of C-Mobile service X optimizes the flow of vehicles	(Weighted) Mean of mean speeds	Mean speed	Significantly higher
		Pooled variance	Speed variance	Significantly smaller
		(Weighted) Mean of mean acceleration/deceleration	Mean acceleration	Significantly lower
		Total/ average number of stops	Stop count (stop defined as spot speed(t) < 2 km/h and spot speed (t-1 sec) > 2 km/h	Significantly lower
		Total/ average number of double stops	Number of double stops	Significantly lower
		Total number of braking events	Number of braking events	Significantly lower
		Average number of vehicles passing point per time lapse	Number of vehicles per time lapse	Significantly higher
HY-15	The use of C-Mobile service X reduces vehicle-kilometres driven	Total vehicle-kilometres driven	Vehicle-kilometres driven	Significantly lower
HY-16	The use of C-Mobile service X increases punctuality	Departure time deviation from scheduled departure time	Departure time at every station, time schedule	Equals zero
HY-17	The use of C-Mobile service X reduces travel time (EV response times, service times)	Average travel time	Travel time, mean speed, route length, time stamp	Significantly lower
HY-18	The use of C-Mobile service X reduces parking space demand	Average parking space occupied	Number of spaces occupied, number of spaces	Significantly lower
HY-19	The use of C-Mobile service X reduces VRUs' waiting time	Average waiting time	Waiting time	Significantly lower
HY-20	The use of C-Mobile service X increases travel times of cars	Average travel time	Travel time, mean speed, route length, time stamp	Significantly higher
HY-21	The use of C-Mobile service X reduces energy consumption (fuel, kW)	Total/Average energy consumed	Energy consumed, distance driven	Significantly lower

HY-22	The use of C-Mobile service X improves the performance of traffic lights	Average queue length	Queue length per traffic light phase	Significantly lower
		Average number of vehicles passing per traffic light phase	Number of vehicles per traffic light phase	Significantly higher
HY-23	The use of C-Mobile service X reduces lost time	Mean percentage time with speeds < 2 km/h	Time speed < 2 km/h, travel time	Significantly lower

Table 15. Hypotheses addressing efficiency (opposing hypotheses address different services)

The efficiency of network elements and deployment sites will be assessed using aggregated data collected on-site for important network elements. Key indicators hereby are traffic volumes per timespan (15-minute intervals, peak hour, average working day, week, etc.) and per vehicle type (cars, HGVs, etc.). Other important indicators are the frequency, duration and length of congestions, frequency of accidents etc. Improvements in the infrastructure efficiency will be measured by indices or percentage changes compared to the baseline scenario.

Example for GLOSA:

GLOSA aims at reducing the number of unnecessary stops at traffic signals. By reducing the number of accelerations, it intends to improve vehicle (fuel) efficiency. GPS data will be collected from the users' vehicles and will be used to calculate individual PIs, such as number of stops, speed variance and total acceleration when approaching an intersection, for each driver. Aggregating over users and non-users will deliver the KPIs average number of stops and total acceleration with/without GLOSA. Using the acceleration profiles and a simple fuel consumption model will provide the efficiency gain in terms of fuel saved for GLOSA users versus non-users (see also next section).

A preliminary table with all hypotheses derived from C-MobILE Deliverable D2.2 "Analysis and Determination of Use Cases" allocated to services and cities can be found in the Annex 2 Table 21. An individual description of every test scenario and available factors for scaling up the results in a specific deployment site are documented in the reports D6.3 "Report on impact of C-ITS on stakeholders and end-users " and D6.4 "Report on impact of C-ITS on surface transport system".

4.9.3.4. Impact on a sustainable environment

The impact on the environment refers to traffic-induced effects (e.g. emissions of pollutants, carbon dioxide, noise, etc.) on society (including non-travelling persons, residents etc.) and nature. There is a close connection between the quantity of travel and mobility, its efficiency and the overall environmental impact. Traffic-related emissions are roughly considered to occur proportionally to distance driven, with the vehicle (energetic) efficiency and emissions per driven km being the correlating factor.

The use cases in C-MobILE are not expected to have an impact on the environmental characteristics of the vehicles itself. However, an increase in the vehicle operation efficiency (see 4.9.3.3) will also have positive effects on the environmental impacts. In the simplest assessment, a reduction of fuel consumption by x% will lead to a reduction of negative environmental effects per km by the same x%. This applies to CO_2 as it is highly correlated with fuel consumption. If available, microsimulation models with an environmental component could be used to enhance the accuracy of this analysis, e.g. EnViVer.

Similar to section 4.9.3.3 (Impact on efficiency), the vehicle speed and acceleration profiles can be derived from GPS data. Using an instantaneous emissions model [37] and information on the vehicle characteristics (propulsion system, size, age, European pollution norm, etc.), the instantaneous emissions at each point in time t can be estimated with the following generic model:

$$E(t) = f(v(t), a(t), v^2(t), a^2(t), \ldots)$$

with

E(t): Instantaneous emissions (e.g. CO_2) at point in time t

v(t): Instantaneous (spot) speed at point in time t

a(t): Instantaneous (spot) acceleration at point in time t

A similar model can be used to evaluate noise, consisting of rolling noise and propulsion noise, which are also a function of speed and acceleration and various control variables [38].

The related hypotheses pertain to the following and every performance indicator is meant to be observed per driver, per time and **within the area of interest**, e.g. approaching and passing an intersection, key performance indicators refer to a group of users within the area of interest:

Hypotheses Index	Hypothesis	KPIs	PIs observed on individual level	KPI Success Criteria
HY-24	The use of C-Mobile service X reduces pollution/emissions	Average CO_2 emitted	CO_2 emission	Significantly lower
		Average noise emitted	Noise emission	Significantly lower
		Number of noise emissions over a threshold of tbd	Noise emission	Significantly lower
		Average peak noise emission	Noise emission	Significantly lower

Table 16. Hypotheses addressing environmental aspects

When considering the overall traffic-related impact on the environment, the positive effects from the reduced emission rates per km driven have to be offset by a potential increase in mobility, as discussed in 4.9.3.1. This means that, even though the environmental impacts per km are reduced, an increase of the overall impacts is still possible when multiplying with the total number of km driven. The relation between a more efficient and better performing transport system and its rebounds for the environment has to be assessed carefully. Overall impacts on the environment will be measured in absolute value differences and percentage changes compared to the baseline scenario.

Example for GLOSA:

As mentioned in the previous section, GLOSA aims at reducing the number of unnecessary stops when approaching a traffic signal. This should not only enhance the energy efficiency of the vehicles, but also reduce their emissions. The acceleration and speed profiles derived from the individual GPS data can be aggregated over all observed users and non-users and then be used to derive average vehicle emissions using the instantaneous emissions model shown above and information on the average vehicle fleet in the respective deployment site. The total reduction in emissions can finally be calculated multiplying the average emissions reduction per vehicle passing with the number of users/ occurrences observed (assuming that data has been collected for each user). For upscaling the results, transport performance, traffic load and vehicle type compositions can be used for a specific area.

A preliminary table with all hypotheses derived from C-MobILE Deliverable D2.2 “Analysis and Determination of Use Cases” allocated to services and cities can be found in the Annex 2 Table 21. An individual description of every test scenario and available factors for scaling up the results in a specific deployment site are documented in the reports D6.3 “Report on impact of C-ITS on stakeholders and end-users ” and D6.4 “Report on impact of C-ITS on surface transport system”.

A full list of performance indicators and situational and control variables measured on individual level can be found in the Annex 2 in Table 18 and Table 20.

4.9.3.5. Bundling services

In C-MobILE different services are bundled in one wrap-application. Indeed, it is intended to evaluate the services separately, but bundling them might cause and promote interactions between services. The detection of (disrupting) interactions can be tackled by deploying the services separately and later in combination. This has to be noted in the deployment design. Alternatively, when two services are active and send information at the same time, these cases can be discarded from the analysis of a single service and the potential of interactions can be derived from the share of occurrences of these cases in the time the application is used. If a reasonable amount of data on interactions is collected, they can be included in one statistical model.

Since two or more independent services in a bundle have the same key performance indicator as dependent variable, weighted total average effects of this bundle might be carried out. An illustrative example (without weights) is given in Figure 15. Effects can be weighted in terms of, for example, network coverage or incident rates.

Mechanism	Effect on injury accidents	Effect, % (EC)		Coefficient of efficiency (EF)
Mechanism 1	decreases	–30%	→	0.70
Mechanism 3	decreases	–0.5%	→	0.995
Mechanism 5	increases	+12%	→	1.12
Total average effect		**–22%**	←	0.70 x 0.995 x 1.12 = 0.78

Figure 15. Estimating total average effects. Mechanism equals C-MobILE service [26]

4.9.3.6. Power analyses for validation

In order to estimate the amount of a statistically relevant number of observations or validate the appropriateness of the sample sizes for every comparison between baseline and treatment data sets, power analyses can be applied.

For a reliable estimation of expected differences in means the following equations apply:

$$n_b = (\sigma_1^2 + \sigma_2^2/K)(1.96+0.84)/\Delta^2$$

$$n_t = K(n_b)$$

with:

n_b = sample size for baseline group

σ_1, σ_2 = standard deviations of mean 1 in the baseline and mean 2 in the treatment

K = ratio of sample size for treatment group to baseline group

$\Delta = |\mu_2-\mu_1|$ = expected absolute difference between two means derived from the percentage of expected change

n_t = sample size for treatment group

assuming false positive Type I errors' α-level=0.05 and false negative Type II errors' β-level=0.8

An example can be given based on expectations stated in the C-MobILE Proposal Part B (2016, 2.1, p.16):

A reliable decrease or increase of the average speed of tracked users in an area of interest of 3% under the hypothetical assumptions that the current or baseline average speed there is 30 km/h with a standard deviation of 10 km/h and the constraint that the treatment group is 3 times larger than the baseline group of obtained data points implies the requirement of 1048 obtained baseline data points/sets and 3142 treatment data points/sets. A reliable decrease or increase of 10% requires a baseline sample size of 117 data points/sets and a treatment sample size of 351 data points/sets under same basic assumptions.

The size of an adequate sample size decreases with a lower ratio for baseline to treatment data points, with smaller standard deviations (in relation to the mean) in the baseline group and with a bigger effect, thus, with a higher decrease or increase. Standard deviations of route sections can be obtained during a baseline period or from traffic management centres or other sources of secondary data.

A reliable estimation of a percent reduction in the absolute numbers and proportions of incidents, e.g. number of rear-end collisions based on obtained data, is not feasible within the project. Whereas, a reliable change in the proportions among users of braking hardly related to an event or using a bicycle for commuting might be possible.

For a reliable estimation of expected differences in these proportions the following equations apply [36] :

$n_b = (1.96+0.84)^2(p_1(1-p_1)+p_2(1-p_2)) / (p_1-p_2)^2$

$n_t = n_b$

with:

n_b = sample size for baseline group

n_t = sample size for treatment group

p_1 and p_2: (expected) proportions in the two samples

assuming false positive Type I errors' α-level=0.05 and false negative Type II errors' β-level=0.8

An example can be given. A reliable indication for the doubling of averagely using a bicycle in 20% of daily commutes (meaning obtained 40% in the treatment period) requires 81 users in the baseline and 81 users in the treatment period under the constraint that the same number of users is observed in a baseline and a treatment period.

The amount of data needed increases with a lower difference of p_1 and p_2.

Finally, for upscaling the results, true population means and proportions should lie within an appropriate confidence interval.

An example based on the previous examples can be given:

For an appropriately (in terms of considering the population level) obtained mean of 30 km/h, meaning the confidence interval does not exceed 2.5%, it needs 700 data points under the basic assumption that the population consists of at least 100000 drivers, the standard deviation is 10 km/h and the confidence level equals 95%. 400 data points suffice to appropriately assess a mean of 31 km/h with a standard deviation of 5 km/h and the same basic population and confidence level. The latter does not necessarily mean that this amount is enough to detect significant differences in the means of 30 km/h and 31 km/h if both were obtained in the two periods baseline and treatment and are compared, see the first paragraph of this subsection.

The amount of data needed increases with higher standard deviations and lower means (keeping standard deviations constant), higher confidence levels. Whereas it remains constant when the population is even bigger.

The maximum number to obtain appropriate proportions for an error margin of 2.5% and a basic population of at least 100000 drivers and a confidence level of 95% is 1515 user data points. This number decreases with deviations of the expected proportion from 50%.

4.9.3.7. Deployment designs

Large-scale demonstration during C-MobILE differs from pilot studies regarding the controllability of drawing samples and determining baseline and treatment periods for a limited number of participants. Although a within-design with repeated measures of the same driver in the same situations over a long period of time is still the preferred way of collecting data, this might not be realistic in a real-life large-scale deployment. Further, the prevalence of incidents some services address and aim to support is not high enough to collect enough data. Thus, the preset approach is to divide up the time periods of collecting baseline and treatment data for a specific service and to consider the data points as being independent, e.g. the first 5 months of deployment every user that has downloaded the app is not provided with information of a service and after that every current user and every new user is provided with information of this service. A more elaborated approach (for services with higher incident prevalence) is to trigger the baseline and treatment period dependent on the time of downloading the app and individually specific, e.g. the first 2 months after downloading the app independent of when the app was downloaded a user is not provided with information of a service and after that this user is. Both phases can change periodically or even based on a probability function and on a daily basis.

Due to implementing different services (and implementations being different) in every deployment site, the C-MobILE deployment sites are going to carry out different approaches to draw samples.

4.9.3.8. General remarks on up-scaling and hypothetical scenarios

The methods to assess the impact of the C-MobILE services presented in the previous sections attempt to identify and quantify the real-world impact of the services as they have been implemented and operated during the deployment phase. This means that the results only reflect the benefits at transport system level for those specific locations (road segments, intersections, etc.) where services have been deployed and only for the specific proportion of users/non-users which happened to drive in those locations during the deployment phase. This is in line with the overall objective of the project, which is to demonstrate the real-world benefits of large-scale deployment of C-ITS services and not to conduct tests and simulations in a controlled environment.

For the decision makers and local authorities, the questions of transferability and scaling up of benefits to larger regions and increased user numbers might nevertheless be of relevance. An example for such a scaling up of benefits would be to answer the hypothetical question of how much larger the benefits would be if the proportion of drivers using a network segment/intersection would double. The trivial solution, in this case the assumption that the benefits would also double, is not necessarily the correct one as the interaction between users and non-users might be more complex and thus the impact would not be linearly correlated to the number of users.

Generally speaking, the scaling up of results can be grouped into two main categories:

a) geographical transferability

b) user transferability

Category a) enables decision makers to estimate the impact of the services if they were also deployed in other locations. Here, the assumption that similar results are to be expected on similar road segments and intersections as with the ones observed in C-MobILE can lead to plausible results, provided the similarity of locations (e.g. same number of lanes on the road section, same geometry intersection, etc.) is ensured.

Category b) considers what would happen if the number/proportion of C-ITS users in traffic would change. This hypothetical scenario is more difficult to assess, as it requires an understanding of the possible interactions between users and non-users. For example, if a certain proportion of drivers benefit from a warning system (e.g. Road Hazards Warning) and accordingly slow down upstream of a hazard, non-users might also be affected if the involuntarily have to slow down and thus might also benefit in terms of traffic safety. If the proportion of users reaches a certain high level, then any additional users on top of that will probably not have any impact at all. To understand and consider these types of interactions, microscopic simulations of the traffic flow are required. When available for the C-MobILE deployment sites, such models should be used for these kinds of hypothetical scenarios.

4.9.4. Evaluation While Doing

In the C-MobILE project, about twenty services will be tested on eight different pilot sites throughout Europe. Because of this large variety (and the fact that probably not all services – and pilot site environments and circumstances – will be exactly the same on all pilot sites) it is desirable to create a methodology that can work with these differences while still delivering results that can be compared between services and pilot sites. This methodology will provide a generic approach that models the main KPI's while allowing and enabling specific evaluators to integrate their individual characteristics, aspects and needs.

Evaluation while doing (EWD) aims to enable evaluation to take place earlier in the process and in a more efficient way (by accelerating data collection, processing and analysis through automation). This will speed up the evaluation process. It coordinates with the data management activities related to data storage, such as data quality tools, described in Section 4.7 and 4.9.4.1.

EWD is based on the FESTA-V. A methodologically sound approach to setting up and evaluating Field Operational Tests (FOTs) and pilots was developed in the FESTA (Field opErational teSt supporT Action) project and updated in 2016 [1]. Evaluation includes both technical evaluation and impact assessment in the areas of safety, traffic efficiency, environment, acceptance and socio-economic cost-benefit analysis. Tests can take different forms, as shown in Figure 16.

FESTA defined the "FESTA-V" (also called "FOT-chain") of evaluation steps based on the FESTA methodology. Figure 16 shows the FESTA-V. The left side of the "V" describes the steps taken to prepare the evaluation of the FOT. The right side of the "V" describes the steps taken to carry out the analysis of the FOT. The left and right-hand sides of the V at the same horizontal level are connected to each other. That is, the preparation of the research questions and hypotheses has a direct link with (and impact on) the analysis of the research questions and testing of hypotheses. An important contribution of the FESTA approach is that it made the explicit link between (1) what has to be evaluated and (2) the preparation of the FOT. A consequence of this link is that evaluation issues were brought to the discussion earlier in the execution of the FOT, in theory guaranteeing that the data necessary for impact assessment is collected during the FOT.

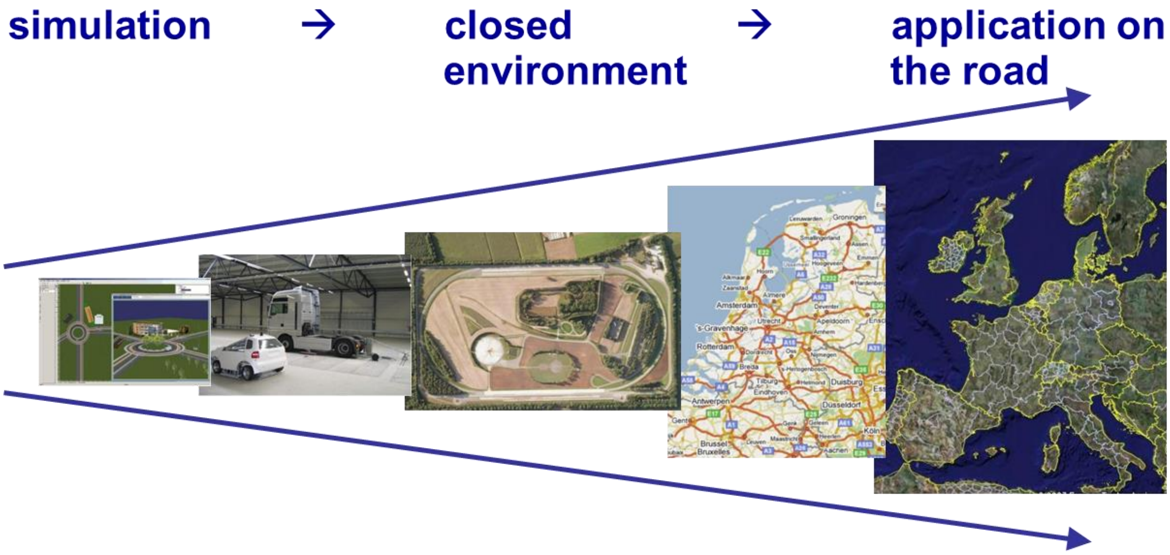

Figure 16: Testing and research environment for ITS

This document is structured as follows. First the context into which the methodology will fit is sketched. The next section presents what the methodology should ideally enable. The examples based on a service follows.

4.9.4.1. Context

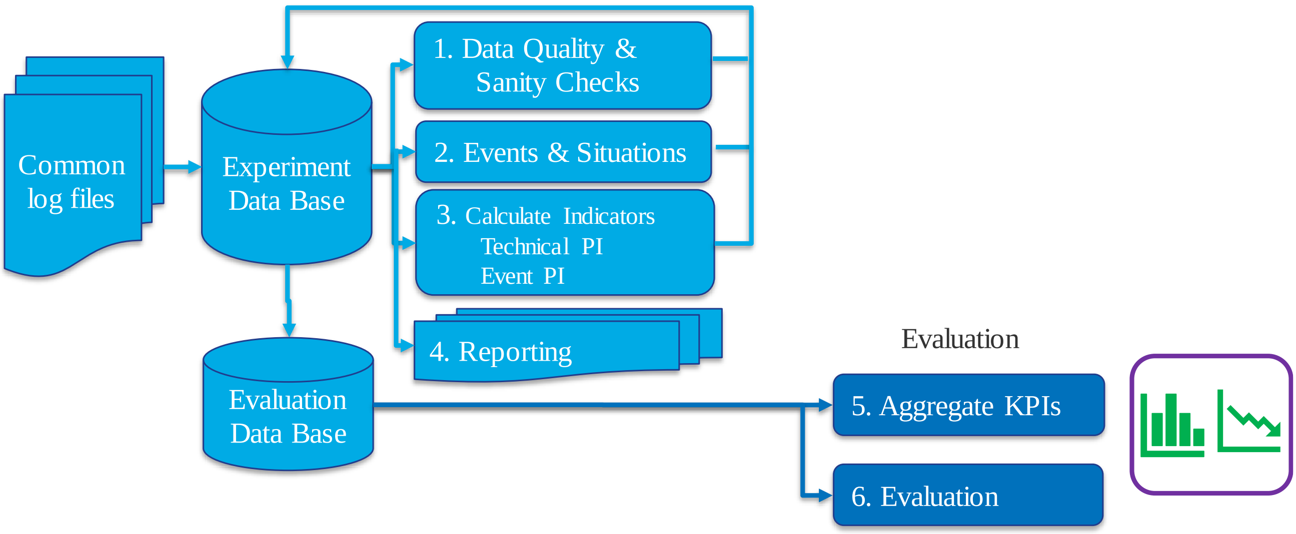

Figure 17 visualizes how the methodology fits into the data collection and processing steps. Data from the deployment sites are collected from various sources and subsequently standardized in a 'common data format' such as the one developed in InterCor, depersonalized and subsequently uploaded into the database, managed by CTAG. The format of the database(s) will be specified in the project and implemented by CTAG. The evaluation toolbox then uses the algorithms developed in this project to produce some of the Performance Indicators (PIs) needed for hypothesis testing and analysis. The PI's will be centrally stored as well.

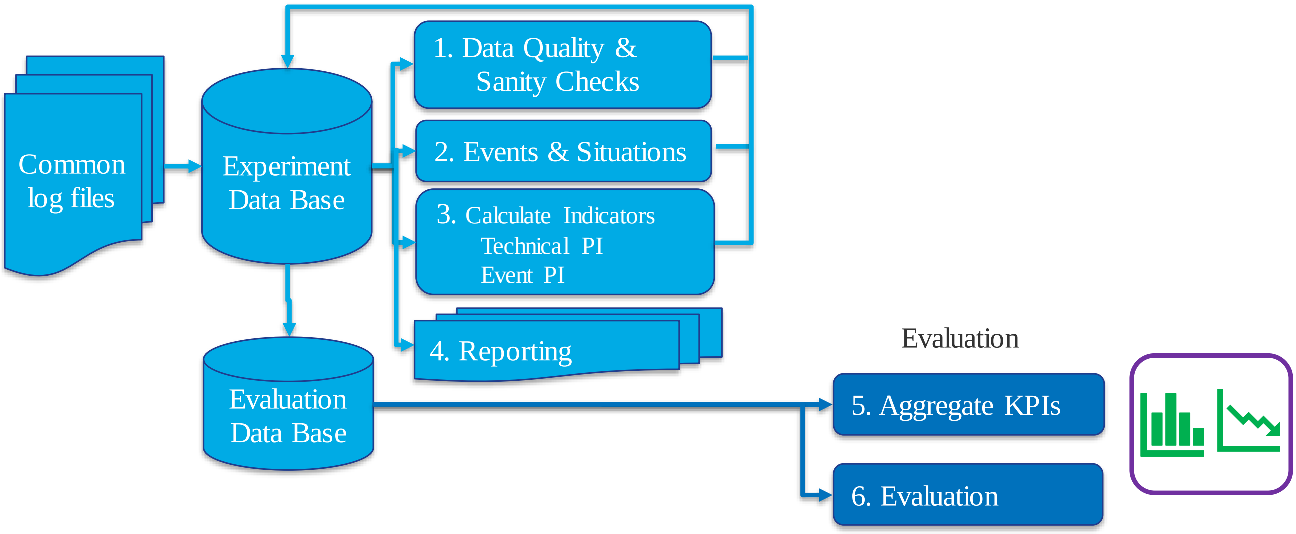

Figure 17: Data management and flow in C-MobILE (diagram shows more than EWD)

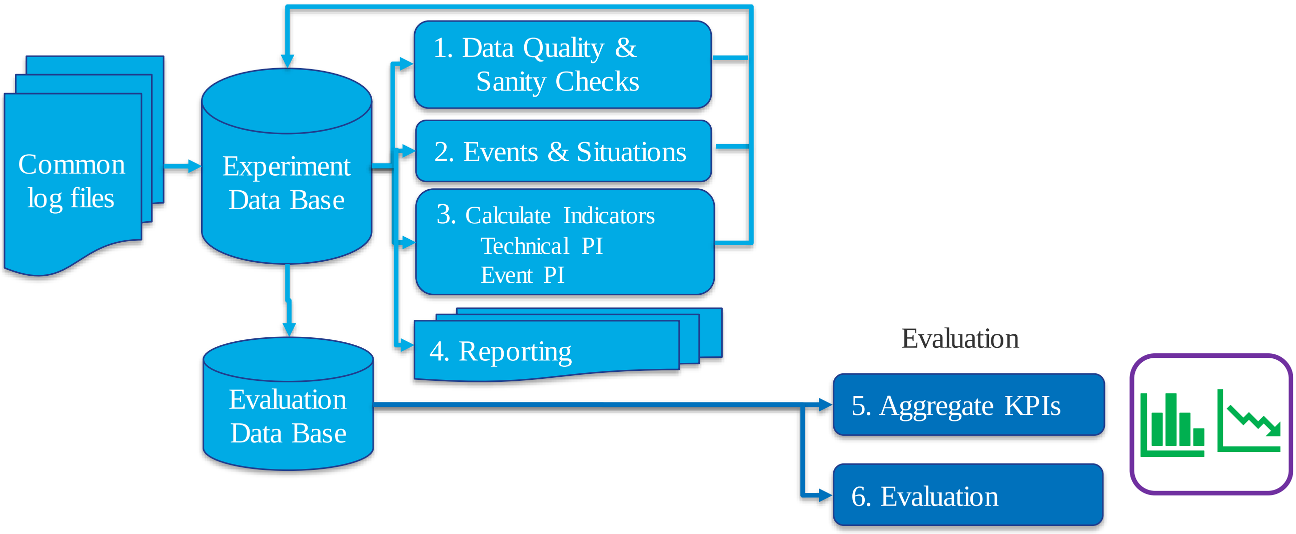

Figure 17 shows the envisaged overall data management and flow from the deployment sites to the evaluators in C-Mobile. Data are collected from the deployment sites and uploaded to an agreed data structure and format to the CTAG database. Tests of data quality and completeness (see Section 4.7), and identification of events, are carried out. For some services (RWW and GLOSA), tools for data quality checks and event identification are available. Performance Indicators are calculated on the processed and screened data, stored in the database for analysis by evaluators. EWD covers step 3 in

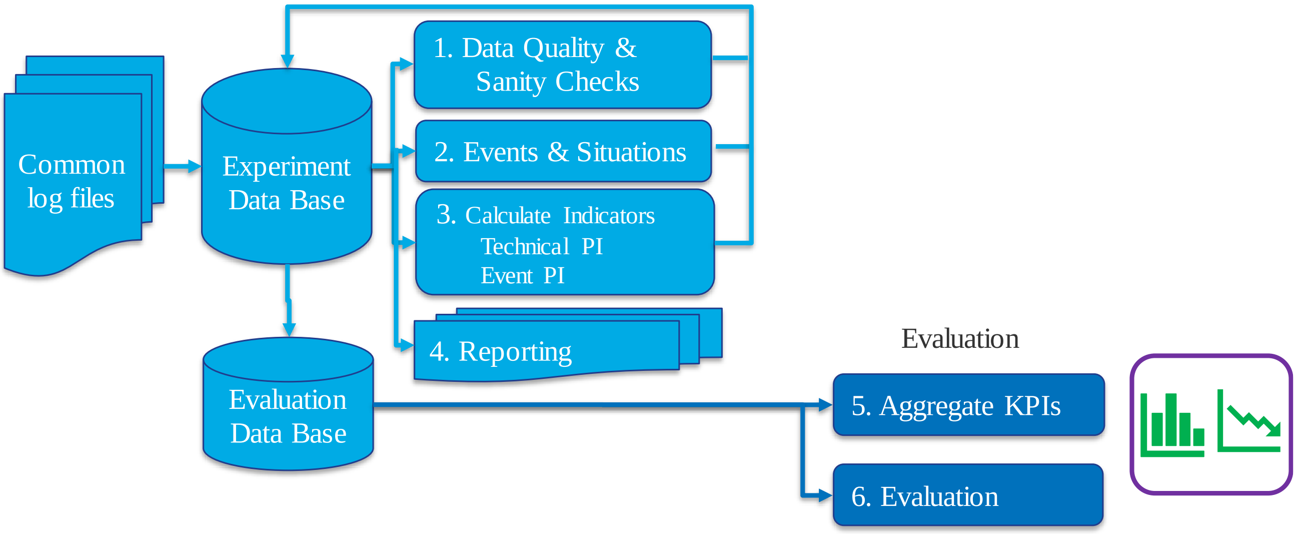

Figure 17.

4.9.4.2. Example

For the Signal Violation Warning Service, examples of Research Questions, Hypotheses, PI's, data to collect, definition of events, and types of analyses are provided.

Signal Violation Warning (SVW) aims to reduce the number and severity of collisions at signalized intersections by warning drivers who are likely -due to high speed or inattention- to violate a red light, or when another vehicle is likely to make a red light violation. Also known as the "Signal violation / Intersection Safety" or "Red Light Violation Warning" (source: D2.2).

The SVW use case is still in development. The following functionality of SVW is assumed:

- The potential red-light violator is provided a warning to stop in a timely way that provides the violator time to stop without violating the red light.

The EWD methodology for Monitoring can provide the following:

- Count the number of warnings (virtual or real) provided by the SVW service per unit of time, e.g., weekly basis.
- In case of a warning for a violator that receives a SVW; count the amount of 'successful' stops (defined as the amount of vehicles that stop within x seconds after receiving the warning) and the fraction of the successful stops in relation to the total number of warnings.

The example focuses on a safety-related research question, Will SVW affect safety? The following hypotheses can be defined to answer this Research Question (this is an incomplete list):

Hypothesis	KPIs
Signal Violation Warning aided me to stop safely before I made a red light violation	Subjective measurement PI: Perceived usefulness of SVW, using the Likert 5-point scale
The use of SVW will reduce the number of red light violations	Objective measurement: PI 2: Number of red light violations

Table 17: Example of hypotheses and PIs for Signal Violation Warning

4.9.4.3. Data

The data to be collected is determined by the PI's needed to test the hypotheses. The PI's can either be measured directly or derived from other measures (raw, derived, self-reported and situational variables), according to the FESTA handbook [1]. Annex 2 contains examples of data that need to be collected for evaluation.

The data that need to be collected at each deployment site depends on many factors, one of which is experimental design. The experimental design affects how data will be collected and how the analysis is carried out. The baseline and treatment need to be defined, as discussed in Section 4.9.3. If small numbers of users (in space and time) in the overall traffic stream are expected, then in-vehicle logging may be necessary. The logging can take place via an on-board unit or via a smartphone, each with its advantages and disadvantages. In the case of in-vehicle logging, the baseline or control group refers to the period during which logging takes place and the services are working in the background, but no message or warning is shown to the driver on the human machine interface. However, the system should log the virtual message and time shown to the driver in the baseline or control group in the same way as when the logging takes place in the treatment period. The phase itself (baseline or treatment) also needs to be logged, either by the in-vehicle device or by a person responsible designated at the deployment site.

4.10. Ex-post cost benefit assessment

The aim of the ex-post cost-benefit analysis (CBA) is to contribute in the post-demonstration assessment of the C-ITS services' impact and value, by providing evidence to decision-makers on the costs and benefits of bundled C-ITS services. The D6.5 "Ex-post cost-benefit analysis and guidance report" will be executed at the end of the project (M40), comprising an evaluation based on the real-world deployment in the deployment sites. The ex-post CBA will complete the validation procedure of the C-ITS services by means of [4]:

- Assessing the added value and economic viability of the C-ITS services bundling for key stakeholders and end-users.
- Elaborating findings for the development of the business and exploitation plans for the large-scale C-ITS deployment within Task 4.3. (Business and exploitation plans for faster market roll-out).
- Providing a model/ guidelines for CBA for the large-scale deployment of C-ITS.

Overall, the ex-post CBA will constitute one of the seven key results of the project, being the assessment of the cumulative real-life benefits of bundling C-ITS applications and integrating multiple transport modes in the C-ITS ecosystem [4].

Prior to the ex-post CBA, an ex-ante CBA (D2.1), was executed at the beginning of the project. It built mainly upon assumptions and estimated values for costs and benefits, based on literature review findings and previous projects' results. For the execution of the ex-ante CBA, the snapshot approach was followed, which suggests the preselection of one or several target years and the calculation of the BCRs for these target years. In this case the costs are transformed to annual values (using a discount rate) and then compared to the target year benefits [26]. Within the C-MobILE framework the baseline scenario was defined as the one describing the deployment sites' current situation, 2017, since the extent of the already existing C-ITS services' deployment differs in each one of them. The with-the-project scenario described a reference situation depicting the C-ITS services deployment in the deployment sites by 2020, i.e. the C-MobILE extensions and updates in each deployment site. The diversity between the deployment sites attributes', depicting the different technology settings, necessary for the C-ITS services deployment, indicated the need to define a set of common assumptions, in order to proceed with the analysis.

As a first step, costs related to investments and to the operation and maintenance of the C-ITS systems were collected from the literature review. Given the fact that there was a diversity between the base years for the costs data originating from the various sources, all pre-2017 costs had to inflate to 2017 levels (C-MobILE baseline), using the Eurostat Harmonised indices of consumer prices (HICPs) for each of the countries represented by the deployment sites (Denmark, France, Greece, Netherlands, Spain and UK) [27]. In order to calculate total annual costs, i.e. investment costs and operation and maintenance costs, for each component of the C-ITS systems, investment costs, which appear only once in the lifetime of the system, had to be annualized. For annualizing the investment costs, information about the discount rate and the lifetime of the systems was necessary. The following equations refer respectively to the calculations needed for costs inflation (1) and annualization (2), (3):

(1) $Cost_{2016} = \dfrac{HICP_{2016} \times Cost_{\,pre-2016}}{HICP_{pre-2016}}$

(2) $System\ cost\ per\ year_{year\ x} = System\ costs\ total_{year\ x} \times AR$, with

(3) $AR = \frac{d \times (1+d)^n}{(1+d)^n - 1}$, with

AR: annuity rate,

d: discount rate, i.e. 4%, and

n: lifetime, i.e. infrastructure: 15 years, and vehicles (different types): according to ACEA data referring to the average age of the EU car fleet by country in 2015 (latest available data) [28].

The second step of the analysis comprised of the benefits' calculation, hence data concerning the impacts of the C-ITS services, at EU level, were extracted from the literature review. The data had to be modified, so as to express impacts at deployment site level. The procedure, followed to produce the most accurate estimates of such impacts, included the scaling down of the impacts of the C-ITS services from EU level to deployment site level. The method used as a starting point the effects on road safety, traffic efficiency and the environment on a large-scale level (EU), distinguished for different situations (surroundings such as traffic state, vehicle type, etc.). Multipliers, necessary for the scaling down of the impacts, were the total annual number of vehicle kilometres (Vkm) driven in a year per road type, i.e. motorways, inter-urban roads and urban roads, in each deployment site. This data was extracted from questionnaires distributed among the deployment sites. The scaling down methodology led to the identification of individual impact rates, concerning safety, traffic efficiency and the environment, per C-ITS service and per deployment site. Thereafter, the physical impacts for each deployment site by 2020, were converted into euros (€), by attributing to them estimates of 2020 inflated market prices, expressing the total benefits of the C-ITS services for each deployment site [29].

The last step of the analysis was comprised of the comparison between the estimated 2020 total costs and total benefits of each deployment site, according to the following equation:

$$BCR = \frac{Total\ Benefits_{2020}}{Total\ Costs_{2020}}$$

The evaluation of the deployment sites' benefit cost ratios (BCRs) was conducted in line with the following classes [30]:

1. 0 < BCR < 1: The BCR is rated "poor" showing the socio-economic inefficiency of the C-ITS services deployment.
2. 1 ≤ BCR < 3: The BCR is rated "acceptable" meaning that the social benefits, associated with the implementation of the C-ITS services, exceed the costs up to almost three-times, which can be labelled as an acceptable absolute efficiency.
3. BCR ≥ 3: The BCR is at least as high as "3" indicating an "excellent" result of the CBA. The C-ITS services should be in first line for market deployment.

For the performance of the ex-post CBA, the inputs of the ex-ante CBA for the reference scenario will be updated by real-world data. This data will be gathered during the real-world deployment of the C-ITS services in the deployment sites of the project and will refer to:

- Measured evidence of the C-ITS deployment parameters, e.g. indicators and figures describing the extent of the C-ITS services' supporting technologies.
- Up-to-date market prices of the supporting technologies, which will be used for the calculation of the C-ITS services' investment and operational costs.
- Impacts of the C-ITS services, which will have a monetary value attributed to them, in order to proceed with the benefits calculation.

The methodology of the ex-post CBA is similar to the classical approach of a CBA, which is based on standard welfare economics, a branch of economic theory mainly considering optimal allocation of resources to increase the welfare of society [8]. In this way a CBA is an analytical tool for analysing the economic advantages or disadvantages of an investment decision by accounting for its costs and benefits, in order to assess the welfare change attributable to it [9]. Within the framework of the ex-post CBA, the potential benefits of the deployment of the bundled C-ITS services will be compared with all the relevant consumption in resources due to their implementation. All the benefits and costs will be measured in monetary terms by multiplying the physical impact units with the accordant cost-unit rates [11].

The first step for the performance of the ex-post CBA will be the identification of the costs, including all investments as well as all tactical and operational costs, year after year, for a given time period. The considered time horizon for the C-MobILE project would be of 10 years, 2030, according to the proposed time horizon for infrastructure projects by the European Commission [9]. The next step will be the identification and qualification of benefits for the same time horizon. The benefits to arise are aligned to specific impact areas, defined in terms of mobility, safety, and efficiency as well as environmental impacts (see section 4.7). Since all costs and benefits are converted into a monetary value, the parameter of the time horizon will be taken into account, in order to compare the two quantities (costs and benefits) at the two different periods (2020 and 2030). This will be achieved by making use of an updating rate “a”. The equation to be used, describing the relationship between the value of a quantity of money “V_t” at a present time “t” and the value of this quantity at a future horizon n “V_n”, is the following one [9]:

$$V_t = V_n/(1+a)n, \text{ where}$$

a: updating rate, e.g. 4%

t: 2020 (present time)

n: 2030 (future horizon)

Year by year the benefits will be compared to the costs and their difference will be updated using the update rate. The final result will be the calculation of an Investment Return Rate (IRR) for the 10-year horizon.

4.11. Ethical and legal issues

According to FESTA; several aspects should be considered for piloting and analysing the data from an ethical and legal perspective such as participant recruitment and agreement as well as data protection, safety, approval for on-road use etc. Relevant for the validation and impact assessment is the processing of private data and how privacy issues are resolved. It is assumed that any sensitive private data will already be removed or anonymized by the deployment sites before providing data to WP6. Note that the evaluation does not intend to give statistics for user ID’s over events and tests. These aspects must be worked out for each deployment site, as regulation and approval procedures may vary amongst countries.

Section 6 of the FESTA handbook [1] gives an overview of legal and ethical issues to be considered for piloting. Issues like participant selection, data protection, safety, approval for on-road use, etc. are primarily the responsibility for pilot operations.

Potentially privacy regulations may pose issues in the distribution of pilot data between public, research and private partners and between member states.

References

[1] FESTA Consortium. (2016). *FESTA Handbook Version 6. Field opErational teSt supporT Action.*

[2] CAR 2 CAR Communication Consortium. (2010). *C2C-CC The handbook for Vehicle-to-X cooperative systems simulation.*

[3] Adaptive Integrated driver-vehicle interface (AIDE) D2.2.5 Driving Performance assessment –methods and metrics

[4] European Commission, Innovation and Networks Executive Agency (2017). Grant Agreement number: 723311 – C-MobILE – H2020-MG-2016-2017/H2020 MG-2016-Two-Stages.

[5] ETSI TR 102 638 V1.1.1 (2009-06) Intelligent Transport Systems (ITS); Vehicular Communications; Basic Set of Applications; Definitions

[6] Regulation (EC) No 561/2006 of the European Parliament and of the Council of 15 March 2006 on the harmonization of certain social legislation relating to road transport and amending Council Regulations (EEC) No 3821/85 and (EC) No 2135/98 and repealing Council Regulation (EEC) No 3820/8

[7] C-ITS Platform final report of January 2016

[8] Boardman, A. E., Greenberg, D. H., Vining, A. R., & Weimer, D. L. (2017). *Cost-benefit analysis: concepts and practice*. Cambridge: Cambridge University Press.

[9] European Commission. (2014). *Guide to Cost-Benefit Analysis of Investment Projects; Economic appraisal tool for Cohesion Policy 2014-2020.*

[10] IEEE Standard for System and Software Verification and Validation. IEEE Std 1012™-2012 (Revision of IEEE Std 1012-2004), IEEE 3 Park Avenue New York, NY 10016-5997 USA. 25 May 2012

[11] Luedeke, A., Schindhelm, R., Francano, M., Levizzani, C., Damiani, S., Manfredi, S., Revellino, M., Mortara, P., Alkim, T., Feenstra, P., de Kievit, M., Wilschut, E., Zwijnenberg, H., Geissler T., & Westerkamp, U. (2010). *SP6 – BLADE – Business models, Legal Aspects, and DEployment; Report on socio-economic, market and financial assessment.* Version 12.

[12] F. Barmpas. 2016. Compass4D; D6.4 Cost Benefit Analysis. Version 1.1.

[13] Hankey, J. M., Perez, M. A., & McClafferty, J. A. (2016). Description of the SHRP 2 Naturalistic Database and the Crash, Near-Crash, and Baseline Data Sets. Virginia Tech Transportation Institute.

[14] Behbahani, H., & Nadimi, N. (2015). A Framework for applying surrogate safety measures for sideswipe conflicts. *International Journal for Traffic and Transport Engineering, 5*(4). 371-383.

[15] Davis, F. D. (1985). *A technology acceptance model for empirically testing new end-user information systems: Theory and results* (Doctoral dissertation, Massachusetts Institute of Technology).

[16] Bertrand M., & Bouchard S. (2008). Applying the technology acceptance model to VR with people who are favourable to its use. *Journal of Cyber Therapy and Rehabilitation, 1*(2), 201-211.

[17] Ajzen, I. (1985). From intentions to actions: A theory of planned behavior. In J. Kuhl, & J. Beckman (Eds.). *Action-control: From cognition to behavior* (pp. 11-39). Heidelberg: Springer.

[18] Kindberg, T., Sellen, A., & Geelhoed, E. (2004). Security and trust in mobile interactions. A study of users' perceptions and reasoning. In N. Davies, E. Mynatt, & I. Siio (Eds.). *Proceedings of Ubicomp 2004: Ubiquitous Computing 6th International Conference* (pp. 196-213). Berlin: Springer.

[19] ISO/IEC 25010:2011 Systems and software engineering -- Systems and software Quality

Requirements and Evaluation (SQuaRE) -- System and software quality models.

[20] Michael, J.B., Riehle, R., & Man-Tak Shing (2009). The verification and validation of software architecture for systems of systems. In Proceedings of the IEEE International Conference on System of Systems Engineering (SoSE 2009) May 30-June 03 2009 (pp.1–6). Institute of Electrical and Electronics Engineers (IEEE).

[21] OMG Systems Modeling Language™ (OMG SysML®). http://www.omgsysml.org/what-is-sysml.htm. Last visited 07/11/2017.

[22] Dobrica, L., & Niemela, E. (2002). A survey on software architecture analysis methods. *IEEE Transactions on Software Engineering archive, 28*(7). 638-653.

[23] Carnegie Mellon Software Engineering Institute. Architecture Tradeoff Analysis Method. https://www.sei.cmu.edu/architecture/tools/evaluate/atam.cfm Last visited 15/11/2017.

[24] Model Driven UML Tools. Enterprise Architect. SPARX Systems. http://www.sparxsystems.eu/start/home/ Last visited 08/11/2017.

[25] Litman, T. (2003-2011). Measuring Transportation: Traffic, Mobility and Accessibility. *Institute of Transportation Engineers (ITE) Journal, 73*(10), 28-52. (Original publication, constantly updated). Retrieved from: http://www.vtpi.org/measure.pdf

[26] Kulmala, R., Leviäkangas, P., Sihvola, N., Rämä, P., Francsics, J., Hardman, E., Ball, S., Smith, B., McCrae, I., Barlow, T., & Stevens, A. (2008). CODIA final study report (CODIA Deliverable 5). VTT Technical Research Centre of Finland.

[27] Eurostat. (2017). *HICP methodology*. [01/06/2017; 05/01/2018]. http://ec.europa.eu/eurostat/statistics-explained/pdfscache/1119.pdf.

[28] ACEA. 2015. Average Vehicle Age. [05/01/2018]. http://www.acea.be/statistics/tag/category/average-vehicle-age.

[29] Mitsakis, E., & Kotsi, A. (2018). C-MobILE D2.1 Ex-ante Cost-Benefit Analysis. Version 1.1.

[30] Schindhelm, R., Luedecke, A., Geissler, T., Westerkamp, U., Feenstra, P., & de Kievit, M. Socio-economic Viability of Safespot Cooperative Safety Systems.

[31] ITS Joint Program Office. (2017). *Connected Vehicle Reference Implementation Architecture (CVRIA)*.

[32] van Sambeek, M., Ophelders, F., Bijlsma, T., Türetken, O., Eshuis, R., Traganos, K., & Grefen, P. W. (2015). Towards an architecture for cooperative-intelligent transport system (C-ITS) applications in the Netherlands. *Eindhoven University of Technology, Tech. Rep. Beta Working Paper series, 485*. 1–134.

[33] Dalla Chiara, B., Deflorio, F., & Diwan, S. (2009). Assessing the effects of inter-vehicle communication systems on road safety. *IET Intelligent Transport Systems, 3*(2), 225-235.

[34] Elvik, R., Christensen, P., & Amundsen, A. (2004). Speed and road accidents. An evaluation of the Power Model. *TØI report, 740*.

[35] Nilsson, G. (2004). Traffic safety dimensions and the power model to describe the effect of speed on safety. *Bulletin-Lunds Tekniska Högskola, Inst för Teknik och Samhälle, Lunds Universitet, 221*.

[36] Wang, H., & Chow, S.-C. (2007). Sample Size Calculation for Comparing Proportions. *Wiley Encyclopedia of Clinical Trials*, 1-11.

[37] Int Panis, L. I., Broekx, S., & Liu, R. (2006). Modelling instantaneous traffic emission and the influence of traffic speed limits. *Science of the total environment, 371*(1-3), 270-285.

[38] Peeters, B., van Blokland, G., & de Graaff, E. (2005). Development of an European road traffic

emission model within the framework of the EU 6th Framework project IMAGINE. *Forum Acusticum Budapest 2005*, 279-284.

[39] https://www.usability.gov/

[40] Adell, E. (2009). Driver Experience and Acceptance of Driver Support Systems - A Case of Speed Adaptation. Department of Technology and Society, Lund University.

[41] Venkatesh, V., Morris, M. G., Davis, G. B., & Davis, F. D. (2003). User acceptance of information technology: Toward a unified view. *MIS quarterly*, 425-478.

[42] Ghazizadeh, M., Lee, J. D., & Boyle, L. N. (2012). Extending the Technology Acceptance Model to assess automation. *Cognition, Technology & Work*, *14*(1), 39-49.

[43] Osswald, S., Wurhofer, D., Trösterer, S., Beck, E., & Tscheligi, M. (2012). Predicting information technology usage in the car: towards a car technology acceptance model. In A.L. Kun (Ed.). *Proceedings of the 4th International Conference on Automotive User Interfaces and Interactive Vehicular Applications* (pp. 51-58). New York, NY: ACM.

[44] Najm, W., Stearns, M., Howarth, H., Koopmann, J., & Hitz, J. S. (2006). *Evaluation of an automotive rear-end collision avoidance system* (No. FHWA-JPO-06-055). United States National Highway Traffic Safety Administration.

[45] Regan, M. A., Triggs, T. J., Young, K. L., Tomasevic, N., Mitsopoulos, E., Stephan, K., & Tingvall, C. (2006). On-Road Evaluation of Intelligent Speed Adaptation. Following Distance Warning and Seatbelt Reminder Systems: Final Results of the TAC SafeCar project, 253, 1-270.

[46] Isa, M. H. M., Deros, B. M., & Kassim, K. A. A. (2015). A Review of empirical studies on user acceptance of driver assistance systems. *Global Journal of Business and Social Science Review*, *4*(1). 182-189.

[47] IEEE Computer Society. Software Engineering Technical Committee. (1998). *IEEE Guide forDdeveloping System Requirements Specifications*.

Annex 1. Technical requirements for evaluation

The following general technical requirements address basic technical issues coming along with measuring the key performance indicators (KPI). General refers to the requirements as being independent of levels of impact evaluation (technical level, user level, traffic system level, business level) and single services. They were derived from partners' expertise involved also in the Work Package 6, aiming to Validation and impact assessment. Moreover, technical requirements that address single levels of impact and single services or specific bundles have been derived from partners' expertise and will be further updated.

To ensure the quality of the requirements, multiple review rounds have been performed. During those rounds, each requirement has been checked by multiple experts from the consortium as well as from experts from the Deployment Sites in order to ensure interoperability with a common understanding.

The following set of requirements is focused on the data level and technical aspects of the evaluation phase.

ID	R/G04-EV-01
Requirement	Every Pilot Site should have a server that logs data from their deployed services.
Description	Every Pilot Site should establish a server for logging data from the services in order to enable evaluation at Pilot Site level.
Affected entities	POMS, PS

ID	R/G04-EV-02
Requirement	Access to logging servers shall be restricted
Description	The access to logging servers must be restricted to certain users, e.g. by username/passwords. Each server admin should be in possession of a list of the users which have access to the data.
Affected entities	C-MobILE database, POMS, PS

ID	R/G04-EV-03
Requirement	Data shall be anonymized at Pilot Site level in order to guarantee participants' privacy but also to guarantee proper evaluation
Description	Privacy data should be removed before sending data to the C-MobILE database. Personal data shall be anonymized as early as possible, e.g. while collecting data. Users/ participants shall be informed about the purpose personal data is logged for (e.g. evaluation) and their informed consent must be given.
Affected entities	POMS, PS

ID	R/G04-EV-04
Requirement	Data collected by Pilot Sites shall be quality checked locally in order to assure meaningful data is logged and to assure right data formats.
Description	Meaningful data shall be logged and logging data which is unnecessary or noisy shall be discarded. The data to be logged comes from the defined KPIs within WP6.
Affected entities	POMS, PS

ID	R/G04-EV-05

Requirement	Quality checked data from Pilot Sites should be synchronized with the SQL database's suitable sub-tables in a regular basis in order to assure and ease the while-doing analysis of it.
Description	In order to assure and ease the while-doing analysis of the data, the synchronization should be once a day.
Affected entities	C-MobILE database, POMS, PS

ID	R/G04-EV-06
Requirement	Databases within C-MobILE shall be hosted and capable to embed local SQL files.
Description	Servers shall exist were data is stored. These servers shall allow receiving data from local POMS, Vehicle ITS-S and Roadside ITS-S.
Affected entities	C-MobILE database

ID	R/G04-EV-07
Requirement	Every Pilot Site and the central databases should have a backup data system in case of unexpected failures in order to avoid loss of data
Description	Periodic backups of data should be carried out.
Affected entities	C-MobILE database, POMS, PS

ID	R/G04-EV-08
Requirement	Geofencing may be enabled and used in order to automate data collecting onsets
Description	Specific events are triggered once a user/participant leaves/enters a geographical area. Logged data may refer only to the fenced areas.
Affected entities	C-MobILE server, C-MobILE database, POMS, PS

ID	R/G04-EV-09
Requirement	Analysis toolboxes must be integrated in evaluation tools used and be adaptable/ accessible via their interface.
Description	Alternatively data itself must be accessible in order to analyze it.
Affected entities	C-MobILE database

ID	R/G04-EV-10
Requirement	Names, units, types of measures, and measuring frequency in data collection and in the database must be standardized in order to realize efficient input and pull of data. Frequencies may be easily adjusted/ matched during input or pull of data.
Description	Names, units and types should refer to FESTA handbook [1]. The exact procedure and position of standardization within the flow of defining, collecting, input and pull of data must be determined. For example, vehicles and infrastructure might receive or send information with different frequencies and to understand and retrace time lapses of user behavior it may be necessary to match this data.
Affected entities	C-MobILE database

ID	R/G04-EV-11
Requirement	Subjective measures should be captured in the same database as all other measures are in order to ease the analysis

Description	To easily pull related data and implement control variables that were assessed by a vehicle or infrastructure, subjective data should be integrated in the same database. If the amount of data escalates, it may be feasible to structure it and to create, for example, a corresponding sub-table.
Affected entities	C-MobILE database

ID	R/G04-EV-12
Requirement	Data from different sources (vehicles, infrastructure) shall be logged and matched spatially in order to do spatial related analyses
Description	For example, it must be ensured that it is logged with which Roadside ITS-S a vehicle is currently close to so that it can be traced if the vehicle/driver is receiving information from Roadside ITS-S and if so what information is communicated (e.g. which signal phase).
Affected entities	C-MobILE server, C-MobILE database

ID	R/G04-EV-13
Requirement	System input and output shall be logged in order to analyze the user's interaction with it
Description	For example, the onsets of turning-on, communication with other entities, or warnings have to be logged.
Affected entities	C-MobILE server, C-MobILE database, Vehicle ITS-S, HMI

ID	R/G04-EV-14
Requirement	All timestamps used in logging should use the same time format and reference as the ETSI messages.
Description	The timestamps should use milliseconds since 2004-01-01, based on TAI, according to the ETSI TS 102 894-2
Affected entities	C-MobILE database, POMS

ID	R/G04-EV-15
Requirement	Data shall be accurately described
Description	Data should be described in a common 'evaluation data glossary'. In this glossary, the name, ID, frequency, data range, type and unit of each data entry should be described. Additionally, it should contain a short description of the data.
Affected entities	C-MobILE database, POMS

ID	R/G04-EV-16
Requirement	Data formats and deviances should be defined in written form and cross-checked before implementation in order to avoid discarding the use of common data formats.
Description	All data formats, as well as the expected ranges of the data should be written in a common data glossary. This data format should be discussed and verified by both, WP5 and WP6 experts. Expected changes in metrics/KPIs should be given, to allow for easier statistical assessment
Affected entities	C-MobILE database, POMS

ID	R/G04-EV-17
Requirement	GNSS accuracy and precision (both temporal and spatial) should be recorded

Description	Being able to know the precision of the data that is being processed is important as it may help discarding erroneous data or focus on the more reliable data.
Affected entities	POMS, C-MobILE server, C-MobILE database

ID	R/G04-EV-18
Requirement	Flow data (and other non-probe vehicle traffic data) should be recorded where possible
Description	Flow data helps to evaluate whether the services are effective or not
Affected entities	POMS, C-MobILE database

ID	R/G04-EV-19
Requirement	Driving behavior should be assessed in order to meet the analyses for some of the use cases' objectives
Description	For example, the reaction to safety critical warnings may require measuring reaction times on a millisecond level, safety critical behavior (e.g. hastiness) may require the analysis of high-resolution steering wheel angle changes within a specific time frame.
Affected entities	C-MobILE systems, Vehicle ITS-S, C-MobILE database

Annex 2. Indicators, variables and hypotheses

PI index	Performance indicator on individual level	Definition	Measures
PI-1	Spot speed	Users' GPS based: $Speed(t) = \frac{Haversine[pos(t), pos(t+1)]}{Duration(t,t+1)}$ $for\ t\ in\ 1,...,t-1$ Camera based: Non-users' data can be collected	Directly from sensors/ derived from GPS/ cameras
PI-2	Mean speed	Mean speed of a user's vehicle in the area of interest	derived from sensors trace or GPS spot speeds
PI-3	Speed variance	Speed variance of a user's vehicle in the area of interest	Derived from trace/ spot speeds and mean speed
PI-4	Initial speed	Spot speed of a user when entering the area of interest	Derived from sensors/ GPS spot speed
PI-5	Speed violation	Binary if user's spot speed exceeds threshold	Derived from spot speeds
PI-6	Spot acceleration	Users' GPS based: $Acceleration(t) = \frac{Speed(t,t+1)}{Duration(t,t+1)}$ $for\ t\ in\ 1,...,t-2$	Sensors/ GPS
PI-7	Mean acceleration	Mean of a single user's spot acceleration in the area of interest	Derived from spot accelerations
PI-8	Maximum deceleration/ acceleration	Peak level of a single user's absolute spot deceleration/acceleration	Derived from spot acceleration
PI-9	Distance of braking event to an incident point	Distance in time and space to incident point where user's deceleration exceeds threshold of 0.2 m/s^2 in the area of interest	Derived from spot acceleration and basic measures
PI-10	Hard braking event	Binary if user's deceleration exceeds threshold of 4.5 m/s^2 in the area of interest	Derived from spot acceleration
PI-11	Time speed < 2 km/h	Time a single user is traveling below 2 km/h in the area of interest	
PI-12	Stop	(spot speeds < 2 for > 2 seconds)	Directly from sensors/ derived from GPS spot speeds
PI-13	Red light violation per traffic light	Binary if red light violated by a user	Sensors/ cameras
PI-14	Red light violation per equipped user	Binary if red light violated by a user	OBU
PI-15	Lane violation per network element	Binary if lane is violated by user	Cameras, GPS

PI-16	Resting time violation	Binary if driving time is exceeded by user	OBU
PI-17	Vehicle-kilometres driven	Total kilometres driven by user	Questionnaire, Directly via OBU
PI-18	Travel time	Time it takes a user passing an area of interest	Derived from length of area of interest and mean speed or time
PI-19	Number of parking spaces occupied	Number of spaces occupied within a period of time	Cameras, reported by third-party
PI-20	Number of parking spaces	Number of permanently available spaces	Secondary data
PI-21	Bicycle Trip	GPS: Binary if location change time-out exceeding a threshold of tbd + 1 (mode detection required)	OBU/ GPS
PI-22	Car trip	OBU: Binary if engine start after time-out exceeding a threshold of tbd GPS: Binary if location change time-out exceeding a threshold of tbd + 1 (mode detection required)	OBU/ GPS
PI-23	Total number of (bicycle/car) trips	Subjectively: self-reported by user OBU/ GPS: sum of all car trips a user made	Questionnaire, derived from trips
PI-24	Average number of (bicycle/ car) trips	Subjectively: self-reported by user OBU/ GPS: mean of all car trips a user made within a period of time	Questionnaire, derived from trips
PI-25	Average trip length	Subjectively: self-reported by user	Questionnaire
PI-26	Length of area of interest	Distance driven in area of interest Users' GPS based: $Dist\left(vehicle\right)=\sum_{t=1}^{t=n-1}Haversine\left[pos(t),pos(t+1)\right]$	Sensors/ GPS
PI-27	Energy consumed	kW or fuel consumed by user in the area of interest	OBU/ GPS
PI-28	Queue length per traffic light phase	Number of vehicles in a queue per phase	Cameras
PI-29	Number of vehicles per traffic light phase	Number of vehicles passing traffic light per phase	Cameras, Induction Loops
PI-30	CO_2 emission	User's CO_2 emitted within the area of interest, or at an incident point	derived from sensors, derived from GPS speed and acceleration, RSU spot CO_2 measure per unit of time
PI-31	Noise emission	User's noise emitted within the area of interest, or at an incident point	Derived from speed and acceleration, RSU spot noise measure per unit of time

Table 18. Performance indicators on individual level for the impact assessment on users and the transport system

BM index	Basic measures
BM-1	Longitude
BM-2	Latitude

BM-3	Altitude
BM-4	Time
BM-5	ID

Table 19. Basic GPS measures on vehicle layer

SC index	Situational/ control variable	Definition
SC-1	RSU ID	Road side unit identifier
SC-2	Vehicle ID	Vehicle identifier
SC-3	Vehicle type	Nominal [Light, heavy…,private, commercial…]
SC-4	Engine type	Nominal [petrol, diesel, electric…]
SC-5	User ID	User identifier
SC-6	User type	Nominal [pedestrian, bicycle, private car…]
SC-7	User within area of interest	Binary if user is in area of interest
SC-8	Service on/off	Binary if service is technically available
SC-9	Service setting	Nominal
SC-10	User switched on/off	Binary if user changed service availability/activity status
SC-11	Service passive active for user	Binary if it makes sense that service provides information
SC-12	Service active for user	Binary if service's information is displayed
SC-13	Service active for RSU	Binary if service gives priority
SC-14	Number of service operations per sequence	Count how many times a service was at least passive active on one trip
SC-15	Weather conditions	Nominal
SC-16	Lighting conditions	Ordinal/ Nominal
SC-17	Traffic conditions	Nominal
SC-18	Other active traffic measures in the area	Nominal
SC-19	Road type	Nominal [urban arterial, main, secondary, residential, motorway…]

SC-20	Speed Limit	Speed Limit on Lane/ Road or in area of interest
SC-21	Number of lanes	
SC-22	Lane ID	Number of lane counted from left to right (reversed in Newcastle)
SC-23	Lane width	Width of the lane driven
SC-24	Road Condition	Nominal
SC-25	Other road characteristics	Nominal
SC-26	Traffic volume in area of interest	Number of cars
SC-27	Total distance travelled per day	
SC-28	Trip category	Nominal [leisure, commute,...]
SC-29	Trip length per trip	
SC-30	States of RSUs	States of e.g. a traffic light during passing an intersection
SC-31	Intersection classification	Nominal
SC-32	Egress From Traffic light/ Route (vehicle manoeuvre during intersection)	Binary/Nominal
SC-33	Distance to incident point	Distance to e.g. a traffic light
SC-34	Having passed incident point	Binary if incident point was passed
SC-35	Equipped non-C-Mobile users	Binary
SC-36	GPS HDOP	
SC-37	GPS quality indicator	
SC-38	Heading	in degrees
SC-39	Date	Day and year

Table 20. Universal situational/ control variables (service specific will be defined in the detailed reports D6.2-D6.4)

Index	Hypotheses	Service X																			
		RTM	MPA	UPA	RWW	RHW	EVW	SVW	WSP	GP	GLOSA	CTLP	FI	IVS	MTTA	PVD	EBL	CACC	SSVW	MAI	BSD
HY-1	The use of C-MobILE service X changes car usage											X			X						
HY-2	The use of C-MobILE service X increases bicycle usage														X						
HY-3	The use of C-Mobile service X increases drivers compliance with driving/resting times	X	X																		
HY-4	The use of C-Mobile service X reduces the risk of accidents		X	X		X	X		X	X				X			X		X	X	X
HY-5	The use of C-Mobile service X reduces the number of collisions/ accidents (with sb./sth.)			X		X		X													
HY-6	The use of C-Mobile service X reduces the severity of accidents							X						X							
HY-7	The use of C-Mobile service X reduces red light violations											X									
HY-8	The use of C-Mobile service X does not change the number of red light violations										X										
HY-9	The use of C-Mobile service X increases the distance (time and path) of braking events to an incident point				X	X											X		X		X
HY-10	The use of C-Mobile service X reduces hard braking events				X	X													X		X
HY-11	The use of C-Mobile service X reduces speed limit violations													X							
HY-12	The use of C-Mobile service X does not elicit lane violations												X								
HY-13	The use of C-Mobile service X reduces parking space search time	X		X																	
HY-14	The use of C-Mobile service X optimizes the flow of vehicles		X	X							X										
HY-15	The use of C-Mobile service X reduces vehicle-kilometres driven		X																		
HY-16	The use of C-Mobile service X increases punctuality									X											
HY-17	The use of C-Mobile service X reduces travel time (EV response times, service times)						X			X		X									
HY-18	The use of C-Mobile service X reduces parking space demand											X									
HY-19	The use of C-Mobile service X reduces VRUs' waiting time											X									
HY-20	The use of C-Mobile service X increases travel times of cars											X									
HY-21	The use of C-Mobile service X reduces energy consumption (fuel, kW)										X							X			
HY-22	The use of C-Mobile service X improves the performance of traffic lights										X										
HY-23	The use of C-Mobile service X reduces lost time										X										
HY-24	The use of C-Mobile service X reduces pollution/emissions			X							X	X			X						
HY-25	The use of C-Mobile service X increases comfort		X						X	X		X			X			X			
HY-26	The use of C-Mobile service X reduces stress		X	X																	
HY-27	The use of C-Mobile service X increases awareness				X	X	X														
HY-28	The use of C-Mobile service X leads to (experience of) driving eco-friendly/energy-efficient										X										
HY-29	The use of C-Mobile service X reduces perceived travel time											X									
HY-30	The use of C-Mobile service X reduces comfort of car users											X									

City	RTM	MPA	UPA	RWW	RHW	EVW	SVW	WSP	GP	GLOSA	CTLP	FI	IVS	MTTA	PVD	EBL	CACC	SSVW	MAI	BSD
Barcelona				X	X	X	X			X		X	X	X	X				X	
Bilbao		X	X	X	X															X
Bordeaux		X	X	X	X	X	X	X	X	X		X	X	X	X	X	X	X	X	X
Copenhagen																				
Newcastle				X	X			X	X	X										X
North Brabant	X	X				X	X	X	X	X	X									X
Thessaloniki				X	X	X	X	X	X	X	X	X	X	X	X					
Vigo				X	X	X	X	X	X				X	X	X	X	X	X	X	

Mobility Behavior and Accessibility
Safety
Efficiency
Environmental Sustainability
Subjective Evaluation

Table 21. Preliminary list of hypotheses linked to the C-MobILE services and cities (only direct links meaning it might be reasonable to test not yet linked hypotheses for some services)