
Risto Öörni

Risto Öörni

A doctoral dissertation completed for the degree of Doctor of Science (Technology) to be defended, with the permission of the Aalto University School of Electrical Engineering, at a public examination held at the lecture hall TU1 of the school on 27 August 2019 at 12.

Aalto University
School of Electrical Engineering
Department of Communications and Networking
Supervising professor
Professor Jyri Hämäläinen, Aalto University, Finland

Thesis advisor
Professor Risto Kulmala, Traficon Ltd, Finland

Preliminary examiners
Professor Andras Varhelyi, Lund University, Sweden
Professor Luca Studer, Polytechnic University of Milan, Italy

Opponent
Professor Dago Antov, Tallinn University of Technology, Estonia

Aalto University publication series
DOCTORAL DISSERTATIONS 142/2019

© 2019 Risto Öörni

ISSN 1799-4934 (printed)
ISSN 1799-4942 (pdf)

Unigrafia Oy
Helsinki 2019

Finland
Abstract

Information on the technical functioning, impacts, interoperability and acceptance of ITS systems is crucial to making decisions on the deployment of these systems. The resources, time and amount of data available for evaluation studies of a potential ITS system are often more or less limited. A need for more agile evaluation methods has been identified. This study provides definitions for different agile characteristics of evaluation methods for ITS systems, measurable criteria for these characteristics, and a definition for an agile evaluation method. The agile characteristics were defined on the basis of a literature study taking into account the Agile Manifesto, literature on measuring the agility of software development methods, and the experience obtained from seven case studies. The agile characteristics found to be relevant for evaluation methods for ITS include flexibility, leanness, lightness and simplicity, responsiveness, learning, speed and sustainability. The descriptions of agile characteristics, criteria for measuring the agility of evaluation methods and the definition of an agile evaluation method were validated by applying them to seven case studies. The case studies involved evaluations of technical functioning, impacts, socio-economic benefits, interoperability and user acceptance of ITS systems. Of the seven evaluated case studies, five were found to be based on agile methods as such or with restrictions.

The study results indicated that agile methods for evaluating the impacts of ITS have many similar characteristics to expert assessment, such as limited requirements for time and resources. However, they can be expected to be more transparent and possibly also more accurate than expert assessment without supporting documentation. Agile methods for evaluating the technical functioning of ITS systems, based on observation of system behaviour, analysis of reliability and statistical analysis of study results, can be used to verify the operation of an ITS system or subsystem in realistic operating conditions. They can be applied also in situations in which no extensive up-to-date documentation on system requirements is produced or maintained, such as when an agile development model is used for developing an ITS system. The study results indicate that not all simple or simplistic evaluation methods are agile methods, as simplicity does not imply that the evaluation method has other essential characteristics of agility such as flexibility and leanness. More validation will likely be needed for descriptions of the agile characteristics of evaluation methods, for the criteria for agility, and for the definition of an agile evaluation method. Further analyses should also be carried out on how to employ agile development, testing and systems engineering methods in the development of ITS systems in an optimum way, and how agile evaluation methods could be connected to them.

Keywords Intelligent transportation systems, evaluation, agile, road safety, advanced driver assistance
Avainsanat: älykäs liikenne, arviointi, ketterä, liikenneturvallisuus, kuljetettaja tukijärjestelmät

ISSN (painettu): 1799-4934 ISSN (pdf): 1799-4942

Vaktuokausi: 25.03.2019 Viestimä: 27.08.2019

Kieli: Englanti

Monografia [ ] Artikkeli-älytöskirja [ ] Essee-älytöskirja

Sivumäärä: 194

This doctoral dissertation was completed at the School of Electrical Engineering of Aalto University. The research included in the case studies was carried out at VTT Technical Research Centre of Finland Ltd. In addition to VTT, the writing process of the dissertation was supported by the Education Fund.

I wish to express my thanks to the supervisor of the thesis, Prof. Jyri Hämäläinen, who guided the writing process with his valuable expertise, and Dr Edward Mutafungwa for his support and encouragement. I also wish to thank the instructor of the thesis, Prof. Risto Kulmala, who contributed to the dissertation through his valuable comments and encouragement to both start and complete the work.

The pre-examiners of the thesis, Prof. András Várhelyi from Lund University (Lunds Universitet) and Prof. Luca Studer from the Polytechnic University of Milan (Politecnico di Milano), contributed to the thesis through their comments. I wish to thank them for their work. In addition, I wish to thank the opponent, Prof. Dago Antov from Tallinn University of Technology (Tallinna Tehnikaülikool), for his efforts.

The case studies included in the dissertation would not have been possible without the support of my co-authors and everyone else who contributed to the projects to which the case studies belonged. Especially, I wish to thank Dr Timo Laakko, who implemented the VTT eCall testbed, my VTT colleagues who implemented a prototype of the in-vehicle warning system for railway level crossings, and Dr Evgeni Meilikhov, who brought his expertise to the study on the interoperability of eCall and ERA-GLONASS. I also wish to express my thanks to Mrs. Adelaide Lönnberg for editing the English language.

The process of writing a dissertation is a long journey. During this process, I received excellent support from family members and friends. Their patience, understanding and encouragement were with me during the thesis work and allowed the writing process to be brought to its conclusion. I especially wish to thank Meskerem for her kindness and support.

Espoo, 8 July 2019
Risto Öörni
Contents

Preface.......................................................................................................... 1
List of Publications...................................................................................... 5
Author's Contribution ............................................................................... 7
List of Abbreviations and Symbols............................................................. 9

1. Introduction.......................................................................................... 11
   1.1 Intelligent transport systems....................................................... 11
   1.2 Technical functioning, impacts and interoperability of ITS.... 13
   1.3 Agile methods .............................................................................. 20
   1.4 Objectives of the thesis................................................................. 25
   1.5 Contributions ............................................................................... 26
   1.6 Structure of the thesis ................................................................. 27

2. Identification of agile characteristics and determining the agility of an evaluation method............................................................................... 29
   2.1 Overview....................................................................................... 29
   2.2 Identification of agile characteristics in ITS evaluation studies 29
   2.3 Determining the agility of an evaluation method...................... 45

3. Overview of current evaluation methods for ITS applications........ 47
   3.1 Frameworks for ITS deployment................................................ 47
   3.2 High-level requirements for assessment methods for ITS.... 48
      3.2.1 Requirements related to scientific research.................... 48
      3.2.2 Requirements from users of study results on ITS .......... 48
   3.3 Methods – technical evaluation.................................................. 50
   3.4 Methods – impact assessment.................................................... 52
   3.5 Methods – socio-economic assessment ..................................... 53
   3.6 Methods – interoperability ......................................................... 54
   3.7 Implementation issues ................................................................. 54

4. Case studies.......................................................................................... 57
   4.1 Model for the safety impacts of road weather information services available to road users and related socio-economic benefits ......................................................... 57
      4.1.1 ITS evaluation problem .................................................... 57
      4.1.2 Evaluation method............................................................ 58
      4.1.3 Experiences from using the method.............................. 60
      4.1.4 Agility assessment of the method .................................. 61
4.2 Reliability of an in-vehicle warning system for railway level crossings – a user-oriented analysis ........................................ 62
  4.2.1 ITS evaluation problem.................................................. 62
  4.2.2 Evaluation method ....................................................... 63
  4.2.3 Experiences from using the method ............................... 66
  4.2.4 Agility assessment of the method................................... 67
4.3 eCall minimum set of data transmission – results from a field test in Finland................................................................. 69
  4.3.1 ITS evaluation problem.................................................. 69
  4.3.2 Evaluation method ....................................................... 69
  4.3.3 Experiences from using the method ............................... 71
  4.3.4 Agility assessment of the method................................... 71
4.4 Interoperability of eCall and ERA-GLONASS in-vehicle emergency call systems.............................................................. 73
  4.4.1 ITS evaluation problem.................................................. 73
  4.4.2 Evaluation method ....................................................... 73
  4.4.3 Experiences from using the method ............................... 74
  4.4.4 Agility assessment of the method................................... 74
4.5 Demand for intelligent vehicle safety systems in Europe........ 76
  4.5.1 ITS evaluation problem.................................................. 76
  4.5.2 Evaluation method ....................................................... 76
  4.5.3 Experiences from using the method ............................... 77
  4.5.4 Agility assessment of the method................................... 78
4.6 Realised safety impacts of electronic stability control in Finland ......................................................................................... 80
  4.6.1 ITS evaluation problem.................................................. 80
  4.6.2 Evaluation method ....................................................... 80
  4.6.3 Experiences from using the method ............................... 82
  4.6.4 Agility assessment of the method................................... 83
4.7 Early adopters of emergency braking and speed alert .......... 85
  4.7.1 ITS evaluation problem.................................................. 85
  4.7.2 Evaluation method ....................................................... 86
  4.7.3 Experiences from using the method ............................... 87
  4.7.4 Agility assessment of the method................................... 88
5. Conclusions.................................................................................... 91
  5.1 Answers to the research questions......................................... 91
  5.2 Contribution to scientific knowledge ..................................... 96
  5.3 Implications to practice ...................................................... 96
  5.4 Discussion ........................................................................... 97
  5.5 Need for further research ................................................... 99
References ................................................................................................ 101
Publications............................................................................................ 113
List of Publications

This thesis is based on the following original publications which are referred to in the text as I–VII.


Author’s Contribution

**Publication I:** Model for the safety impacts of road weather information services available to road users and related socio-economic benefits

The author had the leading role in writing article I. Dr Risto Kulmala provided the original idea and the first draft version of the mathematical model presented in the article.

**Publication II:** Reliability of an in-vehicle warning system for railway level crossings – a user-oriented analysis

The author of the thesis is the only author of article II. The author thanks Ms. Marita Hietikko and the peer reviewers of the article for their valuable comments during the writing process and Mr. Antti Seise for implementation of the video monitoring system.

**Publication III:** eCall minimum set of data transmission – results from a field test in Finland

The author had the leading role in writing article III. Dr Timo O. Korhonen contributed to the article by helping to revise it during the peer review process.

**Publication IV:** Interoperability of eCall and ERA-GLONASS in-vehicle emergency call systems

Article IV is based on contributions from three authors. The author of the thesis had the leading role in the writing process, defined the analysis methods, and had the main responsibility for the analysis of interoperability. Dr Evgeni Meilikhov contributed to the article with his valuable expertise related to ERA-GLONASS, and Dr Timo O. Korhonen provided comments and revisions which helped to improve the article before submission to IET ITS.

**Publication V:** Demand for intelligent vehicle safety systems in Europe

The author of the thesis is the sole author of article V. Ms. Merja Penttinen participated in the work related to article V by designing the questionnaire form used for data collection in the iMobility Challenge study on car users’ awareness and demand for the systems.
**Publication VI:** Realised safety impacts of electronic stability control in Finland

The author had the main responsibility for the work in article VI. Specifically, he designed and conducted the main analyses and reported the results. Dr Juha Luoma contributed to the article with the analysis of ESC fitment data in new registered vehicles and related analysis of vehicle kilometrage data.

**Publication VII:** Early adopters of emergency braking and speed alert

The author of the thesis had the leading role in preparing article VII. Ms. Fanny Malin contributed to the introduction based on the summary of her Master's Thesis. Ms. Merja Penttinen participated in the work related to article VII by designing the questionnaire form used for data collection.
<table>
<thead>
<tr>
<th>Abbr.</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>3G</td>
<td>Third generation</td>
</tr>
<tr>
<td>4-DAT</td>
<td>4-Dimensional Analytical Tool</td>
</tr>
<tr>
<td>ADAS</td>
<td>Advanced Driver Assistance Systems</td>
</tr>
<tr>
<td>AMITRAN</td>
<td>CO₂ Assessment Methodology for ICT in Transport</td>
</tr>
<tr>
<td>ARC-IT</td>
<td>Architecture Reference for Cooperative and Intelligent Transportation</td>
</tr>
<tr>
<td>CO₂</td>
<td>Carbon dioxide</td>
</tr>
<tr>
<td>ERA</td>
<td>European Railway Agency</td>
</tr>
<tr>
<td>ERTRAC</td>
<td>European Road Transport Research Advisory Council</td>
</tr>
<tr>
<td>ERA-GLONASS</td>
<td>Russian in-vehicle emergency call system</td>
</tr>
<tr>
<td>ESC</td>
<td>Electronic Stability Control</td>
</tr>
<tr>
<td>ETSI</td>
<td>European Telecommunications Standards Institute</td>
</tr>
<tr>
<td>EU</td>
<td>European Union</td>
</tr>
<tr>
<td>Euro NCAP</td>
<td>European New Car Assessment Programme</td>
</tr>
<tr>
<td>FRAME</td>
<td>European ITS Framework Architecture</td>
</tr>
<tr>
<td>GDP</td>
<td>Gross Domestic Product</td>
</tr>
<tr>
<td>GNSS</td>
<td>Global Navigation Satellite System</td>
</tr>
<tr>
<td>GPRS</td>
<td>General Packet Radio Service</td>
</tr>
<tr>
<td>GPS</td>
<td>Global Positioning System</td>
</tr>
<tr>
<td>GSM</td>
<td>Global System for Mobile Communications</td>
</tr>
<tr>
<td>HeERO</td>
<td>Harmonised eCall European Pilot</td>
</tr>
<tr>
<td>ICT</td>
<td>Information and Communications Technologies</td>
</tr>
<tr>
<td>IEC</td>
<td>International Electrotechnical Commission</td>
</tr>
<tr>
<td>IEEE</td>
<td>Institute of Electrical and Electronics Engineers</td>
</tr>
<tr>
<td>IP</td>
<td>Internet Protocol</td>
</tr>
<tr>
<td>Acronym</td>
<td>Full Form</td>
</tr>
<tr>
<td>---------</td>
<td>-----------</td>
</tr>
<tr>
<td>ISA</td>
<td>Intelligent Speed Adaptation</td>
</tr>
<tr>
<td>ISO</td>
<td>International Organization for Standardization</td>
</tr>
<tr>
<td>IVSS</td>
<td>Intelligent Vehicle Safety Systems</td>
</tr>
<tr>
<td>ITS</td>
<td>Intelligent Transport Systems</td>
</tr>
<tr>
<td>ITU</td>
<td>International Telecommunication Union</td>
</tr>
<tr>
<td>ITU-T</td>
<td>ITU Telecommunication Standardization Sector</td>
</tr>
<tr>
<td>IVS</td>
<td>In-Vehicle System</td>
</tr>
<tr>
<td>MSD</td>
<td>Minimum Set of Data</td>
</tr>
<tr>
<td>MTBF</td>
<td>Mean Time Between Failures</td>
</tr>
<tr>
<td>MTTR</td>
<td>Mean Time to Repair</td>
</tr>
<tr>
<td>NASA</td>
<td>National Aeronautics and Space Administration</td>
</tr>
<tr>
<td>PReVAL</td>
<td>Evaluation of PReVENT safety functions</td>
</tr>
<tr>
<td>PReVENT</td>
<td>Preventive and Active Safety</td>
</tr>
<tr>
<td>PSAP</td>
<td>Public Safety Answering Point</td>
</tr>
<tr>
<td>RQ</td>
<td>Research Question</td>
</tr>
<tr>
<td>RTTI</td>
<td>Real-Time Traffic Information</td>
</tr>
<tr>
<td>TC</td>
<td>Technical Committee</td>
</tr>
<tr>
<td>TCP</td>
<td>Transmission Control Protocol</td>
</tr>
<tr>
<td>TS</td>
<td>Technical Specification</td>
</tr>
<tr>
<td>TRL</td>
<td>Technical Readiness Level</td>
</tr>
<tr>
<td>VMS</td>
<td>Variable Message Sign</td>
</tr>
<tr>
<td>XP</td>
<td>Extreme Programming</td>
</tr>
<tr>
<td>US DOT</td>
<td>United States Department of Transportation</td>
</tr>
</tbody>
</table>
1. Introduction

1.1 Intelligent transport systems

Several definitions exist for the term Intelligent Transport Systems (ITS). The European Telecommunications Standards Institute (ETSI) defines ITS from both a technical and service point of view (ETSI 2015):

“Intelligent Transport Systems (ITS) add information and communications technology to transport infrastructures and vehicles in an effort to improve their safety, reliability, efficiency and quality.

ITS services are also designed to optimize transport times and fuel consumption thus providing greener and safer transportation. However, the deployment of Intelligent Transport Systems and the provision of corresponding services are not limited to the road transport sector only, but include other domains such as railways, aviation and maritime as well.”

The definition of ITS provided by the IEEE ITS society captures the same meaning but in shorter form (IEEE Intelligent Transportation Systems Society 2015):

“Intelligent Transportation Systems (ITS) are those utilizing synergistic technologies and systems engineering concepts to develop and improve transportation systems of all kinds.”

Both of these definitions highlight that ITS makes use of technologies developed by other engineering sciences. Thus, the theoretical basis of ITS is largely based on other research fields.

ITS applications can be classified by variables such as transport mode, provided functionalities, required level of connectivity or level of automation. An overview of ITS from a communications point of view is illustrated in Figure 1 (originally provided by ETSI).
ITS applications residing in vehicles, roadside, back-office systems or on a platform carried by the traveller can be classified into cooperative and stand-alone systems. While stand-alone systems implement their functionalities without communicating with the outside world, cooperative systems employ communication between ITS stations located in personal, vehicle, roadside and central ITS subsystems (Figure 2).

Figure 1. Types of ITS and related technologies (ETSI TC ITS 2015).\(^1\)

Figure 2. ITS subsystems (adapted from ETSI 2010).

\(^1\) ©European Telecommunications Standards Institute 2012. Further use, modification, copy and/or distribution are strictly prohibited.
Some ITS applications have been identified as priority actions in the European ITS Directive (European Commission 2010). The priority actions include the provision of EU-wide multimodal travel information services, provision of EU-wide real-time traffic information (RTTI) services, data and procedures required for providing road safety-related minimum universal traffic information, harmonised provision for an interoperable EU-wide eCall, and provision of information services for safe and secure parking places for trucks and commercial vehicles.

The deployment of ITS can be accelerated or hampered by several factors. Known barriers to deployment include car users’ low willingness to accept additional costs (Trommer and Höltl 2012), lack of standardisation, lack of interoperability between transport modes and countries, liability issues, and financial and institutional issues (European Commission 2011). These may lead to regional fragmentation of ITS applications or lack of continuity of service. One of the challenges is also uncertainty over the impacts and benefits of a specific ITS application in a given traffic environment. For these reasons, methods for effective evaluation of technical functioning, impacts, socio-economic costs, benefits and interoperability of ITS technologies or individual systems are required.

1.2 Technical functioning, impacts and interoperability of ITS

The priorities for deployment of ITS systems in Europe are described in the ITS Roadmap Outline of the European Commission (European Commission 2007). According to the roadmap, priority should be given in ITS deployment in Europe to a limited number of ITS systems and components. The systems to be prioritised should meet the following criteria (European Commission 2007):

- They contribute “clearly to one or more of the high-level objectives” mentioned in the roadmap outline
- They are “mature or close to reaching maturity”
- They “present a clear benefit to the society and citizen”
- They “have the capability to be rolled out in a consistent manner across Europe”
- They “preferably offer synergies with other applications.”

The high-level objectives in the list refer to the policy objectives mentioned in the ITS Roadmap Outline (European Commission 2007). These include, but are not limited to, safe, efficient and clean transport. Policy objectives for ITS have also been summarised in the European ITS Action Plan. The main policy objectives for ITS mentioned there include “cleaner, more efficient, including energy efficient, safer and more secure transport” (European Commission 2008).

The policy objectives for deployment of ITS in urban environments are summarised in the European Commission working document “Mobilising Intelli-
The policy objectives listed for ITS in urban environments include:

- **Reduce Congestion**
- **Decrease parking pressure**
- **Modal Shift / Increase attractiveness of public transport**
- **Reduce energy consumption / emissions**
- **Enhance road safety**
- **Facilitate freight delivery & servicing**
- **Increase efficiency of transport system**
- **Improve quality of life in cities.**

ITS systems can contribute to policy objectives when they have relevant impacts on transport systems and their users. The relationship between policy objectives and the impacts of ITS systems is illustrated in Figure 3.

### Figure 3. Relationship between ITS functions, target of impacts, main impacts and policy objectives – an example (Kulmala et al. 2002).

The impacts of ITS are realised when the systems are deployed in the transport system. The deployment of ITS systems is affected by the decisions of public authorities acting as regulators of the transport system, infrastructure managers which may represent the public sector or private stakeholders, vehicle manufacturers and their suppliers, companies developing ITS services and products, telecom operators, and other stakeholders.

The decision factors affecting ITS deployment have been studied by Bunch et al. (2012). The results of their study were later supplemented with interviews, stakeholder workshops, site visits and longitudinal observations of ITS deployment in the US (Shah et al. 2013). According to the results, the decisions to implement, maintain, contract or discontinue ITS systems are affected by technology or application-related factors, implementer-related factors, external environment factors and user or market-related factors. The technology-related factors include price, technology readiness and maturity, benefits, compatibility with existing systems, standards, quality and reliability. Implementer factors such as perception of risk, level of knowledge and expertise,
Introduction

organisational structure and adoption by peers are related to the organisation making the deployment decision. External environmental factors include the political environment, budgetary constraints, priorities defined for the organisation, as well as the presence of a regional architecture for the system or a stakeholder pushing the system implementation forward. User and market-related factors include general acceptance towards new technologies as well as understanding the potential benefits and risks. These results are applicable to public sector stakeholders.

The deployment of an ITS system may be guided by an ITS strategy. The terminology used in the definition of an ITS strategy is illustrated in Figure 4. The ITS strategy typically consists of individual actions which may form one or more steps that make up implementation strategies for individual ITS systems.

Figure 4. Terminology in the definition of an ITS strategy, adapted from (Bekiaris et al. 2004, Figure 2).

The development of an ITS strategy also includes the development of the most likely implementation scenarios and the formation of policy recommendations. An example of this process is shown in Figure 5.

---

So far, the role of the regulator has been significant, especially in the case of systems which aim to improve traffic safety. For example, electronic stability control (ESC) (European Commission 2009) and eCall, the European in-vehicle emergency call system (European Commission 2015), have been made mandatory in new vehicle models (M1 and N1 class vehicles, cars and vans) with regulation established at the European level.

The full benefits of an ITS system cannot be realised without implementation of the system that has reached a sufficient level of technological maturity, meets the quality requirements relevant for the system, and corresponds to users’ needs. In summary, there is a need to determine which of the existing ITS systems or services are mature enough for deployment, what are the impacts of the systems, and whether these impacts contribute to the defined policy objectives such as safety or efficiency of the transport system.

The development of a technological innovation from scientific observations to a functional and verified product or system may involve several intermediate steps with different levels of technological maturity. The technological ma-
turity of a product or technology can be measured with Technical Readiness Levels (TRLs). The history and evolution of TRLs has been summarised by Mankins (2009). TRLs were originally developed by NASA for assessment of maturity of new technologies. The concept of TRLs was originally described in the 1970s, and the first standardised definitions of each level were provided in the 1990s. Later, TRLs were applied also in the transport sector (Towery, Machek and Thomas 2017) and in the Horizon 2020 Research Programme of the European Union (European Commission 2017).

Different sources provide slightly different definitions for TRLs. The definitions provided by NASA (2010) and the European Commission (2017) are summarised in Table 1.

**Table 1. Definitions of technical readiness levels (NASA 2010, European Commission 2017).**

<table>
<thead>
<tr>
<th>TRL</th>
<th>NASA</th>
<th>European Commission</th>
</tr>
</thead>
<tbody>
<tr>
<td>TRL1</td>
<td>Basic principles observed and reported</td>
<td>Basic principles observed</td>
</tr>
<tr>
<td>TRL2</td>
<td>Technology concept and/or application formulated</td>
<td>Technology concept formulated</td>
</tr>
<tr>
<td>TRL3</td>
<td>Analytical and experimental critical function and/or characteristic proof of concept</td>
<td>Experimental proof of concept</td>
</tr>
<tr>
<td>TRL4</td>
<td>Component and/or breadboard validation in laboratory environment</td>
<td>Technology validated in lab</td>
</tr>
<tr>
<td>TRL5</td>
<td>Component and/or breadboard validation in relevant environment</td>
<td>Technology validated in relevant environment (industrially relevant environment in the case of key enabling technologies)</td>
</tr>
<tr>
<td>TRL6</td>
<td>System/sub-system model or prototype demonstration in an operational environment</td>
<td>Technology demonstrated in relevant environment (industrially relevant environment in the case of key enabling technologies)</td>
</tr>
<tr>
<td>TRL7</td>
<td>System prototype demonstration in an operational environment</td>
<td>System prototype demonstration in operational environment</td>
</tr>
<tr>
<td>TRL8</td>
<td>Actual system completed and &quot;flight qualified&quot; through test and demonstration</td>
<td>System complete and qualified</td>
</tr>
<tr>
<td>TRL9</td>
<td>Actual system flight proven through successful mission operations</td>
<td>Actual system proven in operational environment (competitive manufacturing in the case of key enabling technologies; or in space)</td>
</tr>
</tbody>
</table>

The NASA documentation of TRLs also includes descriptions of TRLs for both hardware and software and exit criteria for each TRL (Table 2).

**Table 2. Exit criteria for technical readiness levels (NASA 2010).**

<table>
<thead>
<tr>
<th>TRL</th>
<th>Exit criteria</th>
</tr>
</thead>
<tbody>
<tr>
<td>TRL1</td>
<td>Peer reviewed publication of research underlying the proposed concept/application</td>
</tr>
<tr>
<td>TRL2</td>
<td>Documented description of the application/concept that addresses feasibility and benefit</td>
</tr>
<tr>
<td>TRL3</td>
<td>Documented analytical/experimental results validating predictions of key parameters</td>
</tr>
<tr>
<td>TRL4</td>
<td>Documented test performance demonstrating agreement with analytical predictions. Documented definition of relevant environment</td>
</tr>
<tr>
<td>TRL5</td>
<td>Documented test performance demonstrating agreement with analytical predictions. Documented definition of scaling requirements</td>
</tr>
<tr>
<td>TRL6</td>
<td>Documented test performance demonstrating agreement with analytical predictions</td>
</tr>
<tr>
<td>TRL7</td>
<td>Documented test performance demonstrating agreement with analytical predictions</td>
</tr>
<tr>
<td>TRL8</td>
<td>Documented test performance verifying analytical predictions</td>
</tr>
<tr>
<td>TRL9</td>
<td>Documented mission operational results</td>
</tr>
</tbody>
</table>

In practice, technical readiness levels are also applied to research and development projects (European Commission 2017). TRLs also have their limi-
tions: they do not convey information on the resources required to develop an entity from one TRL to another (Kujawski 2013).

The development of an ITS system from a TRL to an upper one includes observations about the technical functioning of the system. In the simplest case, this may be a single observation of whether the system performance was in line with analytical predictions (Table 2) with a binary success or fail criterion. Many ITS systems and environments in which they operate are inherently complex. For example, an ITS system may consist of several hardware and software components and data links. In these cases, a single observation of successful or failed operation of the system is unlikely to give meaningful information about the technical functioning of the system in a real operational environment. It is therefore necessary to study the probability that the system will function correctly in a defined operational environment. In other words, it is necessary to analyse the reliability of the system. The reliability of a system or component can be defined in different ways (Naresky 1970; Abbott 1971; Rausand and Høyland 2004). For example, reliability can be defined as the “ability of a system or component to perform its required functions under stated conditions for a specified period of time” (IEEE 2017). Reliability can also be measured with various indicators such as “number of failures per time unit (failure rate)” or “the probability that the item does not fail in a time interval \( (0, t] \) (survival probability)” (Rausand and Høyland 2004).

In addition to reliability, there are other quality attributes that are relevant for evaluating the quality and technological maturity of an ITS system. These include, but are not limited to, safety and the quality of information provided by the system. In addition, interoperability with other systems and between implementations based on the same specification has an impact on the possibilities of successful implementation of the system.

The importance of interoperability in the deployment of ITS systems is underlined in the European ITS Directive (European Commission 2010). According to the Directive, “ITS should build on interoperable systems which are based on open and public standards.”

The European ITS Directive defines interoperability as

“the capacity of systems and the underlying business processes to exchange data and to share information and knowledge.”

The International Telecommunication Union (ITU) provides the following definition for interoperability:

“The ability of two or more systems or applications to exchange information and to mutually use the information that has been exchanged.” (ITU-T 2000)

In addition to reliability, interoperability and other characteristics related to technical functioning, evaluation studies are needed to provide information on the impacts and socio-economic profitability of ITS technologies. This information is used to support decision making related to the adoption of new ITS technologies, maintenance of existing ITS systems and services, and the discontinuation of existing ITS systems. However, resources such as funding and
manpower availability for analysing ITS technologies and projects and collecting the required data are finite. In addition, constraints to evaluation studies may be imposed by the timeframe available for the task or the availability of data.

Perceived high costs may act as a barrier to initiating an evaluation study of an already realised ITS system or ITS project. If the costs are perceived to be too high, no evaluation study on the impacts, benefits and costs will be done once an ITS system or ITS project has been implemented. In other words, there is a need for methods to evaluate the impacts, costs and benefits of ITS that can be implemented with limited resources and within a short timeframe. The need for such methods is also related to other aspects of ITS such as interoperability and demand for the systems. It is therefore important to explore how the evaluation of impacts, technical functioning and interoperability of ITS systems could be made more affordable and efficient and less time-consuming. Evaluations of the impacts of ITS may utilise different methods and data sources. In this case, the concept of an evaluation method also includes the study design and methods used for data collection.

The need to evaluate the impacts, benefits and costs of ITS was identified a long time ago. The methodology developed by the Transport Research Laboratory (Perrett et al. 1996) was described in 1996 and was used to study the costs, benefits and socio-economic cost-effectiveness of several ITS applications in the same report. The described evaluation process (Table 3) is an example of a method developed to estimate the benefits and costs of ITS systems with finite resources and data available from earlier studies. A closely similar methodology was presented in a later article as a framework for carrying out a socio-economic cost-benefit analysis for an ITS system (Stevens 2004).

Table 3. Transport Research Laboratory methodology (Perrett et al. 1996).

<table>
<thead>
<tr>
<th>TRL methodology</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Define assessment objectives</td>
</tr>
<tr>
<td>• Describe system characteristics</td>
</tr>
<tr>
<td>• Define assumptions concerning the policy and technological contexts</td>
</tr>
<tr>
<td>• Identify impacts</td>
</tr>
<tr>
<td>• Select appropriate indicators</td>
</tr>
<tr>
<td>• Estimate effects on indicators</td>
</tr>
<tr>
<td>• Apply ‘standard’ values</td>
</tr>
<tr>
<td>• Analysis and results</td>
</tr>
</tbody>
</table>

Guidelines for field operational tests and related assessment of impacts and socio-economic profitability of in-vehicle ITS systems are provided in the FESTA handbook (Field opErational teSt supporT Action) (Barnard et al. 2017). Guidelines for evaluation of ITS applications have been provided also in other documents. For example, the C-Roads Evaluation and Assessment Plan provided evaluation guidelines for a number of cooperative ITS (C-ITS) pilots connected to the European C-Roads platform (Studer et al. 2018).

The FESTA handbook describes “the whole process of planning, preparing, executing, analysing and reporting an FOT [field operational test]”. The described approach to planning and implementation is largely similar to the V model of system development and involves a linear process from identification
of ITS functions to impact assessment and socio-economic cost-benefit analysis. The process described involves a number of steps following each other in a linear manner such as planning of use cases, planning of research questions, selection of key performance indicators, data acquisition and other implementation of the FOT followed by data analysis, impact assessment and socio-economic analysis.

The approach in the FESTA handbook has been criticised in a recent paper by Hill et al. (2018). The identified limitations of the handbook’s methodology include a lack of feedback, linear structure of the evaluation process, few possibilities for making changes during the evaluation process, and the prescriptive nature of the methodology which leads to a lack of focus on the needs of its users. The authors also identified a need to develop an agile common evaluation methodology (CEM) for evaluating the impacts of C-ITS. The expected benefits of this more agile evaluation method include the ability to generate and take into account feedback during the evaluation process, improvements in the assessment of individual components of ITS systems, and more flexibility in system deployment. These benefits were expected to lead to improved investment decisions and reduced waste of resources during ITS development and deployment. (Hill et al. 2018)

However, socio-economic cost-benefit analysis or analysis of the impacts of the system is not in itself a sufficient basis for deciding to invest in a new ITS system; the decision-maker also needs to understand the main risks of the project. These include, but are not limited to, technical, organisational, regulatory and market risks. Cost-benefit analysis of ITS applications also involves certain challenges and limitations. Empirical data based on earlier experiences with the systems over a sufficient time period is not always available. Defining the scope and appropriate level of detail of the evaluation may also be a challenge. A detailed evaluation of the impacts of the system is costly, while evaluation studies carried out with financial or time constraints may be lacking important details. Many evaluation studies also focus on the intended effects of the system instead of potential unwanted negative effects. In addition, there is possibly a publication bias towards reporting successful deployments of ITS systems instead of negative experiences (Stevens 2004): “A related issue is that successful implementations may be well-publicised but, in general, disbenefits are less well reported. One motivation for this may be to not adversely influence investments in future ITS projects.” Cost-benefit analysis is also dependent on valuation of impacts. This may be a challenge when there is no consensus on the valuation technique, or even impossible when no suitable valuation method exists, e.g. for distributional effects and social equity. (Stevens 2004)

### 1.3 Agile methods

The need for evaluation methods that provide results with limited resources and timeframe arises from both the characteristics of stakeholders making decisions on ITS deployment and the characteristics of the systems them-
selves. First, the resources to be used for an ex-ante (before system implementation) evaluation of an ITS system need to be proportional to the expected benefits of the ITS system or project. In practice, decisions on the deployment of ITS systems are made on regional, national and local level. Especially when the evaluation study focuses on a small geographical area such as a small country or individual city, the resources for estimating the impacts or other essential characteristics of an ITS system may be limited. Moreover, decision makers may require a study on the impacts or socio-economic profitability of the system carried out for their own infrastructure or administrative region, even if corresponding studies have been carried out elsewhere. Second, ITS technologies typically change rapidly with developments in networking, computing and automation technologies. For example, implementations of real-time traffic information services are now different from implementations ten years ago. Changes in functionality and technical implementation of an ITS system may affect the comparability of results between studies or the relevance of older study results.

The characteristics of agile software development methods such as flexibility, speed and leanness (Qumer and Henderson-Sellers 2006a) are desirable also for evaluation methods for ITS systems. Evaluation methods with agile characteristics can therefore potentially be a way to make the evaluation studies of ITS less time-consuming, more affordable, more flexible and better able to answer the information needs of various stakeholders.

A dictionary of the English language, Merriam Webster, provides two definitions of “agile” as an adjective (https://www.merriam-webster.com/dictionary/agile, 5th May 2018):

“marked by ready ability to move with quick easy grace”
“having a quick resourceful and adaptable character”.

While the first of the definitions refers to physical movement, the latter includes also constructions and activities of the human mind. This latter definition connects agility with adaptability and ability to provide ways and means quickly.

The word “agile” has been used in multiple contexts in the engineering sciences. In addition to its use as an adjective, the word appears in compound terms like agile software development (Qumer and Henderson-Sellers 2006a), agile analytics (Earley 2014), agile manufacturing (Sanchez and Nagi 2010) and agile systems engineering (Stelzmann 2012).

In software engineering, the word “agile” has been used to define agile software development. The principles and values of agile software development are defined in the Agile Manifesto (Beck et al. 2001) (Table 4).
Table 4. Principles and values of agile software development, a summary based on (Beck et al. 2001).

<table>
<thead>
<tr>
<th>Values of agile software development</th>
</tr>
</thead>
<tbody>
<tr>
<td>Individuals and interactions over processes and tools</td>
</tr>
<tr>
<td>Working software over comprehensive documentation</td>
</tr>
<tr>
<td>Customer collaboration over contract negotiation</td>
</tr>
<tr>
<td>Responding to change over following a plan</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Principles behind the Agile Manifesto</th>
</tr>
</thead>
<tbody>
<tr>
<td>Our highest priority is to satisfy the customer through early and continuous delivery of valuable software.</td>
</tr>
<tr>
<td>Welcome changing requirements, even late in development. Agile processes harness change for the customer's competitive advantage.</td>
</tr>
<tr>
<td>Deliver working software frequently, from a couple of weeks to a couple of months, with a preference for the shorter timescale.</td>
</tr>
<tr>
<td>Business people and developers must work together daily throughout the project.</td>
</tr>
<tr>
<td>Build projects around motivated individuals. Give them the environment and support they need, and trust them to get the job done.</td>
</tr>
<tr>
<td>The most efficient and effective method of conveying information to and within a development team is face-to-face conversation.</td>
</tr>
<tr>
<td>Working software is the primary measure of progress.</td>
</tr>
<tr>
<td>Agile processes promote sustainable development. The sponsors, developers, and users should be able to maintain a constant pace indefinitely.</td>
</tr>
<tr>
<td>Continuous attention to technical excellence and good design enhances agility.</td>
</tr>
<tr>
<td>Simplicity—the art of maximizing the amount of work not done—is essential.</td>
</tr>
<tr>
<td>The best architectures, requirements, and designs emerge from self-organising teams.</td>
</tr>
<tr>
<td>At regular intervals, the team reflects on how to become more effective, then tunes and adjusts its behaviour accordingly.</td>
</tr>
</tbody>
</table>

While the Agile Manifesto describes the principles and values of agile software development, it does not provide a comprehensive definition of the word “agile” or agile software development. A definition of “agility” in the context of software development methods has been suggested in a conference paper by Qumer and Henderson-Sellers (2006a):

“Agility is a persistent behaviour or ability of a sensitive entity that exhibits flexibility to accommodate expected or unexpected changes rapidly, follows the shortest time span, uses economical, simple and quality instruments in a dynamic environment and applies updated prior knowledge and experience to learn from the internal and external environment.”

In the context of software development, methods having agile characteristics have been employed to reduce the cost of development work and shorten the time needed for development. The characteristics of agile software development methods include flexibility, speed, leaness, learning and responsiveness (Qumer and Henderson-Sellers 2006a). In addition to software development, agile principles have been applied to systems engineering (Stelzmann 2012) to develop products that consist of both software and hardware. An agile development cycle has been defined as part of this work (Figure 6).
The Agile Manifesto or related publications do not define or prescribe any single development method which would be applicable to all kinds of software development, systems engineering or other development tasks. Instead, there are multiple software development and systems engineering methods which more or less follow the principles defined in the Agile Manifesto.

The 4-DAT model provides a framework which has been developed to measure the agility of different software development methods. The 4-DAT model (Qumer and Henderson-Sellers 2006b) consists of four thematic areas: method scope, agility characterisation, agile values and software process (Table 5). An example of evaluating the agility of methods by using the model appears in a later paper by the same authors (Qumer and Henderson-Sellers 2008).

Table 5. 4-DAT model for measuring the agility of software development methods (Qumer and Henderson-Sellers 2006b); for an example of the use of the model, see (Qumer and Henderson-Sellers 2008).

<table>
<thead>
<tr>
<th>Method scope</th>
<th>Does the method specify support for small, medium or large projects (business or other)?</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Project Size</td>
<td>Does the method support small or large teams (single or multiple teams)?</td>
</tr>
<tr>
<td>2. Team Size</td>
<td>Which development style (iterative, rapid) does the method cover?</td>
</tr>
<tr>
<td>3. Development Style</td>
<td>Does the method specify code style (simple or complex)?</td>
</tr>
<tr>
<td>4. Code Style</td>
<td>Which technology environment (tools, compilers) does the method specify?</td>
</tr>
<tr>
<td>5. Technology Environment</td>
<td>Which physical environment (co-located or distributed) does the method specify?</td>
</tr>
<tr>
<td>6. Physical Environment</td>
<td>What type of business culture (collaborative, cooperative or non-collaborative) does the method specify?</td>
</tr>
<tr>
<td>7. Business Culture</td>
<td>Does the method specify abstraction mechanism (object-oriented, agent-oriented)?</td>
</tr>
<tr>
<td>8. Abstraction Mechanism</td>
<td>Does the method accommodate expected or unexpected changes?</td>
</tr>
<tr>
<td>Agility characterization</td>
<td>Does the method produce results quickly?</td>
</tr>
<tr>
<td>1. Flexibility</td>
<td>Does the method follow shortest time span, use economical, simple and quality instruments for production?</td>
</tr>
<tr>
<td>2. Speed</td>
<td>Does the method apply updated prior knowledge and experience to learn?</td>
</tr>
<tr>
<td>3. Leanness</td>
<td>Does the method exhibit sensitiveness?</td>
</tr>
</tbody>
</table>

Figure 6. Agile development process (adapted from Stelzmann 2012) © 2012 IEEE.
Introduction

<table>
<thead>
<tr>
<th>Agile values</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Individuals and interactions over processes and tools</td>
</tr>
<tr>
<td>2. Working software over comprehensive documentation</td>
</tr>
<tr>
<td>3. Customer collaboration over contract negotiation</td>
</tr>
<tr>
<td>4. Responding to change over following a plan</td>
</tr>
<tr>
<td>5. Keeping the process agile</td>
</tr>
<tr>
<td>6. Keeping the process cost effective</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Software process</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Development process</td>
</tr>
<tr>
<td>2. Project management process</td>
</tr>
<tr>
<td>3. Software configuration control process / support process</td>
</tr>
<tr>
<td>4. Process management process</td>
</tr>
</tbody>
</table>

Criteria for the evaluation of agility have been published as part of an evaluation framework for agile software development methods (Taromirad and Ramsin 2008a) (Table 6). The set of criteria was constructed based on the content of the Agile Manifesto and papers on agile characteristics of software development methods reviewed by the authors (Taromirad and Ramsin 2008a). Table 6 includes the criteria for agility, their descriptions and a proposal for an indicator for each criterion (Domain values).

| Table 6. Agility evaluation criteria (Taromirad and Ramsin 2008a), © 2008 IEEE. |
|-----------------|-----------------|
| **Criterion** | **Description** | **Domain values** |
| Speed | How quickly does the methodology produce results? | 1 / (iteration length (in days) + deployment interval (in days)) |
| Sustainability | Are the speed and quality maintained until the end? Are they controlled or monitored? | Yes (techniques), No |
| Flexibility | Are expected | unexpected changes captured and handled in the project? | Ratio of the number of supporting activities and practices to the total |
| Learning | Does the process "learn" from past projects and previous iterations? | Ratio of the number of supporting activities and practices to the total |
| Responsiveness | Does the method provide feedback? | Ratio of the number of supporting activities and practices to the total |
| Leanness | Does the method value shorter time spans, using economical and simple quality-assured means for production? | Ratio of the number of supporting activities and practices to the total |
| Lightness and simplicity | How light and simple is the development process? | 1 / process complexity |
| Technical quality | How is technical quality monitored and controlled during the development? | Techniques and metrics |
| Active user collaboration | How involved are the customers in the development process? | Related role(s) and responsibilities |

Agile methods in software development and system design, as well as the agile development process, have been characterised in a paper exploring the theoretical background of agile methods (Nerur and Balijepally 2007) (Table 7). According to the authors, “emergent metaphor of design’ in the table is manifest in the agile methods in today’s emerging software-development thinking”.

24
Table 7. Traditional and emerging perspectives of design (Nerur and Balijepally 2007).

<table>
<thead>
<tr>
<th></th>
<th>Traditional View of Design</th>
<th>Emergent Metaphor of Design</th>
</tr>
</thead>
<tbody>
<tr>
<td>Design process</td>
<td>Deliberate and formal, linear sequence of steps, separate formulation and implementation, rule driven</td>
<td>Emergent, iterative and exploratory, knowing and action inseparable, beyond formal rules</td>
</tr>
<tr>
<td>Goal</td>
<td>Optimisation</td>
<td>Adaptation, flexibility, responsiveness</td>
</tr>
<tr>
<td>Problem-solving approach</td>
<td>Selection of the best means to accomplish a given end through well-planned, formalised activities</td>
<td>Learning through experimentation and introspection, constantly reframing the problem and its solution</td>
</tr>
<tr>
<td>View of the environment</td>
<td>Stable, predictable</td>
<td>Turbulent, difficult to predict</td>
</tr>
<tr>
<td>Type of learning</td>
<td>Single-loop</td>
<td>adaptive</td>
</tr>
<tr>
<td>Key characteristics</td>
<td>Control and direction</td>
<td>Collaboration and communication – integrates weltanschauungs, or worldviews</td>
</tr>
<tr>
<td></td>
<td>Avoids conflict</td>
<td>Embraces conflict and dialectics</td>
</tr>
<tr>
<td></td>
<td>Formalises innovation</td>
<td>Encourages exploration and creativity and is opportunistic</td>
</tr>
<tr>
<td></td>
<td>Manager is controller</td>
<td>Manager is facilitator</td>
</tr>
<tr>
<td></td>
<td>Design precedes implementation</td>
<td>Design and implementation are inseparable and evolve iteratively</td>
</tr>
<tr>
<td>Rationality</td>
<td>Technical</td>
<td>functional</td>
</tr>
<tr>
<td>Theoretical (and</td>
<td>or) philosophical roots</td>
<td>Logical positivism, scientific method</td>
</tr>
</tbody>
</table>

The frameworks and characterisations presented earlier are not directly applicable to ITS evaluation studies and should be adapted to make them suitable for assessing the level of agility in such studies, as well as the methods applied in them. It is therefore not clear which methods for evaluation of impacts, technical functioning, interoperability and other aspects of ITS should be considered as agile or non-agile. Consequently, there is no experience on why and where agile evaluation methods and practices should be used for evaluation of ITS.

A relevant question is whether the agile principles identified in the domain of software development could make evaluation of ITS projects more efficient, more affordable and less time-consuming, and reduce the amount of data needed to provide useful and accurate results. It is also somewhat an open question to what extent agile principles are already present in existing evaluation studies for ITS and related working practices, and how the agile principles and agile methods could further contribute to developing better evaluation methods for technical functioning, impacts and interoperability of ITS systems.

1.4 Objectives of the thesis

Based on the gaps in earlier knowledge identified in Chapter 1.3, three research questions were defined for the study:

(1) Why and where should agile evaluation methods and practices be used?

---

(2) Can any simplified or simplistic approaches to the evaluation of impacts, technical performance, interoperability and related aspects of ITS be considered to be agile methods?

(3) How could agile principles contribute towards the development of better evaluation methods and tools for ITS?

These three research questions are approached here by presenting and analysing a set of case studies on the technical functioning, safety impacts, interoperability and user acceptance of ITS applications related to advanced driver assistance systems and connected driving. The case studies are analysed from the agility point of view which allows validation of the terminology related to agile evaluation methods for ITS applications developed on the basis of a literature study. Once the relevant terminology has been defined and validated, the defined research questions can be answered based on the empirical experiences described in the case studies.

Three main objectives corresponding to the research questions were defined for the study:

(1) To study to what extent the case studies and the methods applied in them have agile characteristics

(2) To verify practical evaluation approaches with agile characteristics for technical functioning, impacts, socio-economic benefits and interoperability of ITS applications and provide empirical experience on their use in selected case studies

(3) To study how agile principles could further contribute to the development of better evaluation methods and tools for ITS.

The first of the research questions corresponds to all three objectives. The second research question corresponds to objectives 1 and 2. Research question 3 corresponds to objective 3. The objectives were defined in alignment with the approach based on empirical case studies employed in the thesis.

The thesis aims to provide information that will be relevant also to practitioners outside the scientific community. The work is therefore expected to benefit several stakeholders such as safety regulators, infrastructure managers, system suppliers and other stakeholders investing in ITS, managing the related risks or developing regulation for ITS.

1.5 Contributions

The publications included in the thesis form a set of case studies that are analysed to answer the individual research questions. All case studies contribute to all the research questions and objectives described in Chapter 1.4. The case studies do not have significant dependencies on each other. The objectives of individual case studies and the methods used for data collection and analysis are summarised in Table 8.
Table 8. Articles included as case studies in the thesis – objectives and methods used for data collection and analysis.

<table>
<thead>
<tr>
<th>Publication</th>
<th>Objective</th>
<th>Methods used for data collection and analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td>Publication I</td>
<td>To develop a method for estimating the most likely impact of weather-related information and warning services on the safety of road users and related socio-economic benefits</td>
<td>Impact estimation based on a literature study and mathematical modelling based on the power model described by Nilsson (2004)</td>
</tr>
<tr>
<td>Publication II</td>
<td>To analyse the reliability of an in-vehicle warning system for railway level crossings from the user point of view</td>
<td>Video monitoring of the operation of mobile client and train movements at two level crossings combined with analysis of log files provided by the system, estimation of reliability measures</td>
</tr>
<tr>
<td>Publication III</td>
<td>To analyse the success rate of eCall minimum set of data (MSD) transmission</td>
<td>Test with an eCall in-vehicle system prototype and Finnish eCall testbed, calculation of MSD success rate and estimation of confidence intervals with statistical methods</td>
</tr>
<tr>
<td>Publication IV</td>
<td>To analyse the interoperability of pan-European eCall and the Russian ERA-GLONASS in-vehicle emergency call system</td>
<td>Review of specifications of interworking systems (eCall and ERA-GLONASS), using the framework provided by the ETSI approach for interoperability and definition of interoperability by ITU-T</td>
</tr>
<tr>
<td>Publication V</td>
<td>To analyse the demand for intelligent vehicle safety systems in Europe</td>
<td>Analysis of panel survey data collected in five European countries, estimation of demand curves for the systems based on linear and exponential regression</td>
</tr>
<tr>
<td>Publication VI</td>
<td>To estimate the safety impacts of electronic stability control (ESC) in the current situation in Finland with imperfect fleet penetration</td>
<td>Mathematical modelling based on estimation on the number of ESC equipped vehicles in use, their share of the vehicle fleet or vehicle kilometres travelled and estimates for safety effects of ESC obtained with a literature study</td>
</tr>
<tr>
<td>Publication VII</td>
<td>To analyse the characteristics of early adopters of emergency braking and speed alert in Europe</td>
<td>Analysis of panel survey data collected in five European countries, crosstabulation of demographic characteristics with system use, including chi-square tests and odds ratios with confidence intervals</td>
</tr>
</tbody>
</table>

1.6 Structure of the thesis

Chapter 1 provides a general introduction to ITS, technical functioning, impacts and interoperability of ITS and agile methods. In addition, it describes the objectives of the thesis, the contributions of the individual papers included and the structure of the thesis. Chapter 2 explains the identification of agile characteristics in evaluation methods for ITS, provides a definition for an agile evaluation method, and provides criteria for determining the agility and agile characteristics of evaluation methods. Chapter 3 gives an overview of existing evaluation methods for ITS systems and applications. Chapter 4 provides descriptions of the case studies and agile characteristics identified in the evaluation methods used in the case studies. Chapter 5 includes a discussion of the results, conclusions of the study and recommendations for further research.

The relationships between chapters and their contribution to answering the research questions are illustrated in Figure 7. The agile characteristics of evaluation methods for ITS, related assessment criteria and the definition of an agile evaluation method described in Chapter 2 were developed on the basis of a literature study. The results of the literature study are documented in Chapters 1 and 2. The definitions of agile characteristics, related assessment criteria
and the definition of an agile evaluation method were validated by applying them to seven case studies presented in Chapter 4.

The agile characteristics present in the case studies and the methods applied in those studies could be identified by applying the tools developed in Chapter 2 to a set of case studies described in Chapter 4. Chapter 4 provides practical evaluation approaches with agile characteristics for evaluation of impacts, technical functioning, socio-economic benefits, and interoperability of ITS. The evaluation methods used in the case studies have been verified by applying the methods to different ITS systems and the data available for the purpose and publishing the case studies as peer-reviewed articles. In addition, the empirical experiences of using the described methods with agile characteristics are described in Chapter 4.

A discussion on how agile principles could further contribute to the development of better evaluation methods and tools for ITS is provided in Chapter 5. This includes identification of the advantages and limitations of evaluation methods verified in the case studies and classified as agile methods and develops of recommendations for the use of agile evaluation methods and practices. The identified advantages and limitations are based on the experiences with evaluation methods classified as agile methods (Chapter 4) and a literature study providing an overview of evaluation methods for ITS (Chapter 3).
2. Identification of agile characteristics and determining the agility of an evaluation method

2.1 Overview

The research questions will be answered by analysing seven case studies. The case studies under analysis include seven evaluation studies related to the evaluation of impacts, technical functioning, interoperability or implementation of ITS. The material to be analysed includes published and peer-reviewed versions of the studies.

First, it was necessary to develop a way to identify and measure the agile characteristics present in the methods described in the selected case studies. Second, it was necessary to find a way to determine whether an evaluation method should be called an agile method or not.

2.2 Identification of agile characteristics in ITS evaluation studies

The characterisations of agility originally developed for software development methods, such as the agile principles explained in the Agile Manifesto, were not directly applicable as such to methods used in ITS evaluation studies. For example, ITS evaluation studies aim to produce new information, knowledge and research results on the ITS system or systems under analysis instead of a software product. It was therefore necessary to explore how the principles of the Agile Manifesto could be transposed to the evaluation of ITS applications.

Transposing the agile values and agile principles to ITS evaluation methods included two questions: which of the agile principles and agile values are relevant also for evaluation methods for ITS applications, and in which form are they applicable to evaluation methods for ITS applications? The Agile Manifesto presents the agile values and the principles behind the manifesto in a rather abstract form. It was therefore necessary to describe and understand in more detail how the agile values and agile principles are defined in the context of software development and manifested in practice in different agile software development methods. In other words, it was necessary to develop a more concrete understanding of the agile values and characteristics in the context of software development to allow identification of the corresponding characteristics, processes and other entities in the context of evaluation of ITS systems.
The 4-DAT model (Qumer and Henderson-Sellers 2006b) includes a four-dimensional characterisation of agile software development methods. The characterisation includes method scope characterisation, agility characterisation, agile values characterisation and software process characterisation. As part of this characterisation, the model provides descriptions for attributes related to the scope of the method, agile features of the method, agile values, and attributes related to the software process. For measuring agility, the 4-DAT model proposes five variables: flexibility, speed, leaness, learning and responsiveness. The descriptions in the 4-DAT model can be used to seek interpretations for agile principles and agile values in the context of evaluation of ITS applications.

Agile features of a software development method and their descriptions provided in the 4-DAT model are presented in Table 9. The table also summarises the relevancy of the agile features to evaluation methods for ITS applications, as well as descriptions of agile features developed for evaluation methods for ITS applications.

Table 9. Descriptions of agile features adapted for evaluation methods of ITS applications.

<table>
<thead>
<tr>
<th>Features (Qumer and Henderson-Sellers 2006b)</th>
<th>Description – software development methods (Qumer and Henderson-Sellers 2006b)</th>
<th>Relevancy for evaluation methods for ITS</th>
<th>Description – evaluation methods for ITS applications</th>
</tr>
</thead>
<tbody>
<tr>
<td>Flexibility</td>
<td>Does the method accommodate expected or unexpected changes?</td>
<td>Yes</td>
<td>Does the evaluation method accommodate expected or unexpected changes?</td>
</tr>
<tr>
<td>Speed</td>
<td>Does the method produce results quickly?</td>
<td>Yes</td>
<td>Does the method produce evaluation results quickly?</td>
</tr>
<tr>
<td>Leanness</td>
<td>Does the method follow shortest time span, use economical, simple and quality instruments for production?</td>
<td>Yes</td>
<td>Does the evaluation method follow the shortest time span, use economical, simple and quality means to create evaluation results?</td>
</tr>
<tr>
<td>Learning</td>
<td>Does the method apply updated prior knowledge and experience to learn?</td>
<td>Yes</td>
<td>Does the method apply updated prior knowledge and experience to learn?</td>
</tr>
<tr>
<td>Responsiveness</td>
<td>Does the method exhibit sensitivity?</td>
<td>Yes</td>
<td>Does the method provide feedback during the evaluation process?</td>
</tr>
</tbody>
</table>

In the 4-DAT model developed for the evaluation of agility of software development methods, flexibility has been conceptualised as an ability to accommodate expected or unexpected changes. The evaluation methods for ITS are applicable in a dynamic context which may include changes in technology, physical infrastructure of the transport system, user behaviour or regulation of transport system or ITS services. In addition to these more or less predictable changes, evaluation studies also involve uncertainty in the successfulness of data collection and control of possible confounding factors. It is therefore preferable for an evaluation method used for ITS applications to allow at least a certain level of flexibility. The question developed as a description for software development methods (Table 9) can be adapted with a minor change in form: does the evaluation method accommodate expected or unexpected changes?
The speed with which a software development method provides useful results has been identified as an agile characteristic by Qumer and Henderson-Sellers (2006b) and Taromirad and Ramsin (2008a), even though the word does not appear as such in the Agile Manifesto (Beck et al. 2001) or in the related principles for agile software. In the 4-DAT model, speed as an agile characteristic is determined by the question of whether the software development method produces results quickly. Taromirad and Ramsin (2008a) have defined speed as an agile characteristic with the question how quickly does the method produce the results, and have proposed a quantitative indicator (1 / (iteration length, in days + deployment interval, in days)). The description originally provided for software development methods can be applied to the evaluation of ITS systems with a minor change, by replacing the generic word “results” with a more specific expression “evaluation results” (Table 9).

Leanness is one of the agile features of a software development method identified in papers published by Qumer and Henderson-Sellers (2006b) and Taromirad and Ramsin (2008a). The word “leanness” does not appear as such in the Agile Manifesto or the related Principles for Agile Software (Beck et al. 2001), but the elements of the definition can be identified in the Agile Manifesto and in the literature on agile characteristics. The definition used for leanness by Qumer and Henderson-Sellers (2006b) (“-- shorter time spans, using economical and simple quality-assured means for production”) has been used in a modified form by Taromirad and Ramsin (2008a). The question used as a description of leanness is relevant for evaluation methods for ITS with minor adjustments. Instead of the words “methods” and “results”, the version of the question adapted to the evaluation of ITS uses the more specific terms “evaluation methods” and “evaluation results”.

Learning has been identified as an agile feature of a software development method by both Qumer and Henderson-Sellers (2006b) and Taromirad and Ramsin (2008a). These two papers provide slightly different definitions for learning as an agile characteristic. While Qumer and Henderson-Sellers (2006b) describe learning as “applying updated prior knowledge and experience to learn”, Taromirad and Ramsin (2008a) question whether the development process learns from the past and earlier iteration rounds. Observing the outcomes of the previous iteration round and learning from them is also a part of the agile development cycle described in the context of agile systems engineering (Stelzmann 2012). In the context of evaluation of ITS applications, learning should therefore be understood both as learning from earlier relevant studies and experience, as well as learning from the experience obtained during the same project or evaluation study. Learning is therefore an agile feature for which a meaningful interpretation can be found in the context of evaluation of ITS systems.

Responsiveness has been listed as an agile characteristic of a method by both Qumer and Henderson-Sellers (2006b) and Taromirad and Ramsin (2008a). The descriptions provided in these two papers are different. While Qumer and Henderson-Sellers (2006b) describe responsiveness with the question of whether the method exhibits sensitiveness, Taromirad and Ramsin conceptu-
alise responsiveness as provision of feedback. Even though the word “respon-
siveness” is not used in the Agile Manifesto, it can be understood as an expres-
sion of one of the agile values, “Individuals and interactions over processes
and tools” (Beck et al. 2001). In summary, responsiveness can be understood
as the ability of the development process to provide feedback and the interac-
tion between developers and the customer. In the context of evaluation of ITS
applications, the ability of the evaluation method to provide feedback will be
important in several foreseeable scenarios. First, an early indication of the
result may save valuable time. For example, in the case of evaluation of in-
teroperability, identification of an interoperability issue as early as possible
may allow the development of a fix to start even before the evaluation results
have been fully analysed and reported. Second, the ability of the evaluation
method and evaluation process to provide feedback during the project may
also be beneficial in situations where the evaluation results differ from antici-
pated, and the scope or objectives of the evaluation may need to be adjusted.
The capability to provide feedback during the evaluation process is therefore
an advantage for an evaluation method. In summary, responsiveness is an ag-
ile characteristic that can be expected to be relevant also in the context of eval-
uation of ITS systems.

Agile values characterising agile software development are expressed in the
Agile Manifesto. In addition, the Agile Manifesto includes principles for agile
software. The 4-DAT model provides descriptions for agile values applicable to
software development methods. The following table gives the descriptions of
agile values adapted for evaluation methods for ITS applications, as well as an
assessment of their relevance in the context of evaluation of ITS applications
(Table 10).

<table>
<thead>
<tr>
<th>Agile values (Qumer and Henderson-Sellers 2006b)</th>
<th>Description – software development methods (Qumer and Henderson-Sellers 2006b)</th>
<th>Relevancy for evaluation methods for ITS</th>
<th>Description – evaluation methods for ITS applications</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Individuals and interactions over processes and tools</td>
<td>Which practices value people and interaction over processes and tools?</td>
<td>Yes</td>
<td>Which practices value people and interactions over processes and tools?</td>
</tr>
<tr>
<td>2. Working software over comprehensive documentation</td>
<td>Which practices value working software over comprehensive documentation?</td>
<td>Yes</td>
<td>Which practices value relevant and sound evaluation results over comprehensive reporting?</td>
</tr>
<tr>
<td>3. Customer collaboration over contract negotiation</td>
<td>Which practices value customer collaboration over contract negotiation?</td>
<td>Yes</td>
<td>Which practices value customer collaboration over contract negotiation?</td>
</tr>
<tr>
<td>4. Responding to change over following a plan</td>
<td>Which practices value responding to change over following a plan?</td>
<td>Yes</td>
<td>Which practices value responding to change over following a plan?</td>
</tr>
<tr>
<td>5. Keeping the process agile</td>
<td>Which practices help in keeping the process agile?</td>
<td>Yes</td>
<td>Which practices help in keeping the evaluation process agile?</td>
</tr>
<tr>
<td>6. Keeping the process cost effective</td>
<td>Which practices help in keeping the process cost effective?</td>
<td>Yes</td>
<td>Which practices help in keeping the evaluation process cost-effective?</td>
</tr>
</tbody>
</table>

The preference regarding people and interactions is a principle that has a
meaningful interpretation from the agility point of view, also in the context of
evaluation of ITS applications. It is difficult to find one methodology for evaluation of impacts or technical performance of ITS that would be both documented in detail and applicable to all types of ITS systems; even for the assessment of safety impacts several impact mechanisms have been identified (summarised in Kulmala 2006 and Kulmala 2010, originally listed in Draskóczy et al. (ed.) 1998). While guidelines have been prepared for evaluation of impacts and socio-economic profitability of ITS systems (Kulmala et al. 2002), they need to be applied by a competent expert who will describe the evaluation process, data sources to be used and methods to be applied in a particular case.

The second of the agile values, “working software over comprehensive documentation”, was originally defined for software development. In the context of software development, the authors of the Agile Manifesto have highlighted the importance of the core product – functional software addressing the customer’s needs – over extensive documentation. When applied to the context of evaluation of ITS applications, this principle can be understood as preference for an outcome useful to the customer, ITS stakeholders and the scientific community over extensive reporting that may be related to research projects.

The third of the agile values, “customer collaboration over contract negotiation”, is also relevant for the evaluation of ITS systems. Although a contract with the customer is usually needed before starting an evaluation study, existing procurement and project management practices and distribution rules permit a certain level of flexibility. For example, it may be possible to use the project’s resources to continue the analysis beyond the contractual results of the project or to disseminate the results as scientific publications. Close collaboration with the customer is also needed when the plan originally developed for the evaluation study is found to be unrealistic or impossible to implement within the planned timeframe. Such a scenario can occur if the project involves development of new technology and therefore technical risk, where the quality of data available for analysis is poorer or better than expected, or an unanticipated event forces a change in the project plan. These factors may be difficult or impossible to anticipate in the contract negotiation phase of a project involving an evaluation study or studies.

The fourth of the agile values, “responding to a change over following a plan”, is relevant to the evaluation of ITS for partly the same reasons as the third agile value. The data collected during the evaluation study may be more or less useful for research purposes than originally anticipated, and a need to adjust the scope or objectives of the study may therefore arise.

The fifth of the agile values, “keeping the process agile”, was explained further by Qumer and Henderson-Sellers (2006b), who addressed the question of which practices help keeping the process agile. In software development, this refers to the agile development process. For example, the agile development process includes continuous feedback from customers to developers, which plays a key role in maintaining the agile nature of the process. In the case of agile systems engineering, this feedback and related interactions with the customer have been described as the agile feedback loop (Figure 6). The evaluat-
tion of ITS applications may also involve interactions between the researcher and customers, ITS stakeholders and the scientific community. Keeping the process agile is therefore a relevant part of the definition of agility also in the context of evaluation of ITS systems.

The sixth of the agile values, “keeping the process cost effective”, is also relevant to the evaluation of ITS systems. According to the EasyWay Handbook on Evaluation Best Practice (Tarry et al. 2012), the acceptable cost of an evaluation study will depend on the expected cost of the ITS project implemented or being planned. For large ITS systems with high costs, the handbook recommends reserving 1.0–1.5% of the project costs for performing an evaluation study. A higher percentage, even exceeding 10%, is recommended for ITS projects of small size and projects based on untested or unproven technology. This applies also to projects that serve as a first national demonstration of new technology before deployment on a larger scale. In summary, the cost of the evaluation needs to be proportional to the project size and expected benefits, as well as the type of application and the overall context, while the method should provide sound results answering the information needs present in its context.

Other agility evaluation criteria for software development methods have been provided by Taromirad and Ramsin (2008a) (Table 6). While these criteria are applicable even for assessment for non-agile software development methods (Taromirad and Ramsin 2008a), they have to adapted before they can be used for assessing the agility of evaluation methods for ITS (Table 11). The interpretations of several agile features, such as speed, flexibility, learning, leanness and responsiveness, as well as their relevancy in the contexts of evaluation of ITS systems, have already been discussed with the 4-DAT model. The evaluation criteria for agility provided by Taromirad and Ramsin (2008a) also include parts that are not directly included in the 4-DAT model. These include sustainability, lightness and simplicity, technical quality and active user collaboration.

In addition to the elements mentioned earlier, sustainability is among the agility evaluation criteria suggested by Taromirad and Ramsin (2008a). This can be traced back to one of the principles of agile software expressed together with the agile values in the Agile Manifesto: “Agile processes promote sustainable development. The sponsors, developers, and users should be able to maintain a constant pace indefinitely.” In practice, this refers to working practices that are sustainable over the long term. The Agile Alliance uses the term “sustainable pace”, which refers to the principle that the team should work only at a pace that can be sustained over an indefinite period (Agile Alliance 2018).

In the case of software development, this mostly rules out overtime and weekend work (Agile Alliance 2018). Software developers may experience periods of very high workload called crunches to deliver the planned result until a major milestone or release of a new software product is achieved (Ortega, Guzdial and Reed 2010; Edholm et al. 2017). In addition, a software project
Identification of agile characteristics and determining the agility of an evaluation method

may be in a state of continuous crunch covering most of the lifetime of the project, typically caused by unrealistic schedules (Edholm et al. 2017).

Table 11. Agility evaluation criteria (Taromirad and Ramsin 2008a), adapted for evaluation methods for ITS.

<table>
<thead>
<tr>
<th>Criterion (Taromirad and Ramsin 2008a)</th>
<th>Description – software development methods (Taromirad and Ramsin 2008a)</th>
<th>Relevancy for evaluation methods for ITS as a criterion for evaluation of agility</th>
<th>Characterising question – evaluation methods for ITS applications</th>
</tr>
</thead>
<tbody>
<tr>
<td>Speed</td>
<td>How quickly does the methodology produce results?</td>
<td>Yes</td>
<td>How quickly does the evaluation method produce results?</td>
</tr>
<tr>
<td>Sustainability</td>
<td>Are speed and quality maintained until the end? Are they controlled or monitored?</td>
<td>Yes</td>
<td>Are the speed of the evaluation process and quality of the results maintained over the long term? Are they controlled or monitored? Is the method future-proof?</td>
</tr>
<tr>
<td>Flexibility</td>
<td>Are expected</td>
<td>unexpected changes captured and handled in the project?</td>
<td>Yes</td>
</tr>
<tr>
<td>Learning</td>
<td>Does the process “learn” from past projects and previous iterations?</td>
<td>Yes</td>
<td>Does the evaluation process “learn” from past projects and previous iterations? In other words, does the method “learn” and improve itself when it is applied several times?</td>
</tr>
<tr>
<td>Responsiveness</td>
<td>Does the method provide feedback?</td>
<td>Yes</td>
<td>Does the evaluation provide feedback during the evaluation process?</td>
</tr>
<tr>
<td>Leanness</td>
<td>Does the method value shorter time spans, using economical and simple quality-assured means for production?</td>
<td>Yes</td>
<td>Does the evaluation method value shorter time spans, using economical and simple quality-assured means for producing results?</td>
</tr>
<tr>
<td>Lightness and simplicity</td>
<td>How light and simple is the development process?</td>
<td>Yes</td>
<td>How light and simple are the evaluation process and related procedures?</td>
</tr>
<tr>
<td>Technical quality</td>
<td>How is technical quality monitored and controlled during the development?</td>
<td>No</td>
<td>-</td>
</tr>
<tr>
<td>Active user collaboration</td>
<td>How involved are the customers in the development process?</td>
<td>No</td>
<td>-</td>
</tr>
</tbody>
</table>

In the case of evaluation of ITS systems, the possibility to maintain a sustainable pace of work is usually related to factors other than the evaluation method itself. These may be related, for instance, to time constraints imposed by the customer’s decision-making process, flexible or rigid project management practices of the organisation providing funding for the evaluation work or schedules and internal working practices of the organisations of the expert team. For this reason, the possibility to maintain a sustainable pace of work is not a sufficient interpretation of sustainability in the context of evaluation of ITS systems.

The Agile Manifesto formulates sustainability as promoting sustainable development and the ability of sponsors, developers and users to maintain a constant pace indefinitely. For evaluation methods for ITS, this includes also the anticipated usefulness of the method in the future, i.e. its future-proofness.
The Cambridge English dictionary defines the adjective **future-proof** as the ability of the system or equipment to retain its usefulness regardless of changes in technology: “Future-proof software, computer equipment, etc. is designed so that it can still be used even when technology changes” (https://dictionary.cambridge.org/dictionary/english/future-proof, accessed 5th July 2018). The Oxford English dictionary defines future-proof as the negation of obsolescence in the future: “(of a product or system) unlikely to become obsolete.” (https://en.oxforddictionaries.com/definition/future-proof, accessed 5th July 2018)

The future usefulness of an evaluation method may be affected by developments in scientific knowledge and by technological changes in systems or applications for which the method is intended. The value of an evaluation method may be significantly reduced, for example, if new scientific knowledge challenges the assumptions the method relies on. The method may even become unusable if its results cannot be expected to be accurate or fit-for-purpose any longer, and the model cannot be adapted without substantial difficulties taking the new knowledge or research results into account.

For example, the paradigms on road safety have changed over time (Hagenzieker, Commandeur and Bijleveld 2014; Hakkert and Gitelman 2014) and may continue to do so in the future. An evaluation method may also become obsolete through technological changes, if for example a research question addressed by the method is no longer relevant and there is little or no potential for re-use of the method for similar purposes. For example, estimation of the impacts of real-time traffic information (RTTI) or various ADAS systems on the behaviour of human drivers will cease to be relevant if all human-driven vehicles are replaced by fully autonomous ones.

Lightness and simplicity are mentioned in the set of agility assessment criteria proposed by Taromirad and Ramsin (2008a). In their definition, lightness and simplicity are understood as the inverse of the process complexity of the software development process. Simplicity is one of the principles of agile software attached to the Agile Manifesto (Beck et al. 2001), and it is defined there as “the art of maximizing the amount of work not done.” This can be understood as avoiding unnecessary complexity in the software development process. The word “lightness” used in the agility evaluation criteria proposed by Taromirad and Ramsin (2008a) refers to the same characteristic. When looking at the criteria, two parts – leanness and simplicity and lightness – may seem to overlap. However, their descriptions refer to different characteristics despite the wording similarity. While leanness covers the preference for economical, simple and quality-assured means for production and shorter timespans (Taromirad and Ramsin 2008a), lightness and simplicity refer to the complexity of the development process. This description can be adapted to measure the agility of evaluation methods for ITS applications. In other words, simplicity and lightness should be understood as the opposite of complexity of the evaluation method.

The agility evaluation criteria developed for software development methods by Taromirad and Ramsin (2008a) identify “technical quality” as part of the
assessment criteria. Technical quality is also characterised in the criteria with the question of how technical quality is monitored and controlled during the development process. The principles of agile software identify technical excellence and good design as important enablers for agility: “Continuous attention to technical excellence and good design enhances agility” (Beck et al. 2001). In practice, this principle can be understood as selecting the right and technologically competent developers for the task and paying attention to good design practices (Bless 2011).

In software development, the main outcome of the development process is a software product. For software, it is also possible to measure the technical quality of the product both during the development process and after the development work has ended. In the case of evaluation of ITS systems, the main outcome is the evaluation results with appropriate documentation, but the evaluation process may involve use of technology such as software tools. The concept of technical excellence, as defined for software development, is therefore not sufficient as such for measuring agility in the context of evaluation of ITS systems. Instead, it will be necessary to replace it in the agility evaluation criteria with a different but corresponding element based on the principle expressed in the Agile Manifesto.

The corresponding element for evaluation methods for ITS is “research and technical quality”. Several common characteristics are highly valued both in scientific research and evaluation studies of ITS applications. While not all evaluation studies of ITS aim to produce scientific knowledge, both are expected to provide results with reliability and validity and follow ethical research practices. While research and technical quality is an important characteristic of an evaluation method, it is in fact a requirement for both agile and non-agile evaluation methods. It should therefore not be included in the agility evaluation criteria.

The last part of the agility evaluation criteria proposed by Taromirad and Ramsin (2008) is active user collaboration. This part of the criteria follows from the values expressed in the Agile Manifesto (e.g. “Customer collaboration over contract negotiation” and “Individuals and interactions over processes and tools”). In case of agile software development, active user collaboration provides input to software development during the development process. During the agile development process, collaboration with users may provide information on changes outside the development process as well as new objectives and requirements (Figure 6).

In the context of evaluation of ITS systems, the feedback provided by the method during the evaluation process may be shared with the customer or other stakeholders when necessary. The ability of an evaluation method to provide feedback during the evaluation process has already been covered by another part of the criteria (responsiveness). In addition, flexibility measures the ability of the method to accommodate expected or unexpected changes during the evaluation process, whether they come from the customer or other user of the evaluation results or not. In summary, active user involvement, as a characteristic of an agile evaluation method, is to a substantial degree covered
Identification of agile characteristics and determining the agility of an evaluation method

by flexibility and responsiveness. It was considered preferable to keep the different parts of the criteria orthogonal – in other words, independent and non-overlapping. Active user involvement was therefore not included in the agility evaluation criteria adapted for evaluation methods for ITS systems (Table 11).

Criteria for assessment of the agility of evaluation methods for ITS applications were developed based on the agility evaluation criteria proposed by Taromirad and Ramsin (2008), the 4-DAT model described by Qumer and Henderson-Sellers (2006b), and the interpretations defined for individual parts of the agility assessment criteria in the context of evaluation of ITS applications. The framework proposed for the evaluation of agility of evaluation methods for ITS systems is presented in Table 12. When developing the criteria, meta-criteria developed by Taromirad and Ramsin (2008b) were taken into account to a relevant extent. These meta-criteria were developed for methods evaluating the agility of software development methods. The meta-criteria include precision, simplicity, quantitative metrics, usability and meaningfulness, consistency, minimisation of overlap between individual elements, generality and balance.

Table 12. Definitions of agile characteristics for evaluation methods for ITS applications.

<table>
<thead>
<tr>
<th>Agile feature</th>
<th>Definition (for evaluation methods for ITS applications)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Speed</td>
<td>The speed with which the evaluation produces results, in other words the time between the start of the evaluation process and availability of results</td>
</tr>
<tr>
<td>Sustainability</td>
<td>The extent to which the evaluation method is able to provide relevant evaluation results and remain fit-for-purpose in the future</td>
</tr>
<tr>
<td>Flexibility</td>
<td>The ability of the evaluation method to accommodate expected or unexpected changes occurring during the study</td>
</tr>
<tr>
<td>Learning</td>
<td>The ability of the evaluation method to &quot;learn&quot; from previous iteration rounds and improve itself when applied several times consecutively</td>
</tr>
<tr>
<td>Responsiveness</td>
<td>The ability of the evaluation method to produce feedback during the evaluation process</td>
</tr>
<tr>
<td>Leanness</td>
<td>The extent to which the evaluation method is suitable for implementation with economical, simple and quality-assured means</td>
</tr>
<tr>
<td>Lightness and simplicity</td>
<td>Absence of the complexity in the evaluation method</td>
</tr>
</tbody>
</table>

Several viewpoints had to be considered when developing indicators for measuring the level of agility of evaluation methods for ITS. First, the indicators need to provide valid results when used together to measure the agile characteristics of evaluation methods. In other words, they need to measure the feature or phenomena they are intended to measure. The criteria should be able to determine the level of agility of an individual evaluation method and preferably also to identify the agile characteristics present in the method. The criteria should also cover the values and principles presented in the Agile Manifesto and understood as agile by later authors who continued conceptualising agility. In addition, they need to cover all agile values and agile principles in a balanced way, except values and principles for which no relevant interpretation can be found in the context of evaluation of ITS applications.

To provide valid results, the criteria should measure the characteristics of the evaluation methods documented in the case studies, not factors that are more or less related to the interests and preferences of an individual researcher, his
or her organisation, their working practices or project management methods. The criteria should be defined in such a way that the impact of these confounding factors is minimised.

Second, the criteria had to be measurable. In other words, they had to be defined in a way that allows determining the value of each criterion for all or at least most case studies. A definition for measurability is provided elsewhere (Rossi 2007). The definition considers a characteristic measurable for a class of objects, if possible, as follows: “1. define the class of objects that manifest the characteristic; 2. identify the empirical properties that define the characteristic; 3. construct a reference measurement scale; 4. devise at least one measuring system based on that reference scale” (Rossi 2007).

Third, the need to obtain reliable results had to be considered. For example, it was necessary to define the criteria in such a way that they could be measured without ambiguity and with minimal subjective influence based on the information available in the case studies. Finally, it was considered preferable to define the criteria in a way that the different parts would not be overlapping or redundant.

The criteria proposed for measuring the agile characteristics of evaluation methods for ITS applications and related case studies are presented in Table 12. The related indicators for measuring each criterion are provided in Table 13. The elements included in the criteria are based on agility evaluation criteria defined for software development methods by Taromirad and Ramsin (2008a). The criteria were originally developed for software development methods, and they have been adapted to give them a meaningful interpretation in the context of evaluation of ITS applications.

The proposed criteria are measured with multinomial ordinal variables. This choice was made for multiple reasons. First, most characteristics of the methods are inherently qualitative. In addition, the case studies describing the methods have mostly been described as text. It would therefore be difficult or impossible to map qualitative differences to numbers on a continuous scale in a scientifically sound manner. Second, it was possible to tell which observable properties of the methods correspond to higher or lower levels of agility in terms of the different parts of the criteria, but it was not possible to say whether the different identifiable levels of each individual criterion are equally spaced on a continuous scale.

The first criterion is the speed with which the method provides evaluation results. In this case, speed is understood as the time between the start of the evaluation work and the delivery of the main results with the method. The assessment of speed with which the method will produce results was based on expert assessment and descriptions of the case studies. When determining the speed of an individual method, the reference point was the time that will be required for an experienced professional to apply the method and obtain the main results. When assessing speed, it was assumed that the described methods would be applied in the current situation in which the methods have already been documented and validated in the papers used as case studies.
Five categories were described for speed: very high, high, intermediate, low and very low. Methods which produce results within three months after starting the evaluation process were considered to have very high speed. Methods which provide results in 3–6 months were classified as high speed. Methods requiring 6–12 months to produce results were classified as intermediate speed. Methods requiring 12–24 months or more were considered to have low or very low speed. For evaluation studies of ITS, data collection on the performance and impacts of the system commonly takes several months. In some cases, a study on the impacts of an ITS system will require more than one or two years. For example, this may occur when the number of events to be observed over a period of one year is low. Longer observation periods may also be required when there is reason to expect that the short-term impacts of the system may be affected by “novelty effects” related to the novelty of the system instead of its intended functionality. Novelty effects can be expected, for example, when studying the impacts of variable message signs providing weather-related information and warnings (Rämä and Kulmala 2000).

In the context of agile software development, sustainability refers to the ability of developers and customers to maintain the pace of work indefinitely. For evaluations of ITS applications, sustainability refers to the ability of the method to provide relevant evaluation results and remain fit-for-purpose in the future – its future-proofness. The future-proofness of a method may be affected by the evolution of scientific knowledge, changes in technology, or other factors. Methods with a high risk of becoming obsolete in the near term are not future-proof. On the other hand, the most future-proof methods can be expected to be those with no significant risks of obsolescence, low sensitivity to changes in technology and evolution of scientific knowledge, and possibly with features that can be expected to reduce the risk of obsolescence (Table 13).

Flexibility refers to the ability of the evaluation method to accommodate expected and unexpected changes during the evaluation process. Evaluations of technical performance and impacts of ITS can be affected by a number of expected and unexpected events that may occur during the evaluation study and affect data collection or analysis of the results. First, collection of data on technical performance or impacts of an ITS system may be affected by numerous factors such as equipment failures, exceptional weather or traffic conditions, changes in physical infrastructure or transport services, challenges recruiting test persons for laboratory tests, difficulties getting responses to online surveys, or poor quality data received from an external stakeholder. Unexpected events are also possible during the analysis phase. The number of relevant events may be too small to yield statistically conclusive results, problems that occurred during data collection may be discovered only in the analysis phase, the sample of respondents or relevant events may be found to be biased, or the collected data may be in conflict with the assumptions included in the analysis methods. These include, for example, the assumption of normality involved in some parametric tests of statistical methods, and minimum sample size requirements.
Evaluation methods may involve possibilities to adapt the data collection or analysis once an expected or unexpected change, deviation from a plan or implicit assumption has occurred during the evaluation process. These may include, for example, limiting the scope of the study, extending or repeating the data collection, or use of expert opinion when all input data values required by a method cannot be measured or obtained from earlier studies. However, the options available are highly dependent on the type of expected or unexpected change and specific to the evaluation method. For this reason, the criterion for evaluating flexibility had to be specified in a rather generic way and with five ordinal values indicating very low, low, intermediate, high or very high levels of flexibility exhibited by the method. A method with very low flexibility can be expected to have no practical ways to adapt data collection or analysis during the evaluation process if anticipated or unanticipated changes occur. A method with low flexibility involves only limited possibilities to adapt data collection or analysis during the evaluation process. In case of a method with intermediate flexibility, there is at least one action with substantial impact available to the expert to ensure the method provides meaningful results regardless of anticipated or unanticipated changes. Methods with high flexibility allow the expert responsible for the evaluation to respond effectively to several types of anticipated events. Methods with very high flexibility allow the expert responsible for the evaluation to respond effectively to several types of anticipated events, and they have characteristics which provide resiliency in case of unexpected events during the evaluation process.

Agile software development methods typically include several iteration rounds during the development process instead of delivering only one version of the software. In the context of software development, the capability of the method to learn from previous iteration rounds during the software development process has been identified as a relevant part of agility. In the case of evaluation of ITS applications, the same method is usually not applied more than once to the same or substantially similar research problem within the same project or process of evaluating an ITS system. The aspect of learning from previous iteration rounds can therefore materialise as utilisation of experiences gathered during earlier projects of evaluations of ITS services in which the method has been applied to solve the same or substantially similar research problems.

The potential ability of an evaluation method to learn from previous studies was measured using four levels: (1) none, (2) limited, (3) substantial and (4) extensive. The first value is applicable to methods that have, for example, been so rigidly defined that no possibilities to learn from previous iterations exist without turning the method into something completely different from the original one. The second level (limited) corresponds to situations in which potential improvements due to learning from earlier iterations are limited to non-core aspects of the method, e.g. means to be used for implementation, or only a small improvement can be expected. In the case of level 2, the feedback from previous iteration rounds is unlikely to have a major impact on the results provided by the method. The third level (substantial) corresponds to situations in
which the feedback from previous studies using the method will likely have a substantial impact on the results the method will provide. For example, the parameter values or acceptance criteria used by the method may be built on experiences obtained in studies where the method had been applied earlier. The fourth level (extensive) is applicable to methods for which the feedback from previous iterations has a substantial impact on the results the method will provide, and the description of the method includes an agile development cycle or more than one iteration round during the same study.

The responsiveness of an agile software development method has been characterised as the ability of the method to provide feedback during the development process. Responsiveness is closely related to the third of the agile values, “customer collaboration over contract negotiation”, expressed in the Agile Manifesto. For evaluation methods for ITS applications, responsiveness corresponds to the ability of the evaluation method to produce feedback during the evaluation process. This feedback may consist of preliminary results of the study, information on the progress of the work, and successful completion of different phases of the evaluation study, e.g. successful completion of data collection. The level of feedback the method is capable of providing during the evaluation process was measured with a multinomial variable on an ordinal scale. Four possible levels were identified: (1) no feedback; (2) low: feedback with limited relevancy or information content; (3) substantial: relevant feedback with substantial information content, at least once during the evaluation process; (4) advanced: relevant feedback with substantial information content more than once during the evaluation process.

The combination of lightness and simplicity has been used in an earlier paper (Taromirad and Ramsin 2008a) as an indicator for measuring the agility of software development methods. In that study, lightness and simplicity were defined as lightness and simplicity of the software development process. In the context of evaluation of ITS applications, lightness and simplicity can be understood as the lightness and simplicity of the evaluation process. In practice, this can be understood as the absence of complexity – in other words, the absence of complexity in implementation of the method and analysis of data.

Evaluation methods for technical performance, impacts and interoperability may be complex or simple to implement. While some of the methods can be implemented with calculations carried out by hand or with commonly available spreadsheet software or other software packages, others require specialist software tools or even programming skills. The complexity of implementing the method is relevant for several reasons. First, it has an impact on the time required for preparations needed before the method can be used by an expert. Second, methods that are less complex to implement can be implemented by a wider range of experts and stakeholder groups than methods whose implementation is more complex.

Methods with low complexity of implementation were assigned to the highest category (high) in terms of lightness and simplicity. Methods with intermediate complexity of implementation were classified to the second category
Identification of agile characteristics and determining the agility of an evaluation method (intermediate) in terms of lightness and simplicity. All other methods were classified as having low level of lightness and simplicity.

Finally, an indicator is needed for leanness. The criteria for agility already include an indicator for speed with which the evaluation method will provide results. The criteria also include an indicator for lightness and simplicity which covers the complexity of implementation of the method. In earlier studies, simplicity and speed have also been considered as elements of leanness. However, they were not included now in the criterion for leanness, as they are already covered by other evaluation criteria for agility. Instead, the leanness of an evaluation method was defined as leanness in terms of the use of resources such as working time or financial resources corresponding to the cost of applying the method. In practice, this corresponds to the costs of applying the evaluation method.

The resources available for evaluating an ITS application are limited by at least three factors: (1) the cost of an evaluation study in comparison to the expected costs and benefits of the ITS project, (2) budgetary constraints of organisations implementing and promoting ITS and (3) the complexity of public procurement procedures. The first of the identified constraints has no obvious relationship with the agility of the evaluation method. However, evaluation methods with higher costs and lower levels of leanness are likely to be more difficult to employ in an agile manner, for example as part of an agile feedback loop.

First, the complexity of public procurement procedures and time required for procurement usually increase with the value of the contract. For example, legislation related to public procurement and its interpretation in an individual organisation may require a competitive tendering round to be opened and related documents prepared for all contracts exceeding a defined threshold value. Second, contracts with high value (and high costs) are more likely to require budgeting of the expenses before the start of the year when the costs will be incurred. This may be possible, for example, in the case of organisations with a fixed annual budget for research and development activities.

When the case studies were carried out, the internal guidelines of the Finnish Road Administration, based on the Finnish legislation on public procurement, required a competitive tendering round to be opened for all contracts exceeding €20,000 in value. Contracts valued below this value could be concluded with a simpler procurement procedure, which allowed starting the evaluation study sooner once the need for it had been identified. For this reason, evaluation methods requiring resources below €20,000 have been classified as methods with very high levels of leanness. Other levels for the criterion for leanness have been defined as multiples of this threshold value, taking into account the factors limiting the resources available for evaluating ITS applications. Methods that require resources corresponding to €20,001–40,000 for data collection and analysis were classified having a high level of leanness. Methods that require resources corresponding to €40,001–80,000; €80,001–160,000 or more than €160,000 were considered to have intermediate, low or very low levels of leanness.
Table 13. Measurable indicators for agile features of evaluation methods for ITS systems and related case studies.

<table>
<thead>
<tr>
<th>Agile feature</th>
<th>Measurable indicator(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Speed</strong></td>
<td>Time between the start of the evaluation process and delivery of evaluation results (calendar days or months):</td>
</tr>
<tr>
<td></td>
<td>(1) Very low, &gt;24 months</td>
</tr>
<tr>
<td></td>
<td>(2) Low, 12–24 months</td>
</tr>
<tr>
<td></td>
<td>(3) Intermediate, 6–12 months</td>
</tr>
<tr>
<td></td>
<td>(4) High, 3–6 months</td>
</tr>
<tr>
<td></td>
<td>(5) Very high &lt; 3 months</td>
</tr>
<tr>
<td><strong>Sustainability</strong></td>
<td>Future-proofness of the method, considering the risk of obsolescence:</td>
</tr>
<tr>
<td></td>
<td>(1) Very low: methods with a high risk of being obsolete in the near term, e.g. due to technological changes or expected evolution of scientific knowledge with no factors mitigating the risk of obsolescence</td>
</tr>
<tr>
<td></td>
<td>(2) Low: methods with a substantial risk of being obsolete in the near term, e.g. due to technological changes or expected evolution of scientific knowledge, all significant risks leading to obsolescence of the method during the next 5 years are not covered by possible mitigating factors</td>
</tr>
<tr>
<td></td>
<td>(3) Intermediate: methods with a risk of being obsolete in the near future, risks leading to obsolescence mostly covered by mitigating factors</td>
</tr>
<tr>
<td></td>
<td>(4) High: methods with limited sensitivity to evolution of scientific knowledge and changes in technology and no significant risk of being obsolete in the near future</td>
</tr>
<tr>
<td></td>
<td>(5) Very high: methods with a) limited sensitivity to evolution of scientific knowledge and changes in technology in the near future, b) no significant risk of being obsolete and c) with features that can be expected to mitigate the risks of obsolescence</td>
</tr>
<tr>
<td><strong>Flexibility</strong></td>
<td>Foreseen events (e.g. results different from anticipated, lower than anticipated quality of data, customer priorities changed during the project or external unanticipated event affecting the study) may affect the ability of the method to produce meaningful results. The level of flexibility exhibited by the method has five categories:</td>
</tr>
<tr>
<td></td>
<td>(1) Very low: no practical ways to adapt data collection or analysis during the evaluation process</td>
</tr>
<tr>
<td></td>
<td>(2) Low: limited possibilities to adapt data collection or analysis during the evaluation process</td>
</tr>
<tr>
<td></td>
<td>(3) Intermediate: at least one action with substantial impact available to the expert to ensure the method provides meaningful results</td>
</tr>
<tr>
<td></td>
<td>(4) High: ability to respond effectively to several types of anticipated events</td>
</tr>
<tr>
<td></td>
<td>(5) Very high: ability to respond to several types of anticipated events; in addition, the method has characteristics that provide resiliency in case of unexpected events during the study</td>
</tr>
<tr>
<td></td>
<td>Characteristics of the evaluation method that increase or reduce the ability of the method to respond to anticipated and unanticipated events:</td>
</tr>
<tr>
<td></td>
<td>a) Number of characteristics (N)</td>
</tr>
<tr>
<td></td>
<td>b) Impacts of identified characteristics on flexibility (no impact</td>
</tr>
<tr>
<td><strong>Learning</strong></td>
<td>Ability of the method to learn from earlier evaluations based on similar methods and to improve itself, indicated by the case study or description of an evaluation method:</td>
</tr>
<tr>
<td></td>
<td>(1) None or very limited: no significant potential for learning from previous studies utilising the method</td>
</tr>
<tr>
<td></td>
<td>(2) Limited: improvements limited to non-core aspects of the method, e.g. means to be used for implementation, or only a small improvement expected, no major impact on the results provided by the method</td>
</tr>
<tr>
<td></td>
<td>(3) Substantial: improvements having a substantial impact on the results, e.g. updates to acceptance criteria, parameter values related to the method</td>
</tr>
<tr>
<td></td>
<td>(4) Extensive: improvements having a substantial impact on the results, e.g. updates to acceptance criteria or parameter values and the method involves more than one iteration round during the same evaluation study</td>
</tr>
<tr>
<td><strong>Responsiveness</strong></td>
<td>Level of feedback the method is capable of providing during the evaluation process:</td>
</tr>
<tr>
<td></td>
<td>(1) No feedback</td>
</tr>
<tr>
<td></td>
<td>(2) Low: feedback with limited relevancy or information content</td>
</tr>
<tr>
<td></td>
<td>(3) Substantial: relevant feedback with substantial information content, at least once during the evaluation process</td>
</tr>
<tr>
<td></td>
<td>(4) Advanced: relevant feedback with substantial information content more than once during the evaluation process</td>
</tr>
</tbody>
</table>
2.3 Determining the agility of an evaluation method

Knowing the agile characteristics exhibited by an evaluation method does not tell as such whether the method should be called or classified as an agile one or not. First, a way is needed to combine information provided by the criteria for agility and the analysed case studies. Second, a criterion is needed for classifying an evaluation method as an agile one or other type of method.

Studies on agile software development methods do not provide an unambiguous answer to the question of how to classify a software development method as agile or non-agile. When criteria described by Taromirad and Ramsin (2008a) were applied to a software development method (XP, extreme programming), the authors commented: “Regarding agility, XP seems to be at an acceptable level, as most of the results are high.” In other words, a software development method may be classified as an agile one if it reaches high values with most of the evaluation criteria for agility of software development methods. However, the experience with use of the criteria proposed by Taromirad and Ramsin (2008a) is limited. It is therefore not possible to use the results presented in that paper as the only criterion for determining the agility of a software development method or agility of an evaluation method for ITS applications.

In the context of agile supply chains studied in logistics, agility has been defined as “business-wide capability that embraces organizational structures, information systems, logistics processes, and, in particular, mindsets” (Christopher 2000). In addition, flexibility is defined as the main characteristic of an agile organisation.

The human mind commonly distinguishes words and objects in relation to their binary opposites. While this is an intuitive way to create definitions, it is not sufficient to define an object or entity such as an agile evaluation method. Instead, a positive definition is necessary. It will be necessary to define an agile evaluation method by describing its necessary characteristics – in other words by creating an intensional definition. Intensional definitions are widely used in technical standards. Instructions for creating intensional definitions for standardisation purposes have been provided by the IEC (Mari and Goodwin 2018) and summarised in a related article (Mari 2015). First, all parts of an intensional definition must be necessary to define the concept the definition aims to
represent. Second, the definition should cover all necessary characteristics of the entity represented by the definition.

The following definition is proposed for agile evaluation methods for ITS applications:

Agile evaluation methods for ITS applications are evaluation methods which
1) have an ability to accommodate changes during the evaluation process (flexibility)
2) are lean in terms of consumption of resources (leanness)
3) have high or very high speed which allows them to be invoked in an agile manner (speed)
4) may exhibit also other agile characteristics such as preferring light and simple analysis methods.

Based on the definition above, evaluation methods for ITS applications that exhibit high levels of flexibility and high levels of leanness and provide results with high speed, allowing the method to be invoked in an agile manner, should be classified as agile evaluation methods. Methods with high or very high levels of flexibility fulfil the first part of the definition. Methods with high or very high levels of leanness in terms of resource consumption meet the requirement expressed in the second part. Methods with high or very high speed fulfil the third. Classification of an evaluation method as an agile one requires that the method exhibits all three characteristics included in the definition.
3. Overview of current evaluation methods for ITS applications

3.1 Frameworks for ITS deployment

ITS strategies have been defined in several countries and at European and local level. They typically provide a vision of a transport system and society utilising ITS technologies, and define a limited set of priority actions, priority themes or priority systems. Examples of ITS strategies include the proposal for the national ITS strategy for Finland published in 2009 (Ministry of Transport and Communications 2009), the European ITS Action Plan (European Commission 2008) and the US DOT ITS Strategic Plan for 2015–2019 (Barbaresso et al. 2014).

The ITS strategies of individual countries use slightly different terminology and have different structures. However, the terminology for definition of an ITS strategy was described in a paper published in 2004 (Bekiaris et al. 2004) (Figure 4). An ITS strategy includes a group of individual actions which contribute to one or more steps. An implementation strategy for one or more ITS applications consists of these steps. The implementation strategy or strategies contribute to the achievement of future scenarios defined in the ITS strategy. In addition to strategic level plans for ITS, roadmaps have been created to plan the development and deployment of individual ITS systems, technologies or subsets of ITS technologies.

Technology roadmaps may take many graphical forms, and they are usually expressed as time-based charts which integrate developments in technology, markets, business decisions, organisation and individual projects. Technology roadmaps may consist of multiple layers which express the different thematic areas, e.g. technology, product and market. A technology roadmap may also be expressed with bars within one or more layers. In addition, other forms of expression may be used such as tables, graphs, figures of various types, flowcharts or text. (Phaal, Farrukh and Probert 2004)

Examples of technology roadmaps for ITS include the ERTRAC (European Road Transport Research Advisory Council) roadmap for automated driving (European Road Transport Research Advisory Council 2015) and the Roadmap between automotive industry and infrastructure organisations on initial deployment of Cooperative ITS in Europe (Amsterdam Group 2013) developed by the Amsterdam Group, which is a collaborative platform of infrastructure organisations and vehicle manufacturers.
The planning of ITS deployment may be based on an ITS architecture. ITS architectures may be created globally, regionally, nationally or locally. The European ITS Framework Architecture, FRAME (http://www.frame-online.net), focuses mainly on the functional level. It describes the user needs and stakeholder aspirations for ITS as well as the functional blocks and connections between functions needed for creating combinations of functionalities answering user needs (Bossom et al. 2011, Ebner et al. 2011, Jesty and Bossom 2011). The United States has its own national ITS architecture, the Architecture Reference for Cooperative and Intelligent Transportation (ARC-IT) (http://www.arc-it.net), which includes enterprise, functional, physical and communications views that can be used for describing the implementation of different service packages (US Department of Transportation 2017).

Defining an ITS strategy or an ITS roadmap includes also the selection of ITS technologies or ITS applications which will supported with implementation strategies or roadmaps. In practice, the selection of ITS systems for deployment may start from the list of all possible systems within the scope of the study, such as advanced driver assistance systems (Bekiaris et al. 2004). The systems to be deployed can then be selected, for example, with multicriteria analysis or based on inputs from an expert round table (Bekiaris et al. 2004). This multicriteria analysis of expert assessment will require information on the ITS systems under analysis such as information on impacts, benefits and costs, socio-economic profitability, technical risks and user acceptance.

### 3.2 High-level requirements for assessment methods for ITS

#### 3.2.1 Requirements related to scientific research

Expectations regarding ITS evaluation studies are somewhat similar to those for scientific research, which include e.g. the validity and reliability of research results, repeatability of studies, accuracy and conciseness of language, as well as accuracy and technical soundness of the results required in the engineering sciences. There is also a principle that methods for collecting scientific evidence should be public (Piccinini 2003).

However, there are also substantial differences. While scientific research is expected to provide new research results, evaluation studies of ITS systems do not always aim to achieve this. Instead, they may focus on serving the information needs of industry, regulators, infrastructure managers or other stakeholders.

#### 3.2.2 Requirements from users of study results on ITS

Results on the impacts, costs, socio-economic profitability, interoperability, risks and other aspects of ITS projects are used by a variety of public and private sector decision makers. These include, but are not limited to, public authorities responsible for decisions related to transport and mobility, infrastructure managers and regulators of the transport system. Evaluation results may be used, for example, to make decisions on implementation or disman-
ting of ITS services, development of regulation for the transport system and ITS services or creating incentives for relevant stakeholders such as travellers and system suppliers. Several requirements for evaluation results originate from these user groups and their activities.

The time available for carrying out an evaluation study may be limited, and the ability of the method to provide results in a limited timeframe is therefore a preferable characteristic. For example, budgeting decisions occur every year in many organisations. The deployment of an ITS system may also have connections to development of other than ITS-related regulation. ITS systems may also be planned and procured as part of a construction project. If the evaluation results arrive at too late a stage of the decision-making process, their impact and value may be reduced. Availability of the results only late in the process has been one of the challenges, for example, when cost-benefit analysis has been applied to transport related projects (Mouter 2017a).

Second, the resources available for an evaluation study of an ITS deployment may be limited, and the evaluation methods should be efficient in terms of use of resources. Especially for ex-ante evaluation studies, the ideal characteristics of an evaluation study include efficiency in terms of use of resources and effectiveness, “[meeting] the objectives for a reasonable cost without wider unacceptable side-effects” (Tarry et al. 2002). Among other things, the resources allocated to an evaluation study need to be proportional to the expected benefits of the system, the overall context, the novelty of the ITS system and its technology (Tarry et al. 2002).

Evaluation methods for ITS should be transparent. First, verifying the correctness of the assumptions included in an evaluation method or the results provided by the method will not be possible without a sufficient level of transparency. The same applies to the falsifiability of the results. A critical discussion of the evaluation method and the results would be very difficult if the method were a black box. In the worst case, difficulties with verification or falsification of the evaluation results may contribute to mistrust of the results. Although research results on evaluation of ITS in this regard are limited, experience with the use of cost-benefit analysis for transport projects indicates that its impartiality is not always trusted by political decision-makers (“Politicians don’t trust CBA’s impartiality”), nor do they necessarily agree on the normative choices included in the analysis (“Politicians disagree with normative choices made in CBA”), and its results may be used opportunistically in political debates (“opportunistic use”) (Mouter 2017a). In addition to the method, the need for transparency applies also to the evaluation results (Tarry et al. 2002).

In case of cost-benefit analysis, several recommendations have been made to address the requirement of transparency and perceived problems with the method. The solutions suggested by interviewees in a study carried out by Mouter (2017b) included, for example, ”make the calculations of the CBA verifiable”, “describe the project’s effects on disaggregated levels”, “position CBA as a modest instrument”, “give MPs enough time to verify whether CBAs are
carried out in an impartial way”, and "make it explicitly clear which elements of the political trade-off are (not) covered by a CBA."

The evaluation results should be easy to understand (Tarry et al. 2002). A decision-maker may be less likely to trust results and evaluation methods that are not easily comprehensible. While many civil servants making decisions on ITS have a degree in engineering, this does not necessarily imply expertise with ITS or techniques for their evaluation.

The evaluation results should be “easily comparable with other results” (Tarry et al. 2002). Comparability allows a synthesis to be made of the impacts of an ITS application e.g. in different contexts and with different technical implementations. Comparability can be facilitated by documenting the means used for data collection and analysis and possible external influences on the results.

Evaluation methods for ITS should preferably be adaptable to different operating environments and ITS applications. There may be a need to study the technical functioning or impacts of the same ITS system in different operating environments. An evaluation method with a wider scope is also more valuable than one that is suitable for one ITS application only.

Finally, evaluation methods for ITS should provide accurate results, provide means for comparing ITS to other investments in the transport system, be "objective without any positive or negative bias", "include rigorous sensitivity testing" of results, and not create a false sense of accuracy. (Newman-Askins, Ferreira and Bunker 2003)

3.3 Methods – technical evaluation

The expected impacts of an ITS system cannot be realised without a system being technically functional. The requirements for technical functioning are usually defined during the development phase of the system. They may be related to different characteristics of the system such as reliability, performance, availability, sensitivity to disturbances in the environment, interoperability, fault tolerance (behaviour in the case of erroneous input from the user or other systems), etc. Once the requirements for the system have been defined, it will be necessary to verify that they are met.

The development of an ITS system from basic research to a functional product may proceed in different ways. It may, for example, follow a systems engineering model such as the V model (Cechini, Ice and Binkley 2009) or the waterfall model, or be organised according to the principles of agile systems engineering (Stelzmann 2011). In the case of a project following the V model, validation and verification of the systems starts with unit testing and testing of individual subsystems, and continues with system verification and system validation (Cechini, Ice and Binkley 2009). Based on the V model, an integrated evaluation framework has been developed for ADAS systems with preventive safety features (Scholliers et al. 2008). The framework covers both component level validation and system level verification of the system in regard to technical functioning and human factors.
Verification of individual components or subsystems may be performed against their specified requirements (Cechini, Ice and Binkley 2009) or other technical specification (Scholliers et al. 2008). It may also be part of the development process (Cechini, Ice and Binkley 2009). In the PReVAL framework, technical evaluation consists of verification, checking that all subsystems operate according to related technical specifications, and validation, verifying that the system works as specified and meets its objectives (Scholliers et al. 2008). Different evaluation methods can be used for verification and validation of subsystems and complete systems.

Laboratory tests can be standardised or otherwise documented accurately in terms of the methods applied and the expected results. Their strengths include also repeatability and comparability of the test results and possibility to create different test scenarios. Their weaknesses include their limited scope and possibility for manipulation. For example, laboratory tests do not measure the impacts of failure modes and factors affecting reliability that exist outside the controlled laboratory environment. It is also possible that the performance of the product has been optimised for one test scenario at the expense of others, or even to manipulate the test results with a “defeat device” (Trope and Ressler 2016).

Field operational tests (Barnard et al. 2017) can be used to validate systems in a real operating environment. In regard to technical functioning, this means verifying that the system functions as expected also in realistic operating conditions. Field operational tests may be needed especially where relevant factors affecting the system cannot all be reliably reproduced in a laboratory environment, or when constructing a testing system resembling a realistic operating environment would be uneconomical. Challenges with field operational tests include sensitivity to confounding factors and external disturbances, transferability of results, and obtaining a sufficient number of relevant events while keeping the size of the collected data set manageable.

When standards or specifications have been established for a system or a system component, it will be possible to verify conformance to specifications with conformance testing. Conformance tests may be performed by system suppliers as part of the development process, by third parties for purposes of verifying the conformance of an implementation, or by other stakeholders. Conformance tests may be performed in a real operating environment or in a test laboratory. ITU and ISO have published specifications for conformance testing in the domain of communications engineering. Such specifications are provided in ITU-T recommendations X.290-X.296 and ISO/IEC standard series ISO/IEC 9646. For some ITS systems, conformance testing specifications addressing an individual system have been developed (e.g. EN16454 for eCall and ETSI TS 102 868-102 871 for cooperative ITS). Conformance tests can be used for verifying the conformance of an implementation to a specification, and they can be expected to improve interoperability between implementations of a specification. However, conformance tests do not necessarily measure the suitability of the system for its intended use. For example, intermittent faults or faults occurring only in specific situations or with specific parameter
values are not necessarily detected in a conformance assessment, if they were not anticipated when designing and parametrising the tests to be carried out. The scope of conformance assessment is also tied to the specification to which the system has been claimed to conform. Conformance assessment is not intended to detect the weaknesses of a system that do not manifest themselves as non-conformance to an existing specification.

3.4 Methods – impact assessment

Methods to be used for studying the impacts of ITS systems are more or less specific to the ITS system and types of impacts being studied. Methods needed to study the impacts on safety, environment and emissions, traffic efficiency and mobility are different but share certain common elements.

The impacts of an ITS system on traffic safety can be studied by observing the impact of the system on user behaviour. For a limited number of system users, monitoring of the impacts of the system on driver behaviour can be carried out in a driving simulator (Vollrath, Schleicher and Gelau 2011), in a controlled experiment on a test track (Ruscio, Cicero and Biassoni 2015), in a real traffic environment (Birrell, Fowkes and Jennings 2014) or in a field operational test (Saint Pierre, Tattegrain and Val 2014; Barnard et al. 2017). It may also be possible to monitor the characteristics of the traffic flow and monitor the impact of the system on safety-related parameters such as speed of traffic flow or headways (gaps between vehicles) (Rämä 2001). Observations of user behaviour and traffic conflicts between road users can also be carried out on a limited number of sites (Várhelyi 2018) to study the safety impacts of an ITS system.

Impact assessments may also be carried out based on accident statistics. First, it may be possible to compare the accident rates of vehicles equipped with the system to those of other vehicles with similar characteristics (Highway Loss Data Institute 2012), or to compare accident rates on road links or other physical infrastructure equipped with the system to those on the corresponding infrastructure without it. Second, it may be possible to compare accident frequencies that are sensitive or not to the safety measure being evaluated between vehicles equipped with the system or not (the induced exposure method) (Fildes et al. 2015). The analysis of accident statistics for impact assessment may be hampered by confounding factors such as changes in road user behaviour over time (Ess and Antov 2017) differences in traffic exposure between groups, or low numbers of relevant accidents included in the data set available for analysis.

Traffic simulation models and related surrogate safety indicators (Gettman and Head 2003) have also been used for evaluating the impacts of various ITS systems. Safety impacts of an in-vehicle ITS system can also be studied by monitoring conflicts occurring between the driver in an equipped vehicle and other road users.

Impacts of an ITS system on vehicle emissions can be studied, for example, by monitoring fuel consumption in a field test (Innamaa and Penttinen 2004).
Also traffic simulation models have been used to estimate the impact of ITS applications on CO\textsubscript{2} emissions (Qian and Chung 2011; Elyasi-Pour 2015). The AMITRAN project developed a framework which combines a transport demand model, traffic simulation model and emission model to provide a way to estimate impacts on CO\textsubscript{2} emissions (Mahmod et al. 2014).

Impacts on traffic efficiency are commonly evaluated by measuring travel time and its predictability (Kulmala et al. 2002). Depending on the ITS system and the context of the evaluation study, different technologies such as floating mobile data, automatic licence plate recognition and monitoring of Bluetooth signals can be used for measuring travel times. For measuring the predictability of travel time, variance of travel time (Stockton et al. 2003) and cumulative distribution of travel time can be used as indicators.

In addition to safety, traffic efficiency and emissions, ITS systems may have impacts on other characteristics of the transport system and its users such as user comfort, service level of the transport system, costs of the transport system, efficiency of logistics, vehicle fleet and its costs, and accessibility and mobility. Indicators for these impacts and related evaluation methods are provided e.g. in the EasyWay Viking guidelines for evaluation for ITS projects (Kulmala et al. 2002) and US evaluation guidelines for ITS benefits (Stockton et al. 2003).

3.5 Methods – socio-economic assessment

Cost-benefit analysis has long been used to evaluate the socio-economic profitability of investments in ITS (Schulz and Geis 2014). It is also an established method for evaluating the socio-economic profitability of transport projects. Cost-benefit analysis also allows investments in ITS to be compared to other investments in the transport system.

Cost-benefit analysis requires methods for monetary valuation of various types of impacts such as safety, environmental and traffic efficiency impacts. Most European countries have defined unit cost values for impacts on traffic safety such as fatalities and injuries in traffic accidents (Kasnatscheew et al. 2016, for Finnish unit cost values, see: Finnish Transport Agency 2015; Tervo nen and Metsäranta 2015).

Cost-benefit analysis also has certain limitations. First, it cannot consider the impacts for which there are no commonly accepted valuation techniques and related unit cost values. Second, it may conceal normative choices and should therefore not be used as such for determining a project as acceptable or not or for prioritising projects. For example, a project with traffic efficiency benefits but a detrimental impact on safety may have a positive benefit-cost ratio but would not necessarily be acceptable. Cost-benefit analysis must therefore be combined with other analysis methods such as multicriteria analysis in the socio-economic evaluation of ITS systems.

The use of multicriteria analysis for ITS projects has been discussed by De Brucker, Verbeke and Macharis (2004). Qualitative assessment resembling
Overview of current evaluation methods for ITS applications

3.6 Methods – interoperability

Interoperability testing is usually carried out for implementations of the same specification, or for systems with different specifications but for which interoperability is claimed. The process of interoperability testing has been defined in supplements to ITU-T X.290 series recommendations (ITU-T 2008a, ITU-T 2008b). In addition, ETSI has published the ETSI white paper on interoperability (van der Veer and Wiles 2008) and specifications on interoperability testing.

Typically, interoperability testing involves completing an interoperable features statement which indicates the features of the implementations or systems that are expected to be interoperable or non-interoperable, and documenting the testing architecture and the interoperability test cases and finally the test results.

3.7 Implementation issues

The deployment of an ITS system may be constrained by legal or institutional issues, an insufficient or uncertain level of information security, or lack of user and societal acceptance. These may also be related to corresponding risks of deployment. Successful deployment of an ITS system requires these possible constraints to be addressed and deployment-related risks to be managed. Planning the deployment process and management of risks may therefore require evaluation of legal and institutional issues, information security and user privacy, and user and societal acceptance.

Legal issues related to deployment of ITS have been identified, for example, in topics related to privacy and liability. In the case of liability, legal issues related to ITS have been studied by an approach with three individual parts: 1) identifying and describing the current status of relevant legislation; 2) identifying the applicable "measures, rules and procedures" and 3) identifying the systems most affected by liability issues or systems for which actions are needed to deal with the identified liability issues (Yen et al. 2012).

User acceptance of ITS is commonly studied with questionnaires aimed at users and non-users of the system (e.g. Poli et al. 2012) as well as the general public. In a review by Malin (2015) of theoretical models for measuring user acceptance, most of the models highlighted the importance of the user’s intention to start using the system. A number of factors affecting user intention and actual use of the system were identified. These include, for example, perceived usefulness and ease of use, social influence, conditions facilitating usage, as well as effort expectancy and performance expectancy. Willingness to buy the system was considered an important part of user acceptance although not decisive for actual usage. In the case of intelligent speed adaptation (ISA), “the effectiveness of ISA (1), equity (2), effectiveness of ITS (3) and personal and
social aims (4) were the four variables that had the largest total effect on the acceptability of ISA” (Vlassenroot et al. 2011). User acceptance can be studied with a questionnaire aimed at potential users of the system such as travellers or active car users.

In general, information security can be conceptualised as a combination of confidentiality, integrity and availability. Also other properties such as anonymity may be relevant depending on the system and its characteristics (Anderson 2008). In a US evaluation study on information security of ITS systems, the analysis approach included security assessments of ITS subsystems, ITS communication infrastructure and ITS data flows (Biesecker et al. 1997).

In the context of information security, evaluation has been defined as “the process of assembling evidence that a system meets, or fails to meet, a prescribed assurance target.” A variety of evaluation methods such as reviews of system specifications or source code and security tests (with or without knowledge of system specifications or source code) may be employed. The testing process may look at weaknesses in the architecture of the system, weaknesses in implementation and cover a checklist of less common flaws such as the ones described in relevant literature. (Anderson 2008)
4. Case studies

This chapter presents several case studies analysed from an agility point of view. For each case study, descriptions are provided of the ITS evaluation problem and the evaluation method. These descriptions are based on Publications I-VII and provided as subchapters for each case study. The agile characteristics of evaluation methods used in the case studies are identified in the chapter. The agile characteristics are identified based on the measurable indicators for different agile characteristics provided in Chapter 2. Evaluation methods are also classified in terms of agility based on the assigned indicator values and the definition of an agile evaluation method provided in Chapter 2.

4.1 Model for the safety impacts of road weather information services available to road users and related socio-economic benefits

4.1.1 ITS evaluation problem

Weather and the status of the road surface affect the safety of road users. In many countries, weather poses a significant challenge to road traffic and maintenance. Snow and icy road surfaces reduce the friction of vehicle tyres. Snow, heavy rain or fog affects visibility. Weather and road conditions are highly time-variant phenomena: temperature, rainfall and the state of the road surface may change from one hour to the next. A review of the impacts of winter weather on road safety has been provided by Strong, Ye and Shi (2010). The high accident risks caused by adverse weather conditions can be reduced by providing information, warnings and support to road users. When drivers are informed of weather and road conditions, they can adapt their behaviour accordingly by reducing their speed and increasing their alertness. Currently, warnings and information on related weather and road conditions are disseminated via mass media such as TV, radio and web sites. The operation and implementation of services involve substantial investments and operation costs. The costs are related, for example, to data collection, data processing, operation of the services and dissemination of the information. The benefits of road weather information should be known at least approximately before major decisions are made on the implementation and operation of the services. The decision to implement the service as well as decisions on the quality targets for the service are typically made on a national or regional level.
4.1.2 Evaluation method

A mathematical model for accident reduction related to road weather information and warning services and related socio-economic benefits was built on the basis of estimates made by the authors and available statistics on traffic safety and weather. The model was built with estimates and data related to conditions in Finland, then generalised to other countries. The theoretical framework behind the model is the principle that the number of accidents is dependent on the exposure to risk and probability of accident (Elvik and Vaa 2009). Costs incurred by accidents are dependent on the numbers of different types of accidents during the period under study and the unit cost values of different types of accidents. Information on the effects of road weather information and warning services on the safety of road users was collected by means of a literature review and expert interviews. The results of the literature study are summarised in Publication I. Monetary valuation of safety benefits is based on unit values published by the Finnish Road Administration (2005).

When drivers are informed of weather and road conditions, they can adapt their behaviour accordingly by reducing their speed and increasing their alertness. When developing the model, traffic exposure in various road weather conditions was assumed not to be affected by road weather information services (Publication I). Weather-related information and warning services distributed via mass media to road users were assumed to reduce by 1–2% the number of all accidents involving personal injury or death on public roads in Finland. This estimate was based on the results of evaluations on Finnish VMS-based (variable message signs based) road weather warnings and speed control and on the relationship between speed and the number of accidents involving personal injury or death documented by Nilsson (2004). Similar results were obtained with a Delphi study conducted in 2007, which was answered by 135 Finnish and international experts (Aittoniemi 2007). When the accident reduction expressed in per cent in the current situation and unit costs for fatal and injury accidents are known, it is relatively simple to calculate the socio-economic benefit related to fatal and injury accidents in the current situation. However, more assumptions are needed to calculate the socio-economic benefit related to reduction in the number of accidents involving property damage only.

In addition to injury accidents and fatal accidents, there is an effect on the number of accidents involving only property damage. According to Nilsson (2004), the number of accidents involving personal injury or death increases with the average speed of traffic in relation to the second power of the ratio of average speeds after and before the speed change. However, Nilsson’s study presents no similar relationship for average speed and accidents involving only property damage.

After receiving information or a warning, a driver may lower his or her speed or increase his or her alertness (Luoma et al. 2000; Aittoniemi 2007) to road weather problems ahead. It is reasonable to assume that the effect of weather-related information and warning services on the number of accidents involving only property damage is related partly to speed and partly to the increased...
alertness of road users. Both these factors can be assumed to have equal weight. When the relation between average speeds of the traffic flow after and before the change in the number of fatal and injury accidents is known, it is possible to calculate the ratio of average speeds after and before the accident reduction and its square. When considering the impact of driving speed, the number of accidents involving property damage only were assumed to be directly related to average driving speed. When the square of the ratio of driving speeds before and after a speed change is known, it is possible to calculate its square root which corresponds to the ratio of driving speeds before and after the speed change. It was also assumed that increased alertness and the change in driving speed have impacts of equal weight on the number of accidents involving property damage only. It was then possible to calculate the ratio of property-damage-only accidents before and after the introduction of the service as the average of the two terms representing the impact of increased alertness and change in average driving speed. Next, it was possible to express the reduction in property-damage-only accidents in per cent as a function of the reduction of fatal and injury accidents expressed in per cent by substituting and rearranging the equations. It was then possible to calculate the reduction in the number of property-damage-only accidents as an absolute number when their number in the current situation was known. The socio-economic benefit due to the safety impact of the service was calculated by multiplying the numbers of fatal or injury and property-damage-only accidents by the corresponding unit cost values published for Finland by the Finnish Road Administration. A detailed description of the calculations included in the model is provided in Publication I.

When the model is extended to be applicable to countries other than Finland, the analysis becomes more complex. First, the share of road accidents occurring in adverse weather or road conditions is different in countries with a different climate, transport system and road user behaviour. Secondly, there are differences in the overall level of traffic safety between countries, and commonly agreed unit cost values for different types of road accidents are not available for all countries. These differences had to be considered when the model was generalised for use in other countries than Finland.

The model was extended to other countries and applied to Croatian data. First, differences in the sensitivity of road safety to weather were considered by looking at the share of injury and fatal accidents occurring in adverse road weather conditions in Croatia and Finland. The 1–2% impact on the number of injury and fatal accidents assumed for Finland was scaled with the relative shares of injury and fatal accidents occurring in adverse road weather conditions in Croatia and Finland. In Finland, 21.4% of deaths and 22.3% of injuries in road accidents occur when the road surface is snowy, slushy or icy (Statistics Finland 2006). The number of accidents involving only property damage during one year on public roads (9,951 accidents) was available for Finland in the statistics maintained by the Finnish Motor Insurers’ Centre (2008). In Croatia, 2.54% of fatal accidents and 3.63% of injury accidents occurred in adverse weather conditions. The input values for Croatia were obtained from
Leviäkangas et al. (2007) and Öörni (2007). In some cases, the shares of fatal and injury accidents occurring in adverse weather conditions cannot be obtained from the traffic safety statistics of the country in question. In these cases, one can use values obtained from expert interviews or from other countries with a similar transport system, climate and road user behaviour.

Second, the Finnish unit cost values for different types of accidents had to be scaled to make them applicable to Croatia. The Finnish unit cost values for fatal, injury and property-damage-only accidents were scaled by comparing the gross domestic products per capita adjusted with the purchasing power in Finland and Croatia. This way of calculating the unit cost values involves the assumption that the unit accident costs are proportional to GDP per capita. A large part of the costs associated with the accidents and their consequences are related to GDP as shown in the COST313 Final Report (European Commission 1994). A large amount of costs is also attributed to so-called ‘human costs’ reflecting the amount of pain and suffering in addition to the economic value of lost labour. As the aim is to estimate the costs approximately rather than strictly accurately, this assumption is probably acceptable. Naturally, researchers should use the unit accident costs of the country under analysis when available.

4.1.3 Experiences from using the method

The model can easily be used when the following statistics are available for the country to be analysed: the number of different types of accidents during one year on public roads, shares of different types of road accidents in adverse weather and road conditions, and the GDP and purchasing power parity per capita. The results obtained for Croatia with the model were found to be in line with the values used as inputs for the model. The model provided approximate but not exact values for the safety impacts of road weather information services. It did not provide confidence intervals for its outputs, but allowed calculation of the upper and lower boundary values for the most likely safety impacts and related socio-economic benefits. The model is also based on scaling of the safety impact most probably existing in Finnish conditions. When the model was applied to Croatia, this impact was scaled to take into account differences in risks related to driving in adverse road weather conditions and exposure to risk. The results provided by the model can be expected to be most accurate in countries in which the factors contributing to weather-related accidents are similar to the Nordic countries. In any case, a potential user of the model needs to study the assumptions included in the model and consider whether they are applicable to the country under analysis. Evidently the results provided by the model were approximate and give the correct magnitude of the safety impacts and related socio-economic benefits, but they will not replace a complete and detailed impact assessment that can be carried out for a service for which the content, implementation and user base are known. Instead, the strength of the model lies in its ability to provide quantitative estimates for safety impacts and related socio-economic benefits of road weather
information services with a minimal amount of collected information and the ease with which the model can be adapted to different conditions and services.

4.1.4 Agility assessment of the method

The agility of the method used in the case study was assessed based on the criteria for an agile evaluation method. The values of different parts of the criteria for agility are summarised in Table 14. In addition to indicator values, the table provides short explanations for assigned indicator values. According to the definition of an agile method presented in Chapter 2.3 and the criteria for agility presented in Chapter 2.2, the method can be classified as an agile method.

Table 14. Case study 1 – Indicators for agility of the method.

<table>
<thead>
<tr>
<th>Evaluation criterion</th>
<th>Value</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>Speed</td>
<td>5 (very high)</td>
<td>The method can be applied to a new country in less than 3 months, as the model requires only limited data and no specific software tools or programming for implementation.</td>
</tr>
<tr>
<td>Sustainability</td>
<td>5 (very high)</td>
<td>The method is based on the theories of traffic safety such as exposure, accident rate and accident consequences as determinates of safety and relations between average driving speed and accident occurrence. Theories related to traffic safety change over time although usually over longer periods. The sensitivity of the method to evolution of scientific knowledge is therefore limited in the short term. Sensitivity to technological changes is low, because the method is not sensitive to the dissemination channel of RTTI, and RTTI services are unlikely to become obsolete in the near term. In addition, the method includes characteristics such as high flexibility which can be expected to mitigate the effects of evolution in scientific knowledge.</td>
</tr>
<tr>
<td>Flexibility</td>
<td>4 (high)</td>
<td>If the parameter values for the country are not all available or are known to be inaccurate, values from comparable countries or values obtained from expert opinion may be used instead (e.g. share of injury accidents or fatal accidents occurring in adverse weather conditions). The effectiveness of road weather information services in terms of changing user behaviour may change in the future. Developments are possible, for example, in quality of information, dissemination channels and level of personalisation and context-awareness of the services. These changes can be at least to some extent be accommodated by changing the parameter values. National unit cost values for fatal and injury accidents may be used if they are available. The method allows responding in an effective and meaningful way to several anticipated events. The level of flexibility is therefore high.</td>
</tr>
<tr>
<td>Learning</td>
<td>2 (limited)</td>
<td>The method will take into account the impact estimates used for RTTI when the model has been applied previously. However, the method does not include any internal validation for the results or source data. The ability of the method to learn from earlier evaluation studies and improve itself based on an essentially similar method is therefore limited, as the improvements will likely be related to non-core aspects of the method.</td>
</tr>
<tr>
<td>Responsiveness</td>
<td>2 (low)</td>
<td>The only feedback the method can provide is whether the required input data has been provided or is missing. This feedback has only minimal information content. However, the short time required for applying the method reduces the importance of getting feedback during the estimation process.</td>
</tr>
<tr>
<td>Leanness</td>
<td>5 (very high)</td>
<td>The source data required by the model (data on traffic safety and gross domestic product) can be collected with limited effort. They may be available e.g. from transport safety authorities or the national statistical office for free or with very limited cost. The implementation of the calculations will not take more than one or two days. It is therefore highly probable that the model can be applied to a new country with resources corresponding to less than €20,000.</td>
</tr>
<tr>
<td>Lightness and simplicity</td>
<td>3 (high)</td>
<td>The calculations included in the model can be implemented with almost any spreadsheet software. No specialised software tools or programming skills are therefore required.</td>
</tr>
</tbody>
</table>
4.2 Reliability of an in-vehicle warning system for railway level crossings – a user-oriented analysis

4.2.1 ITS evaluation problem

The safety of level crossings is a major safety issue for railways and an existing traffic safety problem for road users. According to statistics published by the European Railway Agency (ERA), at least 123,000 level crossings exist in the European Union (EU), most of which (71%) are passive level crossings without any active warning or protection devices, such as lights, bells or gates (European Railway Agency 2012). Roughly 45% of level crossing accidents in the EU occur at passive level crossings, and 65% of road users involved in accidents are drivers or occupants of passenger cars or heavy vehicles (European Railway Agency 2012). In 2010, there were 359 level crossing accident fatalities in the EU (European Railway Agency 2012). Most of the direct causes are related to the behaviour of road users such as distraction (European Railway Agency 2012) or other human error (Tordai et al. 2008). According to the Finnish Accident Investigation Authority (2011), most of the accidents at passive level crossings in Finland are caused by the vehicle driver misjudging the situation but not by intentional risk-taking. The awareness and alertness of the driver and the safety of level crossings can potentially be improved by providing a warning to the driver that a train is approaching the level crossing. In-vehicle warning systems for railway level crossings have been under development in recent years at least in Finland (Öörni et al. 2011), the US (Aycin and Benekohal 2002) and Australia (La Trobe University 2012). Analysing the reliability of the system is an integral part of the development process. The warning system also needs to have a sufficiently high level of reliability before the positive safety impacts of the system can be realised. The aim of the analysis carried out in the case study was to analyse the reliability of an in-vehicle warning system for railway level crossings developed in Finland (Öörni et al. 2011), from the user point of view (Publication II). A simplified technical architecture of the system is illustrated in Figure 8.

![Figure 8. In-vehicle warning system for railway level crossings – simplified technical architecture, adapted from Öörni and Virtanen (2007).](image-url)
4.2.2 Evaluation method

The first task when applying the method is to create a system description. The description needs to cover the functionality of the implemented system, its functional requirements, and the requirements for system reliability. In this case, much of the description and functional requirements were available from a technical report describing the system being analysed (Öörni et al. 2011). The reliability requirements for the system in terms of the share of missed alarms of relevant events (false negatives) and share of false alarms of all alarms (false positives) were established based on a literature study described in Publication II.

Next, it was necessary to provide a definition for reliability and to establish reliability requirements for the system. Reliability is a concept that has different definitions in different contexts. Thus, a clear definition of reliability is presented here. An earlier study on the design and validation of advanced driver assistance systems (Gietelink 2007) defines reliability in a way originally mentioned by Storey (1996) as the “probability of a component, subsystem or complete system, functioning correctly over a given period of time under a given set of operating conditions”. The study by Gietelink (2007) also mentions indicators for reliability applicable to warning systems and focuses on system reliability from the human point of view: the performance of the system visible to the user in terms of true positives (correct activation of the system when needed), true negatives (correct suppression of the safety device), false positives (false alarms due to untimely or incorrect decision of the system) and false negatives (for example, late detections and missed alarms) (Table 15).

Table 15. Prediction matrix with number of samples, categorisations as true negatives, true positives, false negatives or false positives (Gietelink 2007), adapted originally from a paper by Lee and Peng.

<table>
<thead>
<tr>
<th>Prediction</th>
<th>Actual data</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Negative (safe)</td>
</tr>
<tr>
<td>Negative (safe)</td>
<td>$N_{TN}$</td>
</tr>
<tr>
<td>Positive (threat)</td>
<td>$N_{FP}$</td>
</tr>
</tbody>
</table>

These four variables can be used to calculate various reliability measures (Table 16). When reliability is defined as in the study by Gietelink (2007), one has to define the operating conditions in which the measurements are made, the length of the observation period, the way the probability of the system functioning correctly is measured, and the criteria for correct functioning of the system.
Table 16. Reliability measures calculated from a prediction matrix (Gietelink 2007).

<table>
<thead>
<tr>
<th>Rate</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>Real occurrence rate, $p$</td>
<td>$(N_{TN} + N_{TP}) / (N_{TN} + N_{TP} + N_{TN} + N_{TP})$</td>
</tr>
<tr>
<td>Accuracy, $p_{\text{accuracy}}$</td>
<td>$(N_{TN} + N_{TP}) / (N_{TN} + N_{TP} + N_{TN} + N_{TP})$</td>
</tr>
<tr>
<td>Precision, $p_{CP}$</td>
<td>$N_{TP} / (N_{TN} + N_{TP})$</td>
</tr>
<tr>
<td>True positive rate, $p_{TP}$</td>
<td>$N_{TP} / (N_{TN} + N_{TP})$</td>
</tr>
<tr>
<td>False negative rate, $p_{FN}$</td>
<td>$N_{TN} / (N_{TN} + N_{TP})$</td>
</tr>
<tr>
<td>True negative rate, $p_{TN}$</td>
<td>$N_{TN} / (N_{TN} + N_{TP})$</td>
</tr>
<tr>
<td>False positive rate, $p_{FP}$</td>
<td>$N_{FP} / (N_{TN} + N_{TP})$</td>
</tr>
<tr>
<td>Reliability, $p_{\text{REL}}$</td>
<td>$\sqrt{N_{TP} / (N_{TP} + N_{TP})(N_{TN} + N_{TP})}$</td>
</tr>
</tbody>
</table>

One of the possible methods for estimating the probability of any event is the relative frequency approach (Milton and Arnold 1995). Basically, this means repeating an experiment many times and calculating the probability of the event by dividing the number of times the event occurred by the number of times the experiment was run. The estimate gets more accurate the more times the experiment is run. In some cases, the data collection methods used and the characteristics of the system to be evaluated do not allow detection of ‘true negative’ as a separate event. This may occur in cases where the system under observation and analysis is normally in an idle state producing no output, and observation of the system is performed continuously rather than in an event-oriented manner. When the prediction matrix is written without $N_{TN}$, some reliability measures presented in Table 16 cannot be calculated when no information is available about the number of TN outcomes. Thus, real occurrence rate, accuracy, true negative rate and false positive rate were excluded from the group of reliability measures to be calculated.

Data collection was carried out at three points in the service chain: (1) where train position information is received by the level crossing server directly from either the train equipment or the mobile gateway, (2) at the interface providing level crossing status information to the in-vehicle device, and (3) between the in-vehicle system and the end-user. Data collection at points 1 and 2 was performed automatically using the data-logging features built into the level crossing server software. Data collection at point 3 was done by monitoring the level crossing and the in-vehicle system with video cameras and using a multiplexer to combine the two video streams into the same video data file (Figures 9 and 10). Since the focus was on studying the reliability from the user point of view, the reliability analysis carried out in this study was performed at point 3. Video monitoring was carried out at two level crossings (Lappohjan satama and Skogbyn seisake) to reduce the possibility that any unexpected random errors or differences in the physical environment would have a disproportionate impact on the results.
The definition of a successful alarm was created based on the high-level user requirements presented earlier, the guidelines available for existing level crossing warning systems, and a brief literature study (Publication II). The timing parameters of a successful alarm are documented in Publication II. The correct alarm belonging to category $N_{TP}$ and various types of unsuccessful alarms are illustrated in Figure 11.

The results of the video monitoring were viewed with Avidemux software, summarised and added to the prediction matrix. The selected reliability
measures (precision, true positive rate, false negative rate, and reliability) could then be calculated for the system.

Confidence intervals were then calculated for the true positive rate for both level crossings where the video monitoring was carried out. When estimating the confidence intervals, individual events (trains passing the level crossing) were treated as Bernoulli trials which may result in either success or failure, have an equal probability of success and are independent of each other. When the sample consists of a number of Bernoulli trials, the number of successes follows binomial distribution. Confidence intervals can be calculated for the true positive rate by approximating the binomial distribution with normal distribution if the sample size is large enough (Milton and Arnold 1995). An upper confidence interval was estimated also for the number of false alarms and corresponding false alarm rate. When estimating the confidence interval, false alarms were assumed to occur randomly and their number within a given period of time was then expected to follow the Poisson distribution. When the number of observed false alarms and the length of the time period are known, the upper limit for the number of false alarms expected to occur during the period can be estimated with help of chi-square distribution as presented in mathematics textbooks (Hald 1952). The statistical methods used are described in detail in Publication II. The statistical analysis was combined with analysis of the causes of unsuccessful alarms. The causes of unsuccessful alarms were studied by analysing the log files collected by the server calculating the statuses of level crossings and records of utilisation of rolling stock such as engines and rail buses obtained from the railway operator.

4.2.3 Experiences from using the method

The selected reliability measures were successfully calculated for both level crossings where video monitoring was carried out. In addition, confidence intervals were successfully estimated for true positive rate and false alarm rate. Even without sophisticated statistical methods, it was possible to conclude that the true positive rate achieved by the system did not meet the required level. Estimation of confidence intervals supported this conclusion. The quantitative methods used in the study pointed out that the challenges with the system analysed in the case study were related to missed alarms (false negatives) rather than false alarms (false positives). The results showed that the expected functionality has been realised, but the true positive rate of the system must be improved.

The use of probabilistic techniques such as those described in Tables 15 and 16 is problematic with events that occur relatively rarely. For example, some potentially hazardous situations may occur only in certain conditions, and for some failures of the system the mean time between failures (MTBF) is longer than or equal to the observation period. However, the results most probably reflected the impact of failure modes that have the greatest impact on the reliability of the system.

A large number of missed alarms occurred early in the test when the service was not operational because of a server failure. The failure occurred after the
The in-vehicle system had been verified to be in an operative state by checking that the in-vehicle system received responses to queries it sent to the level crossing server. Failures of the level crossing server are relatively rare events that do not usually occur within 11 days. For this reason, the measured data does not accurately reflect the impact of that failure mode on the reliability of the system. The time during which the server was not operating was identified on the basis of server log files and video recordings of the display of the in-vehicle client and was excluded from the observation period.

Video monitoring is a labour-intensive data analysis method. For this reason, only a relatively short observation period of 11 days was feasible. Thus, the results had to be interpreted with caution for all events for which the observation period was shorter than or of the same length as the mean time between a particular type of failure (MTBF) or the mean time to repair it (MTTR). In other words, the method could not be used to measure accurately the impact of failure modes that occur only rarely.

The reliability targets for the system and the parameters for a successful alarm were set on the basis of the literature study. While they could be expected to serve as a good starting point for analysing the reliability of the system, they reflected the knowledge available at the time the study was carried out, but they could not be considered as final. Analysis of the causes of unsuccessful alarms allowed many of their causes to be identified. Based on the types of unsuccessful alarms, recommendations were drafted for further research and development of the system.

The method requires understanding of the operation of the system being studied and its main failure modes. Otherwise, it will be difficult to conclude which failure modes are reflected in the results and which have likely not been captured by monitoring. Information on the operation of the system is also required to conclude whether the assumptions made when calculating the confidence intervals are valid in a particular case.

4.2.4 Agility assessment of the method

The agility of the method used in the case study was assessed based on the criteria for an agile evaluation method. The values of different parts of the criteria for agility are summarised in Table 17. In addition to indicator values, the table provides short explanations for assigned indicator values. According to the definition of an agile method presented in Chapter 2.3 and the criteria for agility presented in Chapter 2.2, the method can be classified as an agile method. The method can be employed, for example, as part of the agile feedback loop when the development of the system is carried out according to agile principles.
Table 17. Case study 2 – Indicators for agility of the method.

<table>
<thead>
<tr>
<th>Evaluation criterion</th>
<th>Value</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>Speed</td>
<td>4 (high)</td>
<td>Data collection takes at least 1–2 weeks. In addition, time will be needed for planning the study and selection of test sites (incl. permissions to install equipment), implementation of video monitoring equipment, analysis of study results and reporting. It is expected that the method will produce results in 3–6 months. An observation period of a few weeks, including a sufficient number of relevant events, is most likely enough to reveal the impact of failures that occur frequently in the operation of the system (e.g. many communication and software related issues). It can also be expected to be long enough to conclude whether improvements to the system will be needed or whether the system is ready for further validation, e.g. a longer observation period, with more installed sites or with real end-users.</td>
</tr>
<tr>
<td>Sustainability</td>
<td>4 (high)</td>
<td>The method is based on well-known statistical methods and measures of reliability which are not expected to change in the near future. Even though information on system behaviour can be collected in many ways, there is no direct substitute for video monitoring that would be suitable for all possible implementations of the system and give an equal level of assurance of observations and equally rich information on the failure modes of the system. When more information is accumulated on the failure modes of the system, it will be necessary to check that the assumptions involved in the calculation of confidence intervals are still valid. However, this new information can most likely be taken into account by updating the calculation methods for confidence intervals if necessary.</td>
</tr>
<tr>
<td>Flexibility</td>
<td>5 (very high)</td>
<td>Video monitoring can detect and document both anticipated and unanticipated failure modes of the system, especially when it is combined with analysis of the log files of the system under study. In case of unexpected external disturbances or rare events not considered relevant to the analysis (e.g. failure of the system under study), the corresponding part of the collected data may be excluded. The data collection period may be extended if the number or rate of relevant events is less than expected. The monitoring method is not sensitive to changes made to the system under analysis. No changes are usually needed to the monitoring equipment when the system under study is updated. The method can respond appropriately to several types of anticipated events. It also has characteristics that provide resiliency in case of unexpected events.</td>
</tr>
<tr>
<td>Learning</td>
<td>2 (limited)</td>
<td>Repeating the analysis and data collection will take a substantially shorter time once the required equipment has been assembled and an analysis process for the results has been developed (e.g. criteria for a successful and unsuccessful alarm have been defined). However, improvements in efficiency do not affect the core aspects of the method even though efficiency will likely be improved.</td>
</tr>
<tr>
<td>Responsiveness</td>
<td>2 (low)</td>
<td>The progress with data collection can be reported as feedback during the study. However, this feedback has limited information content.</td>
</tr>
<tr>
<td>Leanness</td>
<td>4 (high)</td>
<td>The costs of equipment such as battery, cameras, video multiplexer will not be higher than €1000–2000. The video files can be viewed, and analysis of results can be implemented with open source software, and no software costs are therefore assumed. Selection of the physical site, implementation of monitoring and analysis of the results will likely not require more than about 2 months of working time. The costs of equipment and working time will most likely be between €20,000 and €40,000.</td>
</tr>
<tr>
<td>Lightness and simplicity</td>
<td>2 (intermediate)</td>
<td>Avidemux or other software will be needed for viewing the video recordings second by second. Skills with hardware and video monitoring will also be needed for setting up the video monitoring hardware. However, implementation of the method requires no software development.</td>
</tr>
</tbody>
</table>
4.3 eCall minimum set of data transmission – results from a field test in Finland

4.3.1 ITS evaluation problem

eCall is the European in-vehicle emergency call system. The system has been made mandatory in new type-approved vehicle models (passenger cars and vans) in Europe. A review of eCall and other in-vehicle emergency call systems has been provided elsewhere (Öörni and Goulart 2017). When the in-vehicle system (IVS) is activated manually or an accident is detected automatically by sensors, the IVS initiates an emergency call which is routed by the mobile network to the most appropriate public safety answering point (PSAP). A data packet called minimum set of data (MSD) is sent with an in-band modem through the voice channel of the circuit-switched connection from the IVS to the PSAP, before a voice connection is opened between the vehicle occupants and the PSAP. The MSD includes information on the incident and the accident vehicle such as location and timestamp. The aim of the service is to reduce the time between accident occurrence and arrival of emergency services to the location of the accident, thereby reducing the severity of accident consequences in many cases. The information in the MSD is safety critical, and users’ expectations of the reliability of the service may be high. Moreover, the ability of eCall to mitigate an accident is dependent on its ability to provide timely information on the accident, such as location, to the PSAP. It was therefore necessary to test the reliability of eCall MSD transmission in realistic operating conditions. The factors affecting the reliability of MSD transmission are discussed in Publication III.

4.3.2 Evaluation method

The successfulness of eCall MSD transmission was studied with a field test in which test calls were launched from fixed locations and from a moving vehicle. Because of the absence of an eCall-capable PSAP at that time in Finland, the test calls were regular mobile-originated calls. The tests were carried out with an eCall IVS prototype capable of initiating a test call, transmitting the eCall MSD with the in-band modem and providing information on the initiation, progress and outcomes of the test calls via a dedicated interface. The IVS had a USB serial port interface, and the data provided by the IVS was collected on a laptop with terminal emulator software (PuTTY). Relevant events were then extracted from the log files provided by the IVS and analysed together with the log files generated by the VTT eCall Testbed, which was used as a PSAP to receive the test calls. The VTT eCall testbed was connected to the fixed line network with an analogue telephone line. The tests were carried out as part of the Finnish eCall pilot organised as part of the HeERO project. The architecture of the field test is described in Publication III.

The outcomes of test calls were classified and recorded according to a classification tree presented in Figure 12. When calculating the MSD success rate, calls with no network registration or no successful call setup were excluded. In
addition to the classification of call outcomes, unsuccessful MSD transmissions were examined based on the log files provided by the in-band modem in the IVS and PSAP.

Confidence intervals for the success rate of MSD transmission were estimated separately for each test location or part of the test route. It was assumed that individual MSD transmissions initiated by the same modem prototype and within the same geographical area to the same back-end system have an equal probability of success, and that the successes or failures of individual MSD transmissions are independent of each other. With this assumption, the transmission attempts of individual MSDs were modelled as Bernoulli trials, and the number of successful receptions were expected to follow binomial distribution (Milton and Arnold 1995). The confidence intervals for the probability of success can be estimated by approximating the binomial distribution with normal distribution as explained by Milton and Arnold (1995), if the sample size is large enough \( n \times p > 5 \) and \( n \times (1-p) > 5 \), where \( n \) is the number of trials in the data set and \( p \) is the probability of success or failure. However, this method is not possible when \( p \) or \( p-1 \) is very small or the number of trials, \( n \), is limited. This is the case, for example, when the success rate of MSD transmissions equals or is close to 100%. In these cases, the confidence intervals for the MSD success rate have to be estimated using the Poisson distribution. The share of failed MSD transmissions can be assumed to follow the Poisson distribution because the probability of success or failure of each MSD transmission within the same route section can be considered to be equal and independent of the outcome of earlier MSD transmissions. In the case of the Poisson distribution, the upper confidence interval for the expected number of failed MSD transmissions can be estimated using chi-squared distribution (Hald 1952).

The study also included an analysis of the impact of the mobile network signal strength on the success rate of MSD transmission. The analysis was carried out only for MSD transmissions initiated from a moving vehicle. First, the signal strength at the time of each MSD transmission was plotted as a function of time. Second, the Wilcoxon rank sum test (Milton and Arnold 1995) was applied to determine whether any statistically significant difference in the mobile network signal strength existed between successful and unsuccessful MSD transmissions.
4.3.3 Experiences from using the method

The data collection and analysis methods allowed quantitative analysis to be carried out of the success rate of eCall MSD transmissions in real-life mobile and fixed-line networks. Moreover, confidence intervals were successfully estimated for the MSD success rate for parts of the test route and fixed locations where the measurements were carried out. However, the results were applicable only to a single pair of eCall IVS and PSAP and one mobile network operator.

The method also allowed analysis of the impact of the mobile network signal strength on the MSD success rate. The results provided by the study were applicable to signal strength values observed during the field measurements (between -80 dBm and -40 dBm) and to the IVS and PSAP equipment and the mobile network used in the study. The sample size was relatively small, which had to be taken into account when interpreting the results. With a small sample size, the Wilcoxon rank sum test does not necessarily detect differences between the categories being analysed. However, it would likely detect major differences. Because the Wilcoxon rank sum test is a non-parametric test, it was not necessary to assume or test the normality of the data.

The test calls made during the tests were initiated automatically by the IVS with a pre-determined interval without direct intervention by a human user. Because the interval between activations of test calls had been configured to too small a value and a pre-paid SIM card was used in the test, some of the initiated test calls failed without a proper network registration and call setup for a reason unrelated to eCall. The classification of test calls used in the study allowed these erroneous test calls to be excluded and the MSD success rate to be calculated only for calls with a proper network registration and call setup.

In addition, the log files of the IVS and PSAP provided further information on unsuccessful MSD transmissions: all unsuccessful MSD transmissions observed during the test were cases in which the maximum time for MSD transmission defined in EN16062:2011 (20 seconds) was exceeded.

The method involves analysis of the log files of the IVS and the PSAP. While it would be possible to analyse the log files for all test calls manually, it would be impractical for a large number of test calls, especially if the data collection and analysis is repeated several times. Different eCall IVS models and different PSAPs (or test beds emulating a PSAP) also produce their log files in different formats. The method therefore requires programming skills if it is applied with a large number of test calls in a single test or repeatedly, e.g. with different equipment models, equipment configurations or network operators.

4.3.4 Agility assessment of the method

The agility of the method used in the case study was assessed based on the criteria for an agile evaluation method. The values of different parts of the criteria are summarised in Table 18. In addition to indicator values, the table provides brief explanations for the assigned indicator values. According to the definition of an agile method presented in Chapter 2.3 and the criteria for agil-
ity presented in Chapter 2.2, the method meets the criteria for an agile evaluation method except in regard to leanness in terms of use of resources. If the implementation of the method starts from a clean table without any equipment, the cost of the study may be more than €40,000. However, if much of the equipment such as an eCall IVS and an eCall PSAP or a similar testbed are already available, the cost of applying the method may be less than €40,000. However, the equipment costs can be expected to be smaller than for methods requiring a fully equipped test laboratory (e.g. with dedicated GSM and 3G test networks).

Table 18. Case study 3 – Indicators for agility of the method.

<table>
<thead>
<tr>
<th>Evaluation criterion</th>
<th>Value</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>Speed</td>
<td>4 (high)</td>
<td>The method requires developing software which analyses the log files of the IVS and those of the PSAP if available. This requires obtaining documentation of the IVS and PSAP and checking that both of them create accurate log files of relevant events. In addition, data collection in a field test is required. It is assumed that the method can realistically be implemented in a timeframe of 3–6 months.</td>
</tr>
<tr>
<td>Sustainability</td>
<td>4 (high)</td>
<td>The method is based on well-known statistical methods. It is also difficult to obtain a high level of assurance on the operation of eCall in real-life mobile and fixed-line networks with methods other than a field test, because a complete documentation of the call path from IVS to PSAP is commonly not available. The need for the method is likely to remain as long as the eCall in-band modem or other modems with similar characteristics are used to transmit safety-critical information through the voice channel over mobile and fixed-line networks.</td>
</tr>
</tbody>
</table>
| Flexibility          | 4 (high)| ▪ Different techniques may be used for estimation of confidence intervals for the MSD success rate, if the results differ from expected (e.g. if the number of failed MSDs is very low and the normal approximation of binomial distribution cannot be used).  
▪ If software tools have been developed for processing and analysis of the log files, data collection and analysis may be repeated with limited cost and effort. This may be necessary, for example, if unexpected problems (e.g. failures of IVS or PSAP) occur during the test or unanticipated test results are obtained. |
| Learning             | 3 (substantial)| ▪ If the method is applied several times to the same system (e.g. eCall), it will be possible to adapt and tailor the parameters of the test to better measure the weaknesses of the system, for example by adjusting the test locations or the test route.  
▪ When the method is applied the first time, only limited information can be expected to be available on MSD success rates. When it is applied a second time or more, it will be possible to adjust the number of test calls to a more optimum value. If the MSD success rate is low, even a smaller number of test calls will be enough to prove that the success rate of MSDs is not acceptable. If the success rate is high, a larger number of calls may be preferable to obtain narrower confidence intervals.  
▪ Software development will be needed when the method is used the first time. The same software can in most cases be re-used if data collection and analysis is repeated.  
▪ The method has an inherent potential for learning from previous iteration rounds. |
<p>| Responsiveness       | 3 (substantial)| It is possible to provide the frequencies of initiated and successful MSDs for each geographical location or part of the test route as feedback when the analysis and reporting of results is still ongoing. |</p>
<table>
<thead>
<tr>
<th>Evaluation criterion</th>
<th>Value</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>Leanness</td>
<td>3–4 (from intermediate to high)</td>
<td>The use of the method requires an eCall in-vehicle system (IVS) and public safety answering point (PSAP) suitable for testing purposes (or a similar testbed). The cost of an IVS can be expected to be around €1000. The number of working hours required will depend on the documentation provided with the IVS as well, the data logging features and interfaces supported by the IVS, and the level of competence of the software developer and his or her knowledge of eCall. An expert with software development skills, familiarity with eCall and a suitable IVS and PSAP (or similar testbed) will need 2 months of working time or less to apply the method. With the equipment costs, this corresponds €20,000–40,000. In other cases, the cost may be more (€40,000–80,000).</td>
</tr>
<tr>
<td>Lightness and simplicity</td>
<td>1 (low)</td>
<td>Programming skills will be needed to analyse the log files generated by the eCall IVS and eCall PSAP.</td>
</tr>
</tbody>
</table>

4.4 Interoperability of eCall and ERA-GLONASS in-vehicle emergency call systems

4.4.1 ITS evaluation problem

eCall, the European in-vehicle emergency call system, is now mandatory in the EU for new models of passenger cars and vans (M1 and N1 class vehicles). ERA-GLONASS is the Russian in-vehicle emergency call system which has been made mandatory in countries of the Eurasian Customs Union (e.g. Russia). Both eCall and ERA-GLONASS provide an in-vehicle emergency call service with a largely similar technical solution. Moreover, some of the countries supporting eCall share land borders with countries supporting ERA-GLONASS (e.g. Finland and Russia), making it necessary to study the interoperability between the systems. The first step of this analysis is to identify the features of systems that can be expected to be interoperable based on the specifications of both systems. The results of the analysis also provide information on features that need further analysis, verification with empirical tests, or changes to system specifications or configuration to achieve interoperability.

The ability to make an emergency call from an in-vehicle system and to communicate the location of the accident to rescue services is a safety-critical functionality which should be available regardless of the user’s physical location, as is the case with mobile-originated emergency calls to a general emergency number (112 in the EU). In case of an accident, the location of the accident vehicle is communicated from the IVS to the PSAP in the MSD. This functionality can be expected to be especially relevant for users travelling outside their own countries, as they may be confronted with language difficulties and problems expressing the accident location in an unfamiliar environment.

4.4.2 Evaluation method

The analysis of interoperability was started by collecting the core standards of both eCall and ERA-GLONASS. Relevant use cases were then identified: an eCall IVS interacting with an ERA-GLONASS back-office system and an ERA-GLONASS IVS interacting with a PSAP supporting pan-European eCall. The next task was to identify the core features of both systems based on their specifications. The use cases, interworking scenarios, were then described as se-
sequence diagrams, with focus on the identified core features. References to the specifications of both systems were provided with each element of the sequence diagram. While sequence diagrams can be used to explain how information is communicated, they are not the best possible tools for explaining state changes of the communicating entities. For this reason, the descriptions of the interworking scenarios involved also text in addition to the sequence diagrams.

While the specifications of the systems were reviewed, the interoperability of different core functionalities of the systems were analysed. Experts on eCall and ERA-GLONASS involved on the team then validated the study results, including the list of core features for both systems, descriptions of interworking use cases and results on interoperability of different features of the systems.

For pan-European eCall, the standards of the system were available to all team members in English. For ERA-GLONASS, English translations of the Russian-language standards were used; this was not considered to pose a major risk to the validity of the conclusions, as the evaluation team involved a Russian expert on the ERA-GLONASS system.

4.4.3 Experiences from using the method

The results of the analysis were based on standard versions available in 2013. Since then, the standards of both eCall and ERA-GLONASS have been updated. While the method itself is not sensitive to changes in technology, the analysis needs to be repeated when the specifications of the interworking systems are changed. If carried out today, the results of the analysis would be different.

Analysis of the standards of the two systems answered the question of which functionalities would likely be available in the analysed interworking scenarios. However, the results were not considered sufficient to guarantee interoperability of the systems in real-life situations. In other words, analysis based on the specifications of the systems provided useful information on interoperability but could not be assumed to replace empirical interoperability tests. First, all features of real-life systems or their configured parameter values are not necessarily described in any publicly available standard or specification. Second, a real-life implementation of the system may differ from its specification. Third, it is difficult to rule out unexpected feature interactions. In telecommunications, feature interaction is understood as a situation in which “new features may interact in unexpected or adverse ways with existing features” (Cameron and Velthuijsen 1993).

4.4.4 Agility assessment of the method

The agility of the method used in the case study was assessed based on the criteria for an agile evaluation method. The values of different parts of the criteria for agility are summarised in Table 19. In addition to indicator values, the table provides brief explanations of the assigned indicator values. According to
the definition of an agile method presented in Chapter 2.3, the method meets the criteria for an agile evaluation method.

Table 19. Case study 4 – Indicators for agility of the method.

<table>
<thead>
<tr>
<th>Evaluation criterion</th>
<th>Value</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>Speed</td>
<td>4–5 (from high to very high)</td>
<td>Obtaining the specifications of both systems and other relevant ones may take some time. If the specifications are available in different languages, it may be necessary to translate core documents or obtain translated versions. In addition, some time will be needed for interaction between experts on the systems under analysis and commenting rounds on the study results. The speed of applying the method depends on the complexity of the systems under analysis, the clarity of their specifications and the level of success of collaboration between the experts. In favourable conditions, the method can provide results in a very short time (&lt;3 months). In other cases, more time (3–6 months) may be needed.</td>
</tr>
<tr>
<td>Sustainability</td>
<td>4 (high)</td>
<td>The method is based on analysis of the specifications of the systems (e.g. specifications supported by systems, messages exchanged and statuses of communicating parties) and interaction between experts familiar with the systems being analysed. In addition to eCall and ERA-GLONASS, the method can be expected to be applicable to other cases in which interoperability is claimed between two systems following different specifications. The method is unlikely to become obsolete in the near term, as it is not technology specific.</td>
</tr>
<tr>
<td>Flexibility</td>
<td>4 (high)</td>
<td>The method is not particularly sensitive to the characteristics of the systems being studied. This means that the method will usually provide correct results even if the systems are found to have characteristics different from originally anticipated. The method has flexible characteristics:</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• If new specifications or other documents are identified during the evaluation process, they can be included in the analysis in a flexible way without major difficulties.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• The method can be expected to provide meaningful results even if the systems under analysis have characteristics different from expected.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• If the scope of the evaluation needs to be changed during the process, use cases may be added or removed.</td>
</tr>
<tr>
<td>Learning</td>
<td>1 (none or very limited)</td>
<td>The method is relatively well documented in the ETSI White Paper on Interoperability. In addition, guidance for interoperability testing is provided in supplements 4 and 5 to ITU-T recommendation series X. Due to extensive documentation, the possibilities to learn from previous evaluation studies are limited. This is especially the case when previous evaluation studies focus on services which have substantially different characteristics than the service under analysis.</td>
</tr>
<tr>
<td>Responsiveness</td>
<td>3 (substantial)</td>
<td>Evaluation of interoperability based on the method documented in the paper can likely provide preliminary results as feedback during the evaluation process. A draft version of the interoperable features statement (features that are expected to be interoperable with or without conditions) and a list of features which are likely not to be interoperable may be provided before completion of the final results of the study. Information on probable challenges with interoperability and a draft version of the interoperable features statement can be expected to be relevant for the decision maker.</td>
</tr>
<tr>
<td>Leanness</td>
<td>4 (high)</td>
<td>The resources required to apply the method are dependent on the scope of the task (e.g. scope and complexity of the interworking systems), the clarity of specifications and the level of cooperation between experts on the systems. Resources corresponding to about 2 person months can be expected to be sufficient in most cases.</td>
</tr>
<tr>
<td>Lightness and simplicity</td>
<td>3 (high)</td>
<td>No specific software tools or programming skills are required for implementation of the method.</td>
</tr>
</tbody>
</table>
4.5 Demand for intelligent vehicle safety systems in Europe

4.5.1 ITS evaluation problem

Intelligent vehicle safety systems (IVSS) such as emergency braking and speed alert have been available on the market for a while. The systems are expected to improve safety by providing information to the driver or directly intervening in the driving task (Wilmink et al. 2008). So far, only electronic stability control, ESC, has been made mandatory in passenger cars by European regulation. The deployment of other IVSS is therefore dependent on the demand for the systems as well as the availability of the systems in new vehicles, and the price. The systems may be included in a vehicle model as standard features, optional features or not available at all for the vehicle model. The deployment of the systems is therefore dependent on market demand as well as the extent to which vehicle manufacturers are offering the systems as standard or optional features in their vehicle models. Information on demand may be used, for example, to estimate the effect of price changes of the systems on penetration of the systems in new vehicles. Analysing the demand for the systems also provides information which can be used to estimate the impacts of various policy measures related to in-vehicle ITS.

4.5.2 Evaluation method

There are different methods for studying and estimating consumers’ willingness to pay for a given product. A summary of these methods by Breidert, Hahsler and Reutterer (2006) mentions four main approaches, namely direct surveys, indirect surveys, laboratory experiments and analysis of market data. In this case, a direct survey was organised and respondents were asked whether they would be willing to pay and how much.

The analysis of demand for IVSS was based on data collected in the iMobility Challenge study of 2014 (Öörni and Penttinen 2014) and the earlier eSafety Challenge consumer study carried out in 2009 (eSafety Challenge 2009). The data for the iMobility Challenge study was collected in five EU member states (Czech Republic, Finland, Germany, the Netherlands and Spain), and at least 1000 responses were received from each member state. The data collection was performed by a consortium of marketing research companies with a standardised Internet survey, and the sample of respondents was taken from their national panels. The survey was answered by active car users aged 18–74 years who had driven more than 1500 km during the last 12 months. More details of the iMobility Challenge study are provided in Publication V and in the report on the study (Öörni and Penttinen 2014). The data for the eSafety Challenge (2009) study was collected with computer-assisted telephone interviews by a marketing research company. In total, 1000 interviews were carried out in five EU member states (France, Germany, Italy, Poland and the United Kingdom, 200 interviews each). More details on the eSafety Challenge study are provided in Publication V. In both consumer surveys, the respondents...
were allowed to select whether they would be willing to pay for a given system, and to select a price range within which they would be willing to pay.

The centre points of the price ranges selected by respondents and the percentages of respondents selecting the price range could be described as points in a diagram presenting the share of respondents (y axis) as a function of the mid-point of each price range (x axis). It was then assumed that all users willing to pay a given price for the system would be willing to pay also any of the lower prices used in the study. With this assumption, it was possible to calculate the share of users willing to pay for any of the price ranges used in surveys by “summing backwards” the shares of users willing to pay a given price. With this procedure, it was possible to obtain coordinate points consisting of the mid-points of the price ranges (x axis) and the shares of respondents willing to pay this price or more expressed as per cent (y axis).

Demand curves were calculated based on the coordinate points mentioned earlier. Two types of demand curve were fitted to the points calculated from consumer surveys: a linear model and an exponential model. The linear curve was fitted to the data points with the least-squares method available in standard textbooks (Milton and Arnold 1995). The exponential model was fitted to the data points by taking natural logarithms of both sides of the equation, applying the least squares method and reversing the transformation, as described in a textbook (Milton and Arnold 1995). A detailed description of the calculations is available in Publication V.

The goodness of fit was tested by calculating the mean absolute error for the percentage of users willing to pay a given price predicted by the model and the percentage of users willing to pay a given price obtained from questionnaire data. The linear and exponential demand models were compared in terms of mean absolute error for all analysed systems to determine which model better reflects car users’ self-reported willingness to pay.

4.5.3 Experiences from using the method

The clear advantages of using a direct survey to measure users’ willingness to pay are the cost-effectiveness of the method and the possibility to collect data and perform an analysis within a relatively short timeframe (Breidert, Hahsler and Reutterer 2006). Its known limitations include the validity of the results obtained from a survey, and the fact that self-reported willingness to pay in a survey does not always reflect the real purchasing behaviour or the observed choice behaviour (Breidert, Hahsler and Reutterer 2006).

Both linear and exponential demand curves (share of users willing to buy as a function of price) were successfully constructed for four intelligent vehicle safety systems: blind spot monitoring, emergency braking, lane support system and speed alert. Linear and exponential models for demand were also compared by calculating the mean absolute error for both models for all analysed systems. The exponential model used in this paper can be used to calculate the price elasticity at various points along the demand curve. However, the model does not assume that the price elasticity is constant within the relevant range.
Although the method presented here has certain advantages such as a limited amount of resources and time needed to implement it, it also has clear limitations such as the potential difference between the self-reported willingness to pay and actual markets and purchasing decisions. This means that the method can be trusted to provide estimates for the parameters described in this paper and their approximately correct magnitudes, but not their exact values.

Most errors in the results are expected to arise from the method used to measure willingness to pay and the methods applied to collect the dataset. These include respondents of surveys over- or under-reporting their willingness to pay, non-perfect representativeness of the sample, collecting the data in a subset of all EU countries and possible random variation due to a limited number of respondents. In addition, the study did not consider the difference in willingness-to-pay between buyers of new or pre-owned cars. Car users intending to buy a new vehicle most likely have more disposable income and therefore more ability and willingness to pay for new safety innovations than car users planning to buy a pre-owned vehicle. Therefore, the method may slightly underestimate the willingness-to-pay of car users intending to buy a new vehicle. In addition, respondents of direct surveys may overestimate or underestimate their willingness to pay (Breidert, Hahsler and Reutterer 2006).

The eSafety Challenge and iMobility Challenge surveys covered only their own samples of EU member states. It is possible or even likely that the results would be slightly different for data collected in all EU member states and weighted appropriately. In addition, the eSafety Challenge study was published in 2009, and the data for the iMobility Challenge study was collected in early 2014. Five years is a relatively long time in a rapidly developing field. The results based on the eSafety Challenge and iMobility Challenge studies were therefore not comparable.

The exponential model fit clearly better than the linear model the demand estimated from car users’ self-reported willingness to pay for in-vehicle safety systems. The results suggested that the demand for the systems may increase rapidly once the price paid by the consumer has dropped to a certain level. In other words, the same change – in terms of absolute price – may have very different impacts when the price paid by the user decreases. The linear model was considered suitable for estimating the relationship between price and demand for IVSS, but only within a small range from a point at which the slope of the line is known or can be estimated.

4.5.4 Agility assessment of the method

The agility of the method used in the case study was assessed based on the criteria for an agile evaluation method. The values of different parts of the criteria for agility are summarised in Table 20. In addition to indicator values, the table provides short explanations for assigned indicator values. According to the definition of an agile method presented in Chapter 2.3 and the criteria for agility presented in Chapter 2.2, the method cannot be considered an agile one. First, the flexibility of the method is limited, as most of the costs are in-
curred in the data collection phase, and the scope of the study and definitions of the systems need to be fixed when the data collection starts. Second, the method is not necessarily lean in terms of use of resources if a data set is collected solely for the purposes of analysing demand for the systems.

Table 20. Case study 5 – Indicators for agility of the method.

<table>
<thead>
<tr>
<th>Evaluation criterion</th>
<th>Value</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>Speed</td>
<td>With collection of the data set: 4 (high) With a pre-existing data set: 5 (very high)</td>
<td>Data collection with an Internet survey can be expected to take 2–3 months. A tendering round for data collection will be necessary (2–4 weeks). In addition, time will be needed for planning the survey and analysis and reporting of results. The method can be expected to provide results in 3–6 months if sufficient resources are available. If a pre-existing data set can be used, the calculations can be implemented and results analysed in under 3 months.</td>
</tr>
<tr>
<td>Sustainability</td>
<td>3 (intermediate)</td>
<td>The method is based on estimating willingness-to-pay with a direct survey and constructing linear and exponential demand curves for the systems with the method of least squares and method of least squares combined with linearisation. The method can be expected to be applicable to also other types of in-vehicle ITS systems in addition to IVSS. However, the described method has been applied to only four systems and two data sets. The experiences from applying the method are therefore limited. If the demand is found to have a relationship to price other than linear or exponential, the method is not necessarily able to construct a demand curve in those cases. However, the exponential model provided a relatively good fit for all four IVSS analysed in the study.</td>
</tr>
<tr>
<td>Flexibility</td>
<td>1 (very low)</td>
<td>A substantial part of the cost of applying the method is incurred in the data collection phase. The specifications and requirements for the data collection need to be established before the tendering round for collecting the data set can start. If the descriptions of the systems to be analysed or the target group of the study need to be changed after this or the quality of the data set turns out to be lower than anticipated, the only option available may be to repeat the data collection. This is both costly and time-consuming.</td>
</tr>
<tr>
<td>Learning</td>
<td>3 (substantial)</td>
<td>When the method is applied several times, it will be possible to target the data collection, e.g. by studying different car user or car buyer segments separately. This may improve the value of the results. Moreover, it will be possible to extend the method by testing the fit of demand data with different functions. If the method is applied several times to similar ITS systems, experience will likely be accumulated on the types of functions best corresponding to the demand for different ITS systems. However, the exponential model tested in the case study already had a good fit.</td>
</tr>
<tr>
<td>Responsiveness</td>
<td>3 (substantial)</td>
<td>Once data collection is completed, it will be possible to provide preliminary results (data in Figure 1 of the article describing the case study) and information on the successfullness of data collection. This information can be considered to be both substantial and relevant for the purposes of the study.</td>
</tr>
<tr>
<td>Leanness</td>
<td>With pre-existing data set: 5 (very high) With data set collected specifically for the purpose: 3 (intermediate)</td>
<td>The costs of applying the method consist of the cost of data collection and the cost of the working time required to plan the study, contract out the data collection and analyse the results. The costs of data collection can be expected to be a few tens of thousand Euro, if the data is collected solely for the purpose of analysing the demand. In this case, the total cost of applying the method is expected to be between €40,000 and €80,000. If a pre-existing data set originally collected for other purposes is used (as in the case study), the cost of applying the method can be expected to be limited (&lt;€20,000).</td>
</tr>
<tr>
<td>Lightness and simplicity</td>
<td>3 (high)</td>
<td>The processing of the results of the questionnaire and the calculations included in the method can be done with spreadsheet software, although GNU Octave and SPSS were used in the case study.</td>
</tr>
</tbody>
</table>
4.6 Realised safety impacts of electronic stability control in Finland

4.6.1 ITS evaluation problem

The impacts of electronic stability control (ESC) on the number of traffic accident fatalities and injuries have been studied based on accident records in Finland (Rämä et al. 2008). The study provided an estimate in per cent of how much ESC would reduce the number of fatalities and injuries in traffic accidents in a year if it were implemented in all cars. However, the result did not answer the question of how many accidents are avoided because of ESC in the current situation, or how much ESC reduces fatalities and injuries if the fleet penetration of the system is less than 100%. This information allows the comparison of ESC with imperfect fleet penetration to other safety measures in terms of impacts. It may also contribute to monitoring the impacts of implemented ITS technologies and the ITS strategy in general.

4.6.2 Evaluation method

The modelling process included two main parts. First, an estimate was made of how many passenger cars with ESC were on the road in a given year in the country under analysis and how large share of the passenger car fleet they represent. In addition, their share of the kilometrage of all passenger cars was estimated. Next, an estimate was made of how many accidents were avoided each year with a given share of passenger cars with ESC and their share of kilometrage.

One of the theoretical frameworks for analysing the impacts presented here are the dimensions of the traffic safety problem described by Nilsson (2004). In this framework, the outcomes of accidents are determined by a product of three factors: the risk (number of accidents with a given amount of exposure), consequences of accidents and amount of exposure. The impacts of ESC on safety have been summarised in a meta-analysis carried out by Høye (2011). When developing the model, ESC was assumed to be effective on several different road types and to some extent in urban environments. It was therefore possible to model the exposure simplistically by assuming that the share of vehicles equipped with ESC or the share of kilometres driven with these vehicles roughly equals the share of exposure to accident risk across the whole vehicle fleet. The modelling of impacts was carried out for passenger cars, given that they form the clear majority of all vehicles in use and data availability is best for this category.

How many passenger cars with ESC are there in Finland?

One of the difficulties facing the study was getting a reliable estimate for the number of passenger cars equipped with ESC in use in Finland. Information on the presence or absence of ESC was not available from the national vehicle registry or directly from the car manufacturers. The number of passenger cars with ESC was first estimated with a linear approximation. This assumed that
the fleet penetration of the system in new passenger cars increased linearly between the year the system was introduced to market in Finland (1997) and the year when all new passenger cars had to have ESC (2015). The linear model for the increase in ESC take-up rate has been studied earlier by Weekes et al. (2009). The applicability of the linear model to the ESC take-up rate to Europe is also supported by industry estimates (Bosch 2008).

In addition to the linear model, data from car importers was used to create another estimate of the share of passenger cars equipped with ESC at the end of 2013. This estimate was also weighed based on the annual kilometrage of ESC equipped cars.

The main data on ESC as standard equipment in passenger cars was collected from a survey sent to importers of 15 car brands covering around 90% of the car fleet in Finland at the end of 2013. However, the resulting data covered only ten car brands, requiring complementary information from a number of sources including Fildes et al. (2013), the Euro new car assessment programme, and records from car brochures and used passenger cars. Since some data sources were less reliable, the data was classified in three categories: (i) ESC as standard equipment, (ii) no documented information available, and (iii) ESC not as standard equipment.

Car fleet data was received from the Finnish Transport Safety Agency (2014). The data included more than 2.6 million vehicles in category M1 (vehicles used for carrying passengers, equipped with no more than eight seats in addition to driver’s seat, mostly passenger cars). On the basis of the data received from car importers, it was assumed that the first passenger cars equipped with ESC in Finland were introduced in 1998. The proportion of older passenger cars in the initial data was lower than in the Finnish car fleet, due to the data collection being biased toward newer cars. Specifically, the car importers reported information about recent car models and it was too challenging to identify all old car models from the car fleet data. This bias was corrected by increasing the data pool by 44,000 passenger cars purchased before 1998. The brand and model of these cars were not identified; the only critical information was that they could not be equipped with ESC. The final data included 75% of the Finnish car fleet in the autumn of 2014.

To assess the penetration of ESC as standard equipment in passenger cars in Finland in the autumn of 2014, weighted by the average kilometrage per car model and age, data was collected from compulsory annual car inspections. The data was collected by one inspection company (40% market share in Finland) from 1999. Missing information was complemented by information from the average change of mileage by car age. This assessment resulted in an average mileage of e.g. 21,000 km during the first 3 years of usage and 2000 km after 14 years.

Model for current safety impacts of ESC in Finland

The model for current safety impacts of ESC in Finland is documented in detail in Publication VI. The fleet penetration of ESC in new passenger cars was calculated first with the linear approximation described above. When the
number of cars registered each year after the introduction of ESC could be obtained from statistics, it was possible to calculate an estimate of how many cars with ESC were registered each year. The number of cars with ESC in use at the beginning of a given year could be calculated by calculating the sum of cars with ESC registered during all previous years since ESC introduction. This calculation assumed that only a negligible number of cars registered after the introduction of ESC had been scrapped before the year for which the fleet penetration was estimated. When the total passenger car fleet in use and the number of cars with ESC in the fleet are known, it is possible to calculate the share of cars with ESC of the whole passenger car fleet in per cent.

Estimation of the number of fatalities avoided started by noting the number of fatalities in the current situation and a situation without ESC as functions of the share of ESC-equipped vehicles and the assumed accident reduction because of ESC with 100% fleet penetration. A system of two equations was written for the number of fatalities in a situation without ESC and number of fatalities with imperfect fleet penetration of ESC. The difference in the number of fatalities was then calculated by subtracting the second equation from the first. It was then possible to solve the number of fatalities avoided due to ESC by rearranging and substituting the equations obtained during the previous steps. The reduction in injuries because of ESC was calculated similarly. Finally, the calculations were performed with an ESC fleet penetration rate estimated with linear approximation, ESC fleet penetration based on data obtained from car importers, and the ESC fleet based on data obtained from car importers weighted with the annual kilometrage of ESC and non-ESC vehicles. The calculations included in the model and the values used as inputs for the model are described in Publication VI.

### 4.6.3 Experiences from using the method

Three methods were applied to estimate the safety impacts of ESC and fleet penetration of ESC in passenger cars in 2014 in Finland: a model based on linear approximation, an estimation based on fleet penetration obtained from a survey of car importers, and an estimation of safety impacts based on a fleet penetration estimate weighted with kilometres driven annually. The last of these three methods is likely to provide the most accurate estimate, because the number of kilometres driven reflects better the exposure to accident risk than the number of passenger cars.

The model based on linear approximation gave a roughly correct magnitude for the fleet penetration of ESC in the year under analysis (2014). Specifically, the actual share of passenger cars equipped with ESC obtained from the survey of car importers (40%) is about 12% larger than the estimated value (35.7%). Data collection by survey was carried out only for 2014 for Finland, and a similar comparison was not possible for other countries. The results indicate that passenger cars equipped with ESC as relatively new cars accumulate far more kilometres per year than other similar cars, and therefore cover a larger share of the exposure than their relative share of the number of all passenger cars. Consequently, the estimate for fleet penetration weighted with the annual kil-
ometrage is 48% larger than the corresponding estimate based on the number of passenger cars in use only.

When estimating the exposure to risk, the model considers the kilometrage of passenger cars equipped with ESC but not the situational variables related to accidents in which ESC can be expected to be effective or not effective. The presented method can be considered a powerful tool due to its ease of implementation in different countries. However, its simplicity gives a less accurate outcome than methods including information on e.g. the distribution of kilometrage on different road types or in different geographical regions for passenger cars with or without ESC.

The model requires a very limited amount of data, which is one of its clear advantages. Specifically, only the year of ESC introduction, year of mandatory deployment, estimate of impact with 100% fleet penetration and data on passenger car registrations and passenger car fleet in use are required. While the statistical offices of most countries can provide the data on car registrations and car fleet in use, obtaining an estimate for the impact of ESC with 100% fleet penetration requires either a country-specific analysis or a literature study focused on countries similar to the one under analysis.

There are some potential sources of error in the assessment of ESC penetration that need to be addressed. First, the penetration of ESC was evaluated as standard equipment, as there is no documented information available for ESC as an option. Second, many new models are introduced prior to the start of the corresponding calendar year, while the car model data was classified by model year or first year of use. These differences resulted in a somewhat conservative assessment. Third, the data set included information about the first year of use in Finland, whereas a car might have been used first in another country and registered later in Finland, causing some random errors. Fourth, the study’s car sample covered 75% of the car fleet including the most common makes and models. Overall, ESC was first introduced in luxury and premium passenger car models and later in family and compact cars. Passenger cars excluded from the data included both types of car models. Finally, some less reliable sources were used for exploring from which year a particular car model was equipped with ESC as a standard. However, this information covered <5% of the total data.

### 4.6.4 Agility assessment of the method

The agility of the method used in the case study was assessed based on the criteria for an agile evaluation method. The values of different parts of the criteria for agility are summarised in Table 21. In addition to indicator values, the table provides short explanations for assigned indicator values. It is debatable whether the method should be classified as an agile one or not. The availability and quality of data on ESC fitment for new cars can be expected to vary between countries. The effort required to combine this information with national vehicle registration data available for research purposes can also be expected to vary between countries. According to the definition of an agile method presented in Chapter 2.3 and the criteria for agility presented in Chapter 2.2, the
The method cannot be considered an agile one in all foreseeable cases. The criterion for leaness in terms of use of resources is not met if difficulties are encountered with quality or availability of ESC fitment data or combining ESC fitment data with national vehicle registration data. However, the method can be expected to meet the criteria for an agile evaluation method if no such difficulties are likely to occur during data collection and analysis.

Table 21. Case study 6 – Indicators for agility of the method.

<table>
<thead>
<tr>
<th>Evaluation criterion</th>
<th>Value</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>Speed</td>
<td>4 (high)</td>
<td>The model requires several types of data to provide the most accurate result. Much of the data required as inputs can be obtained e.g. from the statistical office of the country under analysis or with a literature study from publicly available sources. In addition, information on registered cars equipped with the system under analysis is needed as well as kilometrage data for vehicles. Collecting the data required by the method and performing the analysis can be expected to take 3–6 months, if all required data types are readily available. The calculations included in the model can be implemented in a relatively short time, e.g. with spreadsheet software.</td>
</tr>
<tr>
<td>Sustainability</td>
<td>3 (intermediate)</td>
<td>The need to estimate the safety impacts of in-vehicle ITS systems with imperfect fleet penetration will likely remain. The method can be expected to be relevant also for other in-vehicle ITS systems. More accurate methods for estimating the impacts of the systems may be developed in future, but it is not certain how much data they will require. The method has not been validated or falsified by analysis of the impacts of ESC based on accident statistics. Future tests may increase or reduce confidence in the method. In future, the method has to be updated to take into account scrapping of ESC-equipped vehicles.</td>
</tr>
</tbody>
</table>
| Flexibility          | 4 (high) | The method can be adapted to provide a meaningful result in case of different anticipated or unanticipated events.  
• If no estimate on the impacts of ESC with 100% fleet penetration is available for the country under analysis, an estimate for a country with comparable characteristics may be used.  
• If no data is available on fleet penetration of ESC from car importers or manufacturers, linear approximation may be used to estimate the development of fleet penetration of ESC in new vehicles.  
• If no kilometrage data is available for vehicles in the country under study, a distribution obtained from the literature (e.g. statistics for another country) or assumed by the evaluation team may be used. |
| Learning             | 3 (substantial) | When the method is used, experience will be accumulated on how the required data can be most easily obtained. When the method is used and the results are published, documented parameter values (e.g. distributions of annual vehicle kilometrage and the safety impacts of ESC with 100% fleet penetration) may appear for different types of countries. This would make the method easier to apply to new countries and possibly also improve the accuracy of the results. |
| Responsiveness       | 3 (substantial) | The method can provide intermediate results such as estimates for the development of fleet penetration of ESC in new vehicles and the whole vehicle fleet. Impact estimates based on linear approximation can be provided even before all relevant data has been collected and final results have been calculated. |
Case studies

<table>
<thead>
<tr>
<th>Evaluation criterion</th>
<th>Value</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>Leanness</td>
<td>3-4 (from intermediate to high)</td>
<td>The resources required for applying the method depend on the effort required to process and analyse the ESC fitment data provided by vehicle manufacturers or importers in combination with national vehicle registration data, as well as kilometrage data for vehicles with different years of registration. A certain amount of manual work can be expected when analysing these data sets. The amount of work required will depend on the willingness of vehicle manufacturers or importers to provide information on ESC fitment and the effort required to combine this information with national vehicle registration data available for research purposes. In addition, it is necessary to collect the other input values for the method, to implement the calculations required by the method and to document the results.</td>
</tr>
<tr>
<td>Lightness and simplicity</td>
<td>3 (high)</td>
<td>The calculations and other analyses included in the method can be implemented with spreadsheet software.</td>
</tr>
</tbody>
</table>

4.7 Early adopters of emergency braking and speed alert

4.7.1 ITS evaluation problem

An emergency braking system monitors the surroundings of the vehicle and initiates braking autonomously when a frontal collision is imminent. Generally, the system also warns the driver when a collision is likely (Kulmala and Öörni 2012; Euro NCAP 2018a). Speed alert informs the driver when the vehicle exceeds the current local speed limit. The warning may be visual, auditory, or haptic (Kulmala and Öörni 2012; Euro NCAP 2018b). However, the safety benefits of the systems cannot be realised unless the systems are accepted and adopted by car users and offered by vehicle manufacturers. Some car users adopt the systems or otherwise start to use them before others. The characteristics of early adopters of the systems are relevant information for several reasons. First, the early adopters will also be the first to experience the safety benefits. Second, different driver groups may interact with the systems in different ways. Third, drivers with high annual kilometrage accumulate more exposure to accident risk than other drivers. If any system is first used by drivers with a high annual kilometrage, its coverage of the exposure to risk in terms of kilometres travelled may increase even faster than its share of the passenger car fleet. In summary, information on driver groups who are early adopters of the systems can support the analysis of the systems’ safety impacts, and provide information relevant for the planning both of policy measures intended to accelerate deployment and of safety measures for different driver groups. Analysing the characteristics of early adopters of IVSS, based on self-reported use of the systems, is relevant for several reasons. First, not all car users purchase and personally select their vehicles – they may be using a car owned by a family member, a company car or a shared vehicle. Second, surveys administered on paper or online cannot present exactly the same choices, constraints and information as those inherent in real car purchasing situations.
4.7.2 Evaluation method

Data on the use of the systems by early adopters and other active car users was collected with an online panel survey (Öörni and Penttinen 2014) during the iMobility Challenge project. The original purpose of the survey was to study users’ awareness and demand for a group of selected ITS systems. The survey targeted active car users aged 18–74 years who had driven at least 1500 km during the last 12 months.

The questionnaire is described in detail elsewhere (Öörni and Penttinen 2014). The respondents were first asked to provide driving-related background information such as annual kilometrage, road types used, and use of other transport modes. The second part of the questionnaire focused on car use, car ownership, and purchase patterns, such as the car currently used most by the respondent. This was followed by the main part of the questionnaire, which covered users’ awareness, use and willingness to pay for the systems. For each of the systems covered by the study, the first question was about the user's awareness and experience of the system: whether they had used it, heard about it or did not know the system. If the respondent did not know the system, no further questions on the system were asked (use of the system, its perceived usefulness, willingness to have it in one's next car and willingness to pay). Finally, the respondents were asked to provide information related to socio-economic background such as country of residence, gender, year of birth, number of persons in the household and the category of gross household income. In total, the online questionnaire had 55 questions, but not all questions were answered by all respondents (e.g. when the user had not heard about a system covered by the study).

As data on car users’ awareness and demand could not be collected in all EU member states, data collection was focused on five European countries. This allowed enough responses to be collected for analysis in EU countries with different economic characteristics, passenger car fleet, and physical environment. Responses were collected in the Czech Republic, Germany, Finland, the Netherlands and Spain. This group of countries covered one of the new member states of the EU (Czech Republic), two major passenger car markets (Germany and Spain), one old member state in Central Europe (Netherlands), and one of the Nordic Countries (Finland). A representative sample of about 1000 responses was collected in each member state covered by the study (N = 5048). The respondents were recruited from a panel group maintained by a marketing research company. First, the marketing research company generated a sample of its national panel group which was representative of the country’s population in terms of gender, age, and geographical region. This group was then invited to answer the online questionnaire on a web site. Only respondents belonging to the target group (18–74 years old, active car user with a kilometrage of at least 1500 during the last 12 months) were invited to participate. The respondents were not paid for participating. In some countries covered by the study, small prizes were raffled to motivate respondents.

The first step in identifying which variables affect a respondent’s adoption of emergency braking and speed alert was to calculate the corresponding Chi-
square tests for independence (Sheskin 2011). The test was calculated separately for both systems analysed in this paper (emergency braking and speed alert) and the variables describing car users (gender, age, annual number of kilometres driven and monthly household income). The calculations were performed with the SPSS statistical package. However, the Chi-square test for independence only measures the independence of two variables but does not tell which categories differ in terms of probabilities and by 'how much'.

Then, it was necessary to determine whether the sample size was sufficient for the Chi-square test in all cases. As summarised by Sheskin (2011), there is no common agreement on the acceptable sample size to be used with the Chi-square test for independence. A criterion originally suggested by Cochran and later summarised by Kroonenberg and Verbeek (2018) states that none of the cells (expected frequencies) should have a value of less than one and that at most 20% of the expected frequencies in a contingency table should have values of less than five.

Conditional odds ratios were calculated to highlight the differences between driver groups. Odds can be defined as the ratio of probabilities that the event will occur and the event will not occur (Sheskin 2011). Odds ratios were then calculated by identifying the reference category, the category with the lowest odds, and then dividing the odds calculated for other categories by the odds calculated for the reference category. Finally, confidence intervals were calculated for odds ratios other than the reference category. The means to calculate confidence intervals for an odds ratio are provided by Sheskin (2011) based on earlier studies on the topic. The calculation of odds ratios and confidence intervals for odds ratios is explained in Publication VII.

4.7.3 Experiences from using the method

Chi-square tests of independence were successfully calculated for all combinations of background variables and the two systems under study. Conditional odds ratios and their confidence intervals were calculated to find out which demographic groups indicated by the same background variable differ from each other in a statistically significant way and what is the effect size in case of different background variables attributable to respondents of the study.

However, the odds ratios and their confidence intervals should be interpreted with some caution. Due to the large number of comparisons between the odds ratios in the analysis, the possibility of committing a type I error (i.e. concluding a statistically significant difference when it does not exist) had to be considered. Type I errors when comparing the odds ratios are more likely in cases where the Chi-square test between the system use and the background variable did not show a statistically significant difference, and in cases where the confidence interval of an odds ratio is close to the reference category or the confidence interval of another odds ratio. When interpreting the results, it was also suspected that the background variables included in the study may be correlated with each other. This had to be considered when drawing conclusions based on the odds ratios.
The representativeness of the sample used in the study could not be assumed to be perfect or even close to it. First, the respondents to the questionnaire were recruited from a panel group maintained by a marketing research company. In addition, the responses were collected with an online questionnaire via Internet, which means that no responses could be received from individuals not using the Internet. Still, the sample was most likely representative in terms of gender, age, and geographical regions of the countries covered by the study. The analysis procedures used in the study, the Chi-square test of independence and calculation of odds ratios, are not particularly sensitive to imperfect representativeness of the sample.

The study results provided profiles for early adopters of speed alert and emergency braking, but the study did not even aim to prove a causal relationship between the variables related to the respondent’s background and the observed outcome, experiencing the systems at least once. In addition, the method did not aim to analyse the combined effects of more than one background variable on system use.

4.7.4 Agility assessment of the method

The agility of the method used in the case study was assessed based on the criteria for an agile evaluation method. The values of different parts of the criteria for agility are summarised in Table 22. In addition to indicator values, the table provides short explanations for assigned indicator values. According to the definition of an agile method presented in Chapter 2.3 and the criteria for agility presented in Chapter 2.2, the method cannot be considered an agile one. First, the flexibility of the method is limited, as the scope of the study has to be defined before a tendering round for data collection is launched. The scope of the study may also be limited by the content of the pre-existing data set. Second, the criterion for leanness is not met if a new data set is collected only for the purpose of identifying the early adopters of the systems.

Table 22. Case study 7 – Indicators for agility of the method.

<table>
<thead>
<tr>
<th>Evaluation criterion</th>
<th>Value</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>Speed</td>
<td>With data collection: 4 (high) With pre-existing data set: 5 (very high)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Collecting the data set with an online survey will take 2–3 months. Also, a tendering round taking 2–4 weeks will be required for subcontracting the data collection. In addition, time is needed for planning the online survey and analysing the study results. The method will likely provide results within 3–6 months, if data collection with an online survey is necessary. If a pre-existing data set can be used, the analysis can be carried out and the results reported in less than 3 months.</td>
<td></td>
</tr>
</tbody>
</table>
| Sustainability       | 4 (high) | • The need to study early adopters of new ITS systems is unlikely to disappear. With minor modifications, the method can be used also to study laggards who have not used a system that has long been on the market.  
• The methods used to analyse the results (Chi-square test and odds ratios) are unlikely to be obsolete in the near future and are not sensitive to changes in technology.  
• The analysis methods are not dependent on any specific data collection method, such as online survey. While the technical means for conducting online surveys will probably evolve over time, online surveys will likely remain as a way to collect data. |
<table>
<thead>
<tr>
<th>Evaluation criterion</th>
<th>Value</th>
<th>Comments</th>
</tr>
</thead>
</table>
| Flexibility          | 2 (low) | - The scope of the study (e.g. systems, geographical area or target group) or definitions of the systems cannot be changed once a tendering round for data collection has been launched.  
- Possible biases in the collected data set may be difficult or impossible to deal with, once the online survey has been completed (e.g. respondent groups missing from or overrepresented in the data set). However, the method is based on calculation of odds ratios and their confidence intervals. It is therefore not particularly sensitive to small biases in the sample.  
- If the results provided by the Chi-square test of independence are not statistically significant due to a limited sample size, or if the criteria for the sample size required for Chi-square test are not met, it is possible to combine categories to achieve statistically conclusive results or allow the Chi-square test to be carried out. |
| Learning             | 2 (limited) | The method is extensively documented in the article describing the case study, and it is based on relatively simple and widely used mathematical tools. However, the methods used for collecting a valid and representative data set can potentially improve if the method is applied several times to the same or comparable ITS systems. |
| Responsiveness       | 3 (substantial) | Preliminary results of the study such as the characteristics of the sample and the frequencies of responses indicating use or non-use of the studied systems can be provided relatively quickly after data collection, but before the main results of the study are completed. This feedback can be expected to be relevant in regard to the anticipated purposes of applying the method. |
| Leanness             | With pre-existing data set: 5 (very high)  
With data set collected specifically for the purpose: 3 (intermediate) | The costs of applying the method consist of the cost of data collection and the cost of working time required to plan the study, to contract out the data collection and to analyse the results. The costs of data collection can be expected to be a few tens of thousand Euro, if the data is collected solely for the purpose of studying the characteristics of the early adopters. In this case, the total cost of applying the method is expected to be between €40,000 and €80,000. If a pre-existing data set originally collected for other purposes is used, the cost of applying the method can be expected to be limited (<€20,000). |
| Lightness and simplicity | 3 (high) | The analysis of the online survey and related calculations involved in the method can be carried out with spreadsheet software. Also other tools such as GNU Octave and SPSS can be used. |
Case studies
5. Conclusions

This chapter summarises and discusses the conclusions of the thesis. The research questions and answers to them are discussed first. The contribution of the study to scientific knowledge and its implications for practice are then summarised. Finally, the validity and limitations of the results are discussed and the needs for further research outlined.

5.1 Answers to the research questions

The thesis had three main objectives. The first was to study to what extent the case studies and the evaluation methods applied in them have agile characteristics. The second was to verify practical evaluation approaches with agile characteristics for technical functioning, impacts, socio-economic benefits and interoperability of ITS applications. Finally, the aim of the thesis was to find out why and where agile evaluation methods and practices should be used. Three corresponding research questions were formulated; the answers to the questions are provided below.

Why and where should agile methods and practices be used?

The study has verified seven evaluation methods for ITS (Table 23) and identified five of them as agile, either as such or with restrictions. These five methods have been used in peer-reviewed evaluation studies on impacts, technical functioning, interoperability and user acceptance of ITS. These methods can be expected to be both agile and fit for their purposes. This result indicates that agile evaluation methods can be applied at least to evaluation of impacts, technical functioning and interoperability of ITS. However, seven case studies are unlikely to be enough to identify all ITS evaluation problems, contexts and conditions in which agile evaluation methods would likely be relevant or irrelevant.
The advantages and limitations of the methods classified as agile or agile with limitations were summarised to provide background for answering the question of why and where agile evaluation methods and practices should be used (Tables 24 and 25). The advantages and limitations of agile methods were identified from the case studies in which agile methods were applied and a literature study on evaluation methods for ITS (Chapter 3). Some of the advantages and limitations have been described in the articles describing the case studies, others were identified during the thesis work. The agile characteristics identified in Chapter 4 are not repeated in Tables 24 and 25.

Table 23. Results of the analysis of case studies.

<table>
<thead>
<tr>
<th>Publication</th>
<th>Case study</th>
<th>Agile method?</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td>Model for the safety impacts of road weather information services available to road users and related socio-economic benefits</td>
<td>Yes</td>
<td>The method meets the criteria for an agile evaluation method in terms of speed, flexibility and leanness.</td>
</tr>
<tr>
<td>II</td>
<td>Reliability of an in-vehicle warning system for railway level crossings – a user-oriented analysis</td>
<td>Yes</td>
<td>The method meets the criteria for an agile evaluation method in terms of speed, flexibility and leanness.</td>
</tr>
<tr>
<td>III</td>
<td>eCall minimum set of data transmission – results from a field test in Finland</td>
<td>Yes (with limitations)</td>
<td>Leanness subject to availability of staff with software development skills and familiar with eCall and availability of a PSAP or corresponding testbed for free or at limited cost.</td>
</tr>
<tr>
<td>IV</td>
<td>Interoperability of eCall and ERA-GLONASS in-vehicle emergency call systems</td>
<td>Yes</td>
<td>The method meets the criteria for an agile evaluation method in terms of speed, flexibility and leanness.</td>
</tr>
<tr>
<td>V</td>
<td>Demand for intelligent vehicle safety systems in Europe</td>
<td>No</td>
<td>The criterion for leanness is not met when a data set is collected only for the purpose of analysing the demand. Also flexibility is limited: the scope and target group of the study needs to be fixed before data collection can start.</td>
</tr>
<tr>
<td>VI</td>
<td>Realised safety impacts of electronic stability control in Finland</td>
<td>Yes (with limitations)</td>
<td>The criterion for leanness is not met if difficulties are encountered with combining data on ESC fitment in vehicle models with national car registration data.</td>
</tr>
<tr>
<td>VII</td>
<td>Early adopters of emergency braking and speed alert</td>
<td>No</td>
<td>The criterion for leanness is not met when a data set is collected solely for the purpose of identifying early adopters of systems. Moreover, the flexibility of the method is limited, as the scope of the study and the target group need to be determined before a tendering round for data collection can be opened.</td>
</tr>
</tbody>
</table>

Table 24. Advantages and limitations of agile evaluation methods.

<table>
<thead>
<tr>
<th>Method</th>
<th>Advantages</th>
<th>Limitations</th>
</tr>
</thead>
<tbody>
<tr>
<td>impact estimation based on a literature study and mathematical modelling based on the power model described by Nilsson (2004) (Publication I)</td>
<td>- The model can provide quantitative estimates for safety impacts and related socio-economic benefits of road weather information services with a minimal amount of collected information (Publication I). - The model can easily be adapted to different conditions and services (Publication I). - The results provided by the model are more transparent than expert assessment without supporting documentation.</td>
<td>- The model most likely gives a somewhat correct range of socio-economic benefits but it will not provide an exact result (Publication I). - The results provided by the model will be most accurate in countries in which factors contributing to weather related accidents are similar to Finland (Publication I).</td>
</tr>
</tbody>
</table>
The use of agile evaluation methods for evaluation of impacts, technical functioning and interoperability of ITS should be considered when:

1. the results provided by the method correspond to the information needs present in the case,
2. the advantages and limitations of agile evaluation methods as listed above are well understood,
(3) the requirements specific to the agile evaluation method can be met (e.g. availability of suitable data or experts), and
(4) the characteristics of the agile evaluation method correspond to user preferences better than the characteristics of other methods available for the task, due to the agile characteristics or other advantages.

First, an agile evaluation method needs to be capable of providing information required by the user. For example in the case of impact assessment of ITS, an agile evaluation method may provide a somewhat correct range of safety impacts or an estimate (Publications I and VI) based on mathematical modelling and information available in the literature, but not an exact value. In the case of analysis of interoperability, a review of the specifications of the systems will provide a list of systems’ features that can be expected to be interoperable but it cannot guarantee real-life interoperability of the interworking systems (Publication IV).

Second, the advantages and limitations of the agile evaluation method being considered need to be understood before an informed decision can be made on its use. Experience with agile evaluation methods indicates that their advantages and limitations are largely method-specific: the expert applying the method must be able to understand them before choosing the method for the case at hand. Third, applying an agile evaluation method is possible only when requirements specific to the individual method are met. In practice this means e.g. availability of the required source data and competent experts.

Finally, the characteristics of an agile evaluation method should correspond to user needs better than the characteristics of other agile or non-agile evaluation methods. This condition is fulfilled also in situations in which the agile evaluation method is the only possible one, e.g. due to data, time or resource constraints, and obtaining the same or corresponding results with other evaluation methods is not possible. By definition, agile evaluation methods offer advantages in terms of flexibility, leanness and speed. Because agile evaluation methods for ITS can be used for very different problems in different contexts, this comparison needs to be made each time agile evaluation methods for ITS are considered.

Can any simplified or simplistic approaches for evaluation of ITS be considered to be agile methods?

Simple or simplistic methods have a limited level of complexity in their implementation and application. In this case, lightness and simplicity, part of the criteria for agility, were used to measure the absence of complexity in the implementation and application of an evaluation method (Table 26). The methods applied in the case studies are summarised in the table in terms of their agile or non-agile status and their lightness and simplicity.
Table 26. Lightness and simplicity of agile and non-agile evaluation methods.

<table>
<thead>
<tr>
<th>Evaluation methods</th>
<th>Lightness and simplicity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Agile</td>
<td>1 (low)</td>
</tr>
<tr>
<td>Agile (with limitations)</td>
<td>II</td>
</tr>
<tr>
<td>Non-agile</td>
<td>V, VII</td>
</tr>
</tbody>
</table>

Not all evaluation methods with high levels of lightness and simplicity were found to be agile. The methods applied in case studies V and VII were found to have high levels of lightness and simplicity but did not meet the criteria for an agile evaluation method in terms of leanness and flexibility. This result indicates that simple or simplistic evaluation methods are not necessarily agile, as the simplicity of the evaluation method does not necessarily imply that the method is flexible or lean.

**How could agile principles contribute towards the development of better evaluation methods and tools for ITS?**

The definition of agile characteristics for evaluation methods for ITS, measurable criteria for agility, and the definition of an agile evaluation method potentially improve the understanding of requirements for evaluation methods. They can be used to express the user needs for evaluation methods, and better articulated user needs potentially lead to evaluation methods providing more value to the user.

Agile methods for evaluating the technical functioning of an ITS system can potentially serve as building blocks of an agile development model for ITS systems. Due to their agile characteristics such as leanness and speed, they can be expected to be suitable for development processes proceeding through several iteration rounds.

Agile methods for impact assessment allow a more transparent and systematic impact assessment to be carried out in situations in which expert opinion has so far been the preferred method of impact assessment. This includes, for example, cases in which financial resources, time or available data on the impacts of the system are limited but the data required for applying an agile evaluation method is available. Agile evaluation methods may also be used in combination with expert assessment, for example, when more complete evaluation results are needed. Agile evaluation methods for impact assessment also have the potential to evolve and improve over time with experience and validation. Over the longer term, this potentially allows more accurate impact estimates for ITS to be made under limitations of time, resources or data. Agile evaluation methods for impacts of ITS can also make information available from earlier studies better adapted to the decision-making context. In practice, this means estimating the impacts of an ITS system in a given country or other geographical region with full or imperfect fleet penetration.
5.2 Contribution to scientific knowledge

Agility is a new viewpoint for classification of evaluation methods for ITS. While agile methods have been developed in the contexts of software development and systems engineering, they have not been extensively studied before in the context of development or evaluation of ITS systems. This research has introduced descriptions for agile characteristics of evaluation methods for ITS, developed measurable criteria for the agile characteristics, and suggested a definition for an agile evaluation method. These contributions have also been validated by applying them in seven case studies.

In addition to contributions related to identification of agile characteristics in ITS evaluation methods, the study has verified and assessed seven practical evaluation approaches for ITS. Two of these are related to impact assessment, two to technical functioning of an ITS system or technology, one to interoperability, one to estimating demand and one to identifying early adopters of the systems.

5.3 Implications to practice

When agile software development and systems engineering methods are applied to ITS systems, methods for verifying the technical functionality of systems and subsystems need to be suitable for the agile approach. For example, verifying the technical functionality of a component or subsystem in the agile development model cannot depend on documentation that is not necessarily produced or kept up to date during an agile development process, and applying a method that is expensive or time-consuming to implement can be difficult in an agile development process potentially involving several rounds of development, testing and feedback. It therefore seems that there is a need for testing methods and verification approaches that can be applied as part of an agile development process.

Case studies on the in-vehicle warning system for railway level crossings and eCall MSD transmission focused on analysis of reliability and service quality in combination with empirical observations on system behaviour. The approach used in the case studies for verification of technical functioning requires that one is able to tell acceptable output or system behaviour from unacceptable. However, the approach does not necessarily require extensive documentation on the requirements for the system being analysed, as long as the intended functionality, main requirements and usage context of the system are known. This makes it potentially relevant in situations in which the agile approach is employed in the development of ITS systems.

The need to study the impacts of an ITS system may arise in contexts in which the resources, data or time for performing an evaluation are limited but an approximately correct range or estimate of the safety impact or socio-economic benefits is sufficient. Agile evaluation methods for estimating the safety impacts of ITS can be expected to be answer users’ needs especially in these situations.
5.4 Discussion

The thesis has provided descriptions of agile attributes for evaluation methods for ITS, related assessment criteria and a definition of an agile evaluation method. These contributions have been validated by applying them to seven case studies. In addition, seven evaluation approaches for impacts, technical functioning, interoperability and user acceptance of ITS have been verified. Three of these were identified as agile evaluation methods and two as agile with limitations.

The need for more agile evaluation methods has been identified in earlier research. Most of the agile characteristics of evaluation methods also correspond to requirements for evaluation methods for ITS described more or less directly in earlier studies. Flexibility has been identified as a requirement for evaluation methods for ITS by Newman-Askins, Ferreira and Bunker (2003). Kolosz and Grant-Muller (2015) highlighted that "it is important that balance exists between cost and complexity of evaluation, as well as the cost of the project itself," which indicates a need for simplicity and leanness. Hill (2018) also criticised the FESTA approach for its prescribed and linear nature and lack of feedbacks from later stages of the process to earlier ones. This corresponds to responsiveness and flexibility (Tables 12 and 13). The agile characteristics identified in the thesis are therefore not completely novel – even in the context of evaluation of ITS – or without connections to earlier research. Instead, they are a contribution to the discussion on the requirements for evaluation methods for ITS and the need for more agile evaluation methods being better able to answer users' needs.

The development of definitions of agile attributes, related assessment criteria and the definition of an agile evaluation method involved certain trade-offs. First, the criteria for agility of an evaluation method were a compromise between the need for accuracy and the need to minimise the impact of subjective or contextual influences. While measurement scales with more steps would convey more information, assigning the most appropriate value for an individual evaluation method would also become a more complex task and more open to interpretation. It was also a challenge to develop criteria suitable for evaluation methods with different purposes. For this reason, the agile characteristics of evaluation methods and the criteria for agility had to be described in a rather generic form.

Seven case studies were used to validate the definitions of agile attributes, related assessment criteria and the definition of an agile evaluation method. Even though the number of case studies was relatively small, it was most likely sufficient for at least initial validation of new terminology and related assessment criteria. This conclusion is supported by the diversity of the set of case studies covering different aspects of ITS applications and a wide range of ITS systems.

The assessment criterion for leanness of an evaluation method can be expected to be at least somewhat sensitive to the context in which an evaluation study is carried out. For example, there are differences between countries in the costs of expert work, resources available for evaluating ITS applications...
and procurement processes of public sector and private organisations. The applicability of the monetary values used in the criterion for leanness should therefore be checked when the criterion is applied in future. For example in countries with characteristics different from Finland, the criterion for leanness may need to be adjusted to local conditions.

Experiences gained during the seven case studies and their analysis made it possible to answer the research question of why and where agile methods and practices should be used. For the evaluation methods used as case studies, detailed answers could be provided with information on the limitations, advantages and agile characteristics of the methods. However, the number of analysed case studies was relatively small, and the analysed case studies focused on different aspects of ITS and different ITS systems. This limited the generalisation of the results to all agile evaluation methods with inductive reasoning. It is therefore possible that a more detailed answer could be provided in the future when more experience with the use of agile evaluation methods has been accumulated.

The results (Table 26) indicated that not all simple or simplistic evaluation methods can be called agile methods. This conclusion does not, however, answer the question of whether agile evaluation methods are more likely also to be simple or simplistic. The small number of case studies did not allow any conclusive statistical analysis of this relationship, but the results (Table 26) do not indicate a strong relationship between these two attributes.

There is also the question of whether the objectives defined for the study helped answer the research questions in the best possible way. The objectives were derived from the questions but were identified early in the writing process, before detailed knowledge of the results became available. At the time, it was not yet clear which of the potentially relevant objectives for the study could realistically be achieved based on the case studies available.

For example, the research objectives did not include ascertaining how agile evaluation methods should be designed, raising the question of whether or not this inclusion should have been done. On one hand, doing so would have helped answer the question of how agile principles could contribute to developing better evaluation methods and tools for ITS. On the other, it was necessary to limit the scope of the study, and during the early stages of writing it was still unclear whether the set of case studies would be sufficient for this or not.

One of the tasks during the study was to create a definition of an agile evaluation method. A positive definition indicating a binary conclusion was created. This approach was chosen to keep the terminology of the study as clear, descriptive and understandable as possible. One might ask whether the binary logic of belonging or not belonging to a category is too heavy a generalisation and whether some other type of definition would have been more appropriate.
5.5 Need for further research

More validation will likely be needed for the descriptions of agile characteristics, as well as the criteria for agility of an evaluation method. They were developed mostly on the basis of a literature study and the experience obtained during the seven case studies included in the thesis. However, it is possible that the descriptions of agile characteristics and the criteria for agility can be refined further when they are applied to a larger number of evaluation methods and case studies or when they are compared to experiences obtained from other evaluation studies on impacts, technical functioning, acceptance or interoperability of ITS. This applies also to the definition of an agile evaluation method. The current definition with a binary classification outcome should be validated by applying it to a larger number of evaluation methods and case studies.

Analysis should be carried out on how to employ agile development, testing and systems engineering methods in the development of ITS systems in an optimum way. In addition, further efforts should be made to connect the case studies presented in the thesis to other research on agile testing or the way agile testing should be carried out for ITS systems. In other words, further research will be needed to answer the question of how to design or choose an agile evaluation method for evaluating an ITS application in a given context.

The study included two case studies (Publications I and VI) in which the safety impacts of an ITS system were estimated using a method with agile characteristics. Both of these methods can be expected to provide the correct magnitude of the safety impact, but they are not intended to replace a complete impact assessment based on empirical data, such as accident statistics or traffic conflict counts. The impact estimates obtained with these methods should be validated by comparing them to the results of impact assessments based on accident statistics or data collected on user behaviour.
Conclusions
References


https://www.agilealliance.org/glossary/sustainable/ [accessed 24th May 2018]


https://www.ideals.illinois.edu/bitstream/handle/2142/46345/TES-117.pdf?sequence=2&isAllowed=y [accessed 29th June 2018]


http://www.etsi.org/deliver/etsi_en%5C302600_302699%5C302665%5C01.01 ./01_60%5Cen_302665v010101p.pdf [accessed 10th August 2015]


105

References


References


References


References


Successful deployment of ITS technologies requires information on their impacts, socio-economic profitability, technical functioning and other characteristics. This thesis presents seven case studies on the evaluation of impacts, technical functioning, interoperability and user acceptance of ITS. The work also provides definitions of agile attributes for evaluation methods for ITS, measurable criteria for agility as a characteristic of an evaluation method, and a definition of an agile evaluation method. These contributions are validated by applying them to the case studies included in the thesis.