Petri Mononen

SOCIO-ECONOMIC IMPACTS OF A PUBLIC AGENCY – ENHANCING DECISION SUPPORT FOR PERFORMANCE MANAGEMENT
Mononen, Petri, Socio-economic impacts of a public agency – enhancing decision support for performance management.
University of Oulu Graduate School; University of Oulu, Faculty of Technology
University of Oulu, P.O. Box 8000, FI-90014 University of Oulu, Finland

**Abstract**

Pressures to cut public expenditure and to maximize value-adding use of scarce governmental resources are evident across the globe. Decision support tools for pinpointing when the activity is net beneficial are lacking, i.e. accountability is urgently called for but the process for recognizing and validating the best choices in direction and control are yet in their development. The objective of this research is to provide the science in general, and national administrations’ management and overseeing bodies in particular, a new understanding of how well the tasks and missions of the administration are fulfilled in terms of enhanced socio-economic well-being and adding to the value of different functions of society.

This study approaches public performance management from the viewpoint of impact evaluation of a service oriented public product portfolio. The adopted research approach combines both qualitative and quantitative research methods. The research presents a process for evaluating a public organization’s overall benefit to cost ratio, with a case study within transport safety administration.

The key findings of this research are that the socio-economic efficiency of a public entity can be quantified and that an operation of a public agency can be socio-economically net beneficial. Practical implications include re-allocation of resources to other more beneficial fields of responsibility, redesign of performance agreements and their indicators and initiating discussion in joint design of sustainable and balanced operational and strategic targets across government. Further research can exploit the methods presented in this study, as it has overcome some of the hurdles in socio-economic evaluation of public entities. With a clearer understanding of performance, government is able to better achieve its mission and its task of ensuring that agencies and other public bodies are functioning efficiently and properly.

**Keywords:** administration, cost-benefit analysis, decision support, evaluation, evaluation criteria, evaluation process, impact analysis, performance management, policy objective, public agency, public service, transport
Mononen, Petri, Julkisen organisaation yhteiskuntataloudelliset vaikutukset – päätöksenteon työkaluja tulosohjaukseen.

Oulun yliopiston tutkijakoulu; Oulun yliopisto, Teknillinen tiedekunta


Oulun yliopisto, PL 8000, 90014 Oulun yliopisto

Tiivistelmä

Paineet julkisten menojen leikkaamiseksi tai hallitsemiseksi sekä julkisten resurssien tehokkaamman käytön varmistamiseksi kasvavat maailmanlaajuisesti. Julkisen toiminnan tulosohjausta tukevat päätöksenteon työkalut ovat vielä puutteellisia ja varhaisessa kehitysvaiheessa, samalla kun toiminnan kustannusvastaavuutta ja tehokkuutta vaaditaan. Tämän tutkimuksen päätavoite on tuottaa uutta tietoa tieteelle ja myös julkishallinnolle siitä, kuinka hyvin viranomaisille asetetut tavoitteet ja tehtävät toteutuvat yhteiskuntataloudelle tuotettavan lisäarvon kannalta, sekä siitä miten lisäarvoa kyetään mittaamaan.

Tämä tutkimus lähestyy tulosohjauksen tutkimusta ja kehittämistä palvelukeskeisen julkosorganisaation sekä sen palveluvalikoiman yhteiskunta-taloudellisen vaikuttavuuden arvioinnin kentän. Valittu lähestymistapa yhdistee laadullisia ja määrällisiä tutkimusmenetelmiä. Tutkimuksessa esitettävät prosessi ja menetelmävalikoimaa julkisviranomaisen kustannushyötyystähteen määrittelemiseksi Liikenteen turvallisuusviraston tapauksesta käsittelevän tapakesimerkin avulla.


Asiasanat: arviointi, arviointikriteeri, arviointiprosessi, hallinto, hyöty-kustannus- analyysi, julkinen palvelu, julkinen virasto, liikenne, politiikan tavoite, päätöksenteon tuki, tulosohjaus, vaikutusarviointi
“Unless we know and settle these things, we shall set out without chart or compass.”

Woodrow Wilson 1887, detailing the needs for and the prerequisites of structured research on administration efficiency in his essay *The Study of Administration*.

*Dedicated to my dear family. Henrika, Aleksi and Julius – without your love and support this journey would never have reached its destination.*
Acknowledgements

I would first like to express my humblest and deepest gratitude to my supervisors, Professor Harri Haapasalo and Docent Pekka Leviäkangas for giving me this chance, but also for showing me the way and supporting my research throughout the length of it and for your extremely helpful contributions to the articles and to this dissertation. You two were the ones who originally entertained the idea and the opportunity to commence this kind of research entity, and you were the ones who thought I would be the best person to carry it out. I hope that you can now conclude that your trust was well placed – thank you ever so much.

I am very grateful to several other people, including all my colleagues from University of Oulu, as well as to all the Trafi representatives who have offered their support, comments, and contribution to my research work and publications. I would like to extend my gratitude especially to Trafi management for a true desire for this kind of research to be carried out and, furthermore, for the opportunity to do all of it with full academic freedom. This key group included Kari Wihlman, Sami Mynttinen, Marko Sillanpää, Juhani Nikula, Anu Ylä-Pietilä, Iina Rinne, Mia Nykopp, Yrjö Mäkelä, Tom Wilenius, Taneli Pelttari and Matti Tupamäki, just to name the most focal persons that I had the pleasure of actively working with. (Matti, remember our long discussion – you as the ‘been-there-done-that’ and me as the ‘up-to-my-eyebrows-in-it’ – on dissertation writing and thinking processes, the challenges, the inevitable moments of despair and the promises or at least hopes for rewards looming at the end and driving us forward? Priceless…)

Special thanks belong to the members of my follow-up group Assistant Professor Kirsi Aaltonen and Docent Jukka Majava for your contribution within and between our meetings. The support and input of Professor Jaakko Kujala, Dr. Janne Härkönen and Professor Josef Jablonsky should not be forgotten either; thank you. Thanks also to Tapio Kinnunen for our mutually fruitful cooperation with regard to the theoretical discussion areas and practices in the field of cost benefit analysis.

And to the rest of my former colleagues at IEM, thank you for your cooperation. It was a pleasure to work with you and I hope that occasions keep arising so that we can continue doing so also in the future. Do keep up the nice and pleasant but still exceptionally progressive ambience! I would like to thank Trafi, the University of Oulu and IEM for funding my research, as well as the University of Oulu Graduate School for providing the general framework for the studies. I would also like to express my gratitude to Tauno Tönning Foundation, which granted me
financial support during my doctoral studies and thesis work. Thank you all for your support for my research.

Special thanks go to the pre-examiners Professor Corinne Mulley (The University of Sydney) and Docent Jari Kauppila (OECD and University of Turku) for their effort and supportive critique of my thesis. It was truly a privilege to have such dedicated pre-examiners who devoted their undivided time and attention to this work and brought in exceptional value adding insight with their contribution.

The most important assets that have helped me to reach this goal are my family and my network of close ones. Without you, I could not have been able to do this and go this far. Parents and in-laws, your support and encouragement have been invaluable, not only for this dissertation but for everything that you have done for me. I know that you are aware of the worth of this endeavor and I hope you can be proud of me. I also highly value the friendship that the rest of my family and friends have shown me during these years.

Last – but as far from the least as possible – thanks go to my dear wife Henrika and my beloved sons Aleksi and Julius for all the support and love you have given me. Henrika, in addition to this having been an opportunity, the decision to embark upon this massive process also called for some challenging sacrifices. Therefore, I am especially happy that you fully supported me and we could make this decision together as the team that we are. All three of you, thank you from the bottom of my heart for believing in me, motivating me and making all those countless hours and numerous months and years during this endeavor possible and worthwhile. Without a question it would have been really wonderful to be able to devote to you all of that ‘spare time’ of mine, including all of the vacations and many weekends and nights during the past years – but as they say, you cannot both save the cake and eat it. But finally, now it is time to eat the cake and enjoy!

Oulu, Finland, September 2017

Petri Mononen
# List of abbreviations

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Full Form</th>
</tr>
</thead>
<tbody>
<tr>
<td>AHP</td>
<td>Analytical Hierarchy Process</td>
</tr>
<tr>
<td>B/C</td>
<td>Benefit to Cost (ratio)</td>
</tr>
<tr>
<td>BCA</td>
<td>Benefit to Cost Analysis</td>
</tr>
<tr>
<td>CBA</td>
<td>Cost-Benefit Analysis</td>
</tr>
<tr>
<td>CO2</td>
<td>Carbon dioxide</td>
</tr>
<tr>
<td>OECD</td>
<td>Organisation for Economic Co-operation and Development</td>
</tr>
<tr>
<td>DM</td>
<td>Decision Matrix</td>
</tr>
<tr>
<td>EC</td>
<td>European Commission</td>
</tr>
<tr>
<td>EP</td>
<td>European Parliament</td>
</tr>
<tr>
<td>ERP</td>
<td>Enterprise Resource Planning and Management System</td>
</tr>
<tr>
<td>EU</td>
<td>European Union</td>
</tr>
<tr>
<td>FMS</td>
<td>Financial Management System</td>
</tr>
<tr>
<td>FMA</td>
<td>Finnish Maritime Administration</td>
</tr>
<tr>
<td>FTA</td>
<td>Finnish Transport Administration</td>
</tr>
<tr>
<td>GHG</td>
<td>Green House Gas</td>
</tr>
<tr>
<td>IATA</td>
<td>International Air Transport Association</td>
</tr>
<tr>
<td>ICAO</td>
<td>International Civil Aviation Organization</td>
</tr>
<tr>
<td>ICT</td>
<td>Information and Communication Technologies</td>
</tr>
<tr>
<td>IEM</td>
<td>Industrial Engineering and Management</td>
</tr>
<tr>
<td>IMO</td>
<td>International Maritime Organization</td>
</tr>
<tr>
<td>IT</td>
<td>Information Technology</td>
</tr>
<tr>
<td>ITS</td>
<td>Intelligent Transportation System</td>
</tr>
<tr>
<td>MBO</td>
<td>Management by Objective</td>
</tr>
<tr>
<td>MCDA</td>
<td>Multiple Criteria Decision Analysis</td>
</tr>
<tr>
<td>MoF</td>
<td>Ministry of Finances Finland</td>
</tr>
<tr>
<td>MoJ</td>
<td>Ministry of Justice Finland</td>
</tr>
<tr>
<td>MoTC</td>
<td>Ministry of Transport and Communications Finland</td>
</tr>
<tr>
<td>NPM</td>
<td>New Public Management</td>
</tr>
<tr>
<td>RQ</td>
<td>Research Question</td>
</tr>
<tr>
<td>SAFA</td>
<td>Safety Assessment of Foreign Aircrafts</td>
</tr>
<tr>
<td>Trafì</td>
<td>Finnish Transport Safety Agency</td>
</tr>
<tr>
<td>UN</td>
<td>United Nations</td>
</tr>
<tr>
<td>UniOGS</td>
<td>University of Oulu Graduate School</td>
</tr>
<tr>
<td>WTA</td>
<td>Willingness to Accept</td>
</tr>
<tr>
<td>WTP</td>
<td>Willingness to Pay</td>
</tr>
<tr>
<td>Acronym</td>
<td>Full Name</td>
</tr>
<tr>
<td>---------</td>
<td>---------------------------</td>
</tr>
<tr>
<td>WEF</td>
<td>World Economic Forum</td>
</tr>
<tr>
<td>WHO</td>
<td>World Health Organization</td>
</tr>
</tbody>
</table>
List of original publications

This dissertation is based on the following publications:


The author of this doctoral dissertation is the corresponding and primary author of all of the above listed original publications. He has also had the primary responsibility for the entire research process, from forming the research problems, reviewing and collecting the theoretical base, and forming the research questions, to collecting and analyzing the empirical data and material. However, the roles and efforts of the co-authors have been valuable and have mainly included participating in the design process of the study, as well as planning, reviewing and commenting the article manuscripts of the corresponding author.
## Contents

Abstract

Tiivistelmä

Acknowledgements 9

List of abbreviations 11

List of original publications 13

Contents 15

### 1 Introduction

1.1 Background and research environment ................................................... 17

1.2 Research problem, objective and scope .................................................. 22

1.3 Research design ....................................................................................... 23

1.4 Research process, methodology and dissertation structure ..................... 27

1.5 Research data collection .......................................................................... 33

### 2 Theoretical foundation 37

2.1 Theoretical framework ............................................................................ 37

2.2 Public administration and public policy .................................................. 38

   2.2.1 Public administration: reform and research .................................. 39

   2.2.2 Public policy: design, implementation and analysis ..................... 41

2.3 New public management and performance management regimes in public domain ................................................................. 44

   2.3.1 New Public Management ............................................................. 46

   2.3.2 Management by Objective ............................................................ 49

   2.3.3 Evaluating success or failure of New Public Management .......... 51

2.4 Measuring impacts, impact evaluation of organizations ......................... 54

   2.4.1 Measuring impacts and impact evaluation ................................... 55

   2.4.2 Cost Benefit Analysis ................................................................... 61

2.5 Synthesis of literature review .................................................................. 71

### 3 Research contribution 77

3.1 Performance management system architecture and its coverage .......... 77

3.2 Service architecture and the first steps towards investigating socio-economic impacts ........................................................... 84

3.3 Methods for determining cost efficiency of public services .......... 90

3.4 Cost efficiency of a public agency ......................................................... 94

3.5 Results synthesis ................................................................................... 101

### 4 Discussion 105

4.1 Theoretical implications ......................................................................... 105

15
4.2 Practical implications ................................................................. 122
4.3 Reliability, validity and transferability ....................................... 128
4.4 Recommendations for further research ..................................... 142

5 Summary ....................................................................................... 145
References ....................................................................................... 147
Original publications ................................................................. 159
1 Introduction

1.1 Background and research environment

Background

Most economies are facing increasing challenges in providing public services. Resources for public service production do not seem to grow along same gradient with other forms of economic growth. The challenges differ among the various economies and so do the avenues of solutions. Public agencies and their services are under greater scrutiny than ever before. One key reason for this is the trend of new public management (NPM) that has taken rapidly advancing steps across the globe, including in Europe (Hague 2001; Gruening 2001). The ideological shift towards more market-oriented economy also with regard to public services is often dated to Margaret Thatcher’s governments in the UK. The ideas of new models of public governance also brought along the metrics of private sector management, meaning that the outputs and effectiveness of public services need to be assessed properly in order to facilitate efficient management control, distinguishing it from political control.

Finland in general, and the Finnish transport sector in particular, are prime examples of these phenomena. The NPM approaches were adopted by the Finnish government in the late 1980’s, when the first waves of privatization and commercialization of former administrative and public functions took place. Since then Finland’s successive governments have persistently followed this path, and restructuring of public services have been carried out rather systematically (Finnish Ministry of Justice 2009; Finnish Ministry of Finance 2012; Finnish National Audit Office 2013). The reforms or re-structuring measures have brought with them many performance management tools and principles, the applicability and effectiveness of which is still under debate (Dan 2014; Maugeri and Metzker 2013). The majority of the services are perceived to be important and necessary, but there is no overall understanding of which services are critically necessary, which services could be considered beneficial but not absolutely critical, and which maybe even relics of past administrations.

In general, the transport sector has been in the forefront of the renewal processes of different administrative sectors. For example, the main unit of analysis of this research (Finnish Transport Safety Agency, hereafter referred to as Trafi or
‘the agency’) is a fairly new entity in the national transport sector governance architecture. The historical evolution of the agency to its present form has been a series of radical privatizations, restructuring measures and mergers of agencies (Mononen et al. 2014).

Despite the systematic restructuring and search for well-functioning administrative architectures, several challenges remain in the structure, such as the following:

– challenges in assessing the efficiency and effectiveness of public services;
– difficulties in drawing clear-cut lines between fully public, semi-public, and private (or to-be-private) services;
– lack of clarity and understanding on how to identify non-value adding services (e.g., artefacts from previous governments that are no longer valid due to the surrounding operational and political environment having changed over time);

and

– problems in setting effective, unequivocal, transparent and fair performance targets to manage the agency in line with policies and strategies.

The evaluation of any public agency’s services would need to serve the purpose of providing administrations’ management and the overseeing body with the information on how well the tasks and missions of the agency are fulfilled for example, in terms of enhanced socio-economic well-being and adding to the value of different functions of the society. Without a clear perception of performance, the government as a whole is in fact unable to do its job in ensuring that agencies and other public bodies are functioning properly. This management control is in part associated with the need for continuous change, as pointed out by the Organisation for Economic Co-operation and Development OECD. OECD (2010) complimented Finland’s agility and pioneering in renewing the public administration sector but simultaneously underscored the importance of continued development efforts in the public domain to retain a government that has a good reputation and meets its responsibilities.

In Finland, prior evaluation of the transport sector’s administration has been carried out, for example, by the Finnish National Audit Office (2013) and the Finnish Ministry of Transport and Communications (2013). These evaluations focused on the roles and responsibilities of the country’s different traffic safety actors’ internal efficiency. These evaluations did not cover the agencies’ various services, nor their overall performance, but did identify that the benefits and costs of agencies’ functions would need to be assessed.
Performance contracts between public agencies and their controlling bodies – ministries or their equivalent – are increasingly commonly used as tools for performance management (Binderkrantz et al. 2010). Management by Objectives (MBO) is a well-defined method of setting objectives to achieve the mission of an organization (Drucker 1974). The Finnish state applies MBO throughout the administration to operationalize its mission and specify the targets of each ministry, agency and state institution.

While there is a history in research in evaluating the performance of businesses, projects (Proost et al. 2014) and programs (Fuller et al. 2013), evaluating impacts of an entire agency and its socio-economic impacts is not commonplace and lacks tradition. This is the case especially with regard to service intensive agencies. Namely, it is important to acknowledge that public organizations dealing with investments, such as infrastructure maintenance or construction procurement, have a long and established tradition of submitting investment options to cost-benefit appraisal (Gramlich 1998; Layard and Glaister 1994; Brent 1996). Hence, the term ‘lack of tradition’ here specifically refers to the regulatory and administrative functions and services of public organizations, and not to the practices within investment decision justification.

When reviewing the relevant literature on service oriented public organizations, potential references on headline level are available (e.g. European Environment Agency 2013) but these evaluations have not provided quantitative (monetary or otherwise) results. Advances have been made and are underway (e.g. Saetren 2014) but the need for quantitative methods and metrics has been recognized by numerous initiatives and scholars (Papadimitriou and Yannis 2014; Howlett et al. 2014; United Nations 2010; National Audit Office 2001). Furthermore, there is not enough understanding of either what form the benefits take or how they spill over to different beneficiaries and sectors. It is not self-evident that public services, such as permits to transport certain goods or licenses to educate young drivers, are equally beneficial (or harmful) for all. These are the research gaps this research work contributes to by presenting an approach to constructing an evaluation of an entire service oriented public agency. The research gaps are rather generic in terms of presenting an approach but the approach is demonstrated in the context of a safety agency.
Research environment

The majority of the research presented here as well as the data used revolves around the main study subject, or the main unit of analysis, namely, the Finnish Transport Safety Agency Trafi. This sub-chapter (research environment) summarizes the basic characteristics of the agency. The current form of the agency was created at the beginning of 2010, when all mode-specific transport agencies (road, rail, maritime, aviation) were merged into two new agency entities. Trafi was mandated to govern and manage transport safety and environment and reliability issues, whereas infrastructure ownership and management activities were mandated to the Finnish Transport Administration (FTA). Most of the administrative duties were allocated to the agency, including vehicle taxation implementation and the hosting of most of the transport-related registers and databases. New mandates and responsibilities were issued as a new law on transport sector governance, which was specified further with a decree that mainly includes definitions regarding the director general’s mandate and eligibility to make decisions (Mononen et al. 2014).

According to the official documentation (Finnish Ministry of Transport and Communications 2009), the long-term objective behind the restructuring of the administration was to support transport system-oriented governance and strategic planning. The evolution of the Finnish transportation administration is presented in Mononen et al. (2014). As part of the build-up process of Trafi and FTA, extensive restructuring process had taken place tracing back to mid-1990s. The mandate and role of each modal agency stayed relatively stable up until the 1990s. Since then, the transportation administration sector has been under constant change. Three distinctive phases or waves of transition can be witnessed, including (i) deregulating, commercializing and outsourcing activities that could stand alone outside the governmental ownership and the associated governance and control; (ii) separating administrative duties from infrastructure owner, manager and developer roles, while still keeping each transport mode within its silo; and (iii) a shift toward a transportation system perspective in an attempt to abolish the boundaries between transport modes. Similar developments in transportation administration restructuring have also taken place, for example, in Sweden (Swedish Transport Agency 2014; Swedish Transport Administration 2014) and in Denmark, where administrative and regulatory duties for all modes have been merged into the Danish Transport Authority, but the infrastructure managers still reside in modal silos (Danish Transport Authority 2014; Mononen et al. 2014).
The tasks and mandate of the agency are based on legislation, but it does not define specific operational guidelines and so many of the tasks and services are left for the agency to design and implement in collaboration with the overseeing ministry. The degrees of freedom in this respect are substantial. The legislated tasks include 1) oversee the safety and security of the transport system, 2) limit environmental harms related to transport, 3) ensure prerequisites of maritime and aviation sectors, 4) ensure that markets work efficiently in the transport sector, 5) enforce that the agency’s own regulations are obeyed, 6) ensure that drivers’ education and licensing are organized appropriately, 7) perform taxation, registering and information service duties, 8) grant licenses, permits and decisions within its mandate, 9) specify regulations, 10) commit to international agreements of technical nature, 11) represent Finland in international co-operation, and 12) ensure the functionality and reliability of the transport system in exceptional circumstances. The seventh task, acquiring tax revenues, is important for the treasury as all vehicle and register taxes flow via the agency. However, the agency acts only as a technical tool and the funds flow ultimately to the state treasurer. The agency’s tax collection duty is a legislated and practical arrangement originating from the agency’s predecessors being historically responsible for maintaining the national vehicle registries, that is, the data needed for the correct and appropriate collection of taxes and other fiscal revenues. Based on its extensive registers the agency is able to act as tax collector on behalf of the Ministry of Finance (MoF), and these revenues form a substantial share of MoF’s revenue accumulation. In 2013, road sector taxes alone amounted to approximately five billion euros, which is over 10% of the state’s total tax revenues (Association of Tax Payers 2014). The annual budget for the agency was 136 million euros for fiscal year 2013 (Trafi 2014). As the agency is a net-budgeted agency, it is allowed to cover a major share of its expenses from service revenues such as permits and licenses.

During the process of carrying out this research work, the agency comprised four main functional sections: Regulation and Development, Permits and Approvals, Oversight/Control, and Registers. These were the vertical lines of command under the Director General. The shared functions are divided into Strategy, Communications and Administration. Additionally, the board consists of four mode-specific directors (Mononen et al. 2014).

In addition to traffic safety, the environment is another cornerstone area of activity because as mentioned, the agency collects taxes from the transport sector (vehicle tax, registration tax and several other taxes and fees), and the Finish tax
regime is very much oriented towards carbon-based charging (see e.g. Leviäkangas and Hautala 2011 and Nygrén et al. 2012).

The third main task of the agency is reliability and resilience enhancement of the entire transport system. This area is still in its development phase, and precise objectives are still in their infancy.

1.2 Research problem, objective and scope

The main research problem addressed in this study deals with the public financial or other investment into public agencies and with how, what kind and how much impact referring to wellbeing and benefits is generated with those investments for the nation. More precisely, this problem relates to the deficiencies within the tradition and methodology to investigate the quality and quantity of these impacts. In a way, all public bodies can be seen as societal investments that are expected to produce sufficient quantities of benefits to the citizens, organizations, businesses, and the nation as a whole. Performance management largely builds upon the knowledge of when and where state monetary or other type of input into the operation of a public body yields overall (and preferably maximal) net positive socio-economic benefits. Net benefits emerge from generated impacts that exceed the input. The main research problem addressed in this study can therefore be formulated as follows:

“The current understanding and methods for deducing the overall performance of a public agency are insufficient. Access to sound and transparent socio-economic performance metrics is a prerequisite for efficient performance management. Further research is needed in order to improve the understanding of and the methods for identifying these metrics.”

The main objective of the research is to investigate whether it is possible to construct a research arrangement and a set of methods through which the above mentioned research problem can be adequately addressed.

The main scope of this research is to derive and present the construction of a feasible process for evaluating the socio-economic benefits of a service oriented public organization. This study presents the socio-economic impact evaluation process via a real-life public agency. The entire research process and its development is covered, including the data collection, the research design, the sampling of the services to undergo detailed assessment, and the sample’s impact analysis as well as scaling up the results to cover the agency as a whole.
Traditionally a research problem is addressed with one or more research questions. In this study, the process of addressing the research problem took the form of four research questions (RQ) such that the first three questions divide the research problem into sub-sections and the fourth research question corresponds directly to the main research problem. Three of the four RQs include sub-questions. The research questions are as follows (Table 1):

<table>
<thead>
<tr>
<th>RQ#</th>
<th>Research Questions (RQ)</th>
</tr>
</thead>
<tbody>
<tr>
<td>RQ1</td>
<td>What is the performance management architecture surrounding and including the agency? What is the operational environment and the national performance management system? How is the study subject’s performance managed as part of the national system? How effective and comprehensive is the study subject’s performance management?</td>
</tr>
<tr>
<td>RQ2</td>
<td>With which methods can socio-economic impacts of a service oriented public organization be investigated and quantified?</td>
</tr>
<tr>
<td>RQ3</td>
<td>What is the efficiency of individual public services? Which services or which type of services yield positive B/C ratios, and what are these ratios? Which services, or which type of services yield negative or neutral B/C ratios?</td>
</tr>
<tr>
<td>RQ4</td>
<td>What is the overall performance? Is it possible to construct a research process for deducing a B/C ratio for a public agency? What is the overall B/C ratio of the agency?</td>
</tr>
</tbody>
</table>

The main contribution of this research is methodological and focuses on all stages of the analysis, including the process for quantifying impacts; the analysis of the impacts; generalization; a discussion on the spill-over and distribution effects of the socio-economic impacts; and the results’ scientific implications for the academia along with the results’ practical implications with regard to the management and control of the agency – and alike.

### 1.3 Research design

In this research, a single case study approach was applied, with multiple levels of analysis. Instead of, for example, attempting to observe several cases or analysis units in a comparative manner, the single case approach is mainly justified based on the limited amount of previous research of this kind (Eisenhardt 1989). The main unit of analysis in this research was the Finnish Transport Safety Agency Trafi, which was studied via multiple levels of analysis. Flyvbjerg (2006) states that a
single case study can meaningfully and substantially contribute to scientific development. In addition to case specific results, a single case study can yield generalizable results and can be used for testing hypotheses. Single case studies are capable of producing systematic exemplars that on the whole strengthen the science in the field where they are applied (Flyvbjerg 2006).

This research entity can be characterized as an exploratory study. Stebbins (2008) states, that researchers explore when they have limited prior scientific knowledge on the specific phenomenon (or organization, group, process, activity, etc.) they wish to examine, yet they have reason to believe the research object contains elements that are worth discovering. To be effective in this exploration, the researcher must approach it with innovativeness and open-mindedness in searching out data (Stebbins 2008). This research is built on a premise according to which the performance and socio-economic benefits of a service oriented public organization can be explored and explained via a thorough impact evaluation research exercise.

The research was designed around the above described canvas in the following way. The phenomenon under investigation comprised the performance and socio-economic benefits of a service oriented public organization. The means to investigate the phenomenon entailed the unit of analysis, that is, the agency which was investigated by thoroughly observing the agency’s services (i.e., units of observation) and drawing conclusions from these observations. The research design is depicted in Figure 1.

The reasons for selecting the mentioned unit of analysis related to both the scientific needs and the practical feasibility of the research. The phenomenon under investigation is significant, and the earlier research and literature on the phenomenon is limited, or at least sporadic. The main unit of analysis represents at the same time an extremity and a model example within the global sector of transport administration. The unit of analysis is a highly visible and transparent agency within a developed modern European society. Among the world’s many nations, Finland can in many ways be considered a model example in terms of, for example, the transparency, robustness and soundness with which it conducts its national administration (World Economic Forum 2014). The unit of analysis was ready and willing to have its performance analyzed thoroughly, transparently and publicly, in order to learn and develop. Therefore the selected unit of analysis provides a valid example for this research.
The analyzed data cover a wide array of documents comprehensively describing the agency and its tasks, objectives and performance indicators, as described in detail in sub-chapter 1.5. Furthermore, all available documentation and statistics with regard to planning, production and delivery of the agency’s services were utilized. The impact analysis of individual services within the sample relied on previously published research.

This dissertation contains four separate original publications. An overview of these articles is presented in Table 2. The column labelled ‘RQ’ indicates the most focal research questions for each original publication.
Table 2. Original publications, overview.

<table>
<thead>
<tr>
<th>Article</th>
<th>RQ</th>
<th>Article title</th>
<th>Publication; Publisher</th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td>RQ1</td>
<td>Transport Safety Agency’s Success Indicators – How Well Does a Performance Management System Perform?</td>
<td>Transport Policy; Elsevier</td>
</tr>
<tr>
<td>II</td>
<td>RQ2</td>
<td>Decision Matrix for Prioritizing Services – First Steps towards Full-Scale Impact Analysis of a Public Agency</td>
<td>International Journal of Public Sector Performance Management; Inderscience</td>
</tr>
<tr>
<td>III</td>
<td>RQ2,3</td>
<td>Socio-Economic Evaluation of a Transport Agency – Validating Accountability and Benefits</td>
<td>World Road Congress 2015, proceedings; World Road Association</td>
</tr>
<tr>
<td>IV</td>
<td>RQ2,3,4</td>
<td>From internal efficiency to societal benefits – multi modal transport safety agency’s socio-economic impact analysis.</td>
<td>Research in Transportation Economics; Elsevier</td>
</tr>
</tbody>
</table>

Article I contains a qualitative analysis of the agency’s mission, strategy, institutional architecture, and operational architecture as well as its steering and performance system. Additionally, the goodness-of-fit and efficiency of the performance management system is analyzed within article I. Article II comprises the service architecture analysis, the service cataloguing and classification, the evaluation criteria design and the compilation of the decision matrix. Article III includes the sample definition with Multiple Criteria Decision Analysis and studies and provides examples of the logical constructs for impact creation mechanisms as well as the quantification of the impacts and presents examples of the cost-benefit analysis. Article III also discusses the impact distribution and spill-over effects. Finally, article IV examines the validity of the sample selection and impact evaluations, encloses the results’ scale-up process, and presents the socio-economic balance sheet of the agency.

Figure 2 illustrates the operationalization of the research design, that is, the relations between the original publications, this dissertation synthesis and the research questions. In other words, Figure 2 illustrates which documents contain the responses to each research question. In the figure, the arrows with intact lines indicate major contributions, and the arrows with dotted lines indicate minor or indirect contributions to each research question or to addressing the research problem.
1.4 Research process, methodology and dissertation structure

The research process was largely built on a simplistic assumption of a mechanism of how a public agency is supposed to create benefits with its service provision. As input, the agency receives financial resources from the state budget and also from service revenues. With these resources the agency produces a variety of services for the society as direct but still intermediate outputs. The services in turn are designed to induce (positive) impacts, and if they do so, the impacts yield socio-economic benefits as final outputs. The socio-economic efficiency of the service production can be evaluated via benefit to cost analysis by comparing the monetary value of the input to the monetary value of the output. If the output exceeds the input, the operation is considered net-beneficial. This is an over-simplification since, as will be shown later, there may be impacts that can be identified as positive or negative but cannot be quantified.

To be able to evaluate the efficiency of the services and, in the end, to assess the efficiency of the agency as a whole, the research needed to investigate all of the abovementioned elements. Figure 3 illustrates this chain.

Fig. 3. The simplified chain from inputs to evaluation of efficiency.
The right-hand part of Figure 3 – “evaluation of efficiency” – in fact contains the main bulk of the research process of this work. Next, this process is described in detail and later summarized in Figure 4.

Importantly, the central and leading principle applied in this research, and especially in the impact evaluations for individual services was to focus only on impacts induced solely by the public service under investigation. This was imperative, since in addition to the agency’s contribution, in reality the entire operational environment including the impact generation is influenced (i.e., supported, enhanced, diluted or undermined) by many externalities. To name a few, these externalities can be changes in the economy, development of technology and/or the actions of other actors, typically. Therefore the impacts, and also the performance of the entire sector, are dictated by a complex set of factors, and seldom if ever by the actions of one single actor. In this research the guiding principle was at all times to evade the bias and distortion caused by the externalities and the changes in the surrounding operational environment and (while noting the externalities where they were found) focus on the sovereign impacts of the agency.

The research data, which are described in full detail below in sub-chapter 1.5, were used in and analyzed through a variety of methods, which are described next.

A full list of the agency’s service portfolio was compiled systematically. The list was valuated with regard to 20 different impact criteria using an interactive workshop method (e.g. European Commission 2010). Important aspects to be aware of throughout this process were (i) to keep the size of the list manageable, and at the same time (ii) make sure that in light of the needs of the subsequent research phases the cost accountability (or financial allocation) of the committed resources to these services is not lost at any point along the way. The term ‘cost accountability’ here refers to the fact that later on in the research process there was a need to pinpoint the exact monetary, human and temporal resources that are committed to and spent in producing each evaluated service, together with data on client payments returning as income back to the agency upon the delivery of that service. Some services provided by this agency to the general public or other stakeholders are free of charge, but notably others are not – there is a unit price tag on many of the services. Combining points (i) and (ii) above led to the need to maintain an equilibrium between the aggregation level of the list and the cost accountability factor, because otherwise the cost-benefit evaluation later on in the research would have been deemed difficult or even impossible to perform.

The valuated data were used in the selection of the sample with Multiple Criteria Decision Analysis (MCDA) method ELECTRE III (e.g. Bernard 1973;
Clemen 1996; Belton and Stewart 2002; Jablonsky and Dlouhy 2003). In other words, due to the high number of services (149), it was not realistic nor meaningful to analyze all of them in great detail. Instead, a sample of the entire set was used, and later the general conclusions were deduced leaning on the findings within the sample analysis. MCDA was applied to cordon the services of the public agency and the services were clustered to formulate the agency’s service impact profile. The profiles were constructed using key performance indicators set for the agency under its performance management contract with the governing ministry. A selected sample based on MCDA were used as the cores around which the rest of the agency’s services that carried similar profiles were clustered. The core services’ impacts were evaluated individually, applying traditional cost-benefit analysis, on the basis of which the scaling-up of impacts was finally performed.

The sample validity was analyzed with service clustering and cluster impact profiles, which were tested for uniformity with Spearman’s rank correlation coefficients (e.g. Marascuilo and Serlin 1988), supported by visual inspection.

Several methods were applied to the constructs of impact mechanisms, impact evaluation, benefit to cost (B/C) ratio determination with cost benefit analysis (CBA) (e.g. Gramlich 1998; Layard and Glaister 1994; Brent 1996), and benefit quantification/monetizing. These are explained in detail alongside the relevant research phase description.

When applying CBA, the research at hand focused on net benefits and net costs. Effectively the type of CBA applied was ex post, since the evaluated services were running and operational during the analysis. Chronologically the applied CBA research set-up represents a cross-section. In other words, the research was not designed as longitudinal but instead a single point in time - a snap shot - was examined. Consequently, the impacts are observed and reported on an annual level, although they might keep reproducing during the following years as well. The selected cross-section represents the year 2013, that is the monetary values portray the present annual values of 2013 (Gramlich 1998) and thus no discounting rules were applicable with regard to the calculations. (See also the section “Definition and general theory of Cost Benefit Analysis”.) The reason for selecting this particular year for the cross-section was that 2013 was the most recent completely reported fiscal year upon commencing this research’s calculations with monetary values.

The validity of the results from sample services’ impact evaluations were analyzed with non-parametric tests (e.g. Marascuilo and Serlin 1988).
The sample results were scaled-up to reflect the agency as a whole by first looking at the public funder’s return on their investment and second, by analyzing conservative approximations of the observed benefit cost ratios of the evaluated sample services.

The entire impact evaluation process as constructed in this research is depicted in Figure 4. Although the presentation in Figure 4 implies full linearity, it only depicts the final version and the end result of the design – in reality the entire process and its construction were of a highly iterative nature.

Fig. 4. The impact evaluation process of the research.

The research questions (Table 1) and the related original publications, that is, the articles (Table 2), were at the core of operationalizing and implementing the research and evaluation process of this study and of aligning it pragmatically with what is presented in Figure 4. Figure 5 illustrates the structure of this research.
Fig. 5. Structure of the research and positioning of original publications.

The following describes how this dissertation manuscript is organized. The convention adopted in referring and pointing to different parts of this manuscript is that the main levels, that is, the first level parts (e.g., 1) are referred to as ‘chapters’. The second level parts (e.g., 2.2) are referred to as sub-chapters, and the third levels (e.g., 3.3.3) are referred to as sections. Parts under non-numbered headings are also referred to as sections, with the associated heading quoted and written in italics (e.g. section ‘Policy and its design’). Table 3 below summarizes the chapters, their main contents and how they link to the succeeding parts of the dissertation.
<table>
<thead>
<tr>
<th>Chapter</th>
<th>Content</th>
<th>Link to the following chapter</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 Introduction</td>
<td>Motivation to this research, challenges that need to be addressed, a brief description of the agency, presenting the objective and scope of the study, presenting the research questions in conjunction with the identified research problem from which the research questions were derived, description of research design, process, methodology, dissertation structure, explaining the data that was utilized</td>
<td>Sets the background and reasoning for which theoretical discussion areas should be covered so that they are relevant in tackling the research problem.</td>
</tr>
<tr>
<td>2 Theoretical foundation</td>
<td>Covers a literature review for the three main theoretical discussion areas of (i) public administration and public policy; (ii) new public management and performance management in public domain, and (iii) impact evaluation of organizations, including the concept of cost benefit analysis. Summarizes the theoretical foundation and reflects it against the research problem and research questions.</td>
<td>Defines the theoretical canvas based on which the research contribution needed to be realized, mainly within the research work carried out and reported in the original publications.</td>
</tr>
<tr>
<td>3 Research contribution</td>
<td>Summarizes the work that is reported in the four original publications and details the original publication's contribution to the research questions. Synthesizes the contribution and provides the responses to research questions and to the research problem.</td>
<td>Research contribution constitutes the backbone to the following discussion and to deducing both the theoretical and the practical implications of this research.</td>
</tr>
<tr>
<td>4 Discussion</td>
<td>Extends on the observations presented in the preceding chapter. Details the theoretical and the practical implications of this research. Addresses research reliability and validity issues, as well as the limitations and transferability of this study.</td>
<td>Provides a comprehensive view to the research contribution and its implications and limitation, based on which a summary of the value and usefulness of the study can be drawn.</td>
</tr>
</tbody>
</table>
1.5 Research data collection

The research data and material consisted of scientific literature, published official documents and internal documents extracted from the agency’s enterprise resource planning system and financial management system, at their courtesy. The data cover a wide array of documents comprehensively describing the agency and its tasks, objectives and performance indicators. Furthermore, all available documentation and statistics with regard to the planning, production and delivery of the agency’s services were utilized.

The public data included laws, degrees, statutes, material available at the agency’s, ministry’s and other public online sources, scientific literature and scientific publications. The internal (i.e. not public) material included the documentation that was provided by the agency or that was retrieved from the agency’s enterprise resource planning (ERP) and management systems. In the following, the data and its collection are described in more detail.

The study and description of the functional architecture and performance management system utilized publicly available data. The data included the legislation, annual reports (budget, income and revenues); performance management contracts (and all relevant related supporting documentation), literature on the Finnish transport administration and general administration history, international regulation documentation (EU, IMO, ICAO, etc.); and statistics on the operational environment and its development/change trends (traffic safety, traffic volumes, energy consumption, greenhouse gas (GHG) and other emissions, etc.).

Service related data, that is, the general knowledge of the agency as a whole and its service and function repertoire, was collected next from both public sources and from the agency’s internal sources. A conclusive list of the agency’s services and functions was accumulated through a multi-phase process. The data collection process comprised an interview round in January 2014 involving experts and decision makers from Trafi and the Ministry of Transport and Communications, project steering group meetings and other correspondence.
In addition to the publicly available data, the service description and analysis phase of this research utilized data provided by the agency. These included the architecture descriptions of the agency’s core processes, service descriptions, practical working guidance documentation for the staff and various statistics on the production of services. The statistics included those of service production volumes, of databases on anomalies or deviations in service production, and statistics and data on discovered discrepancies in audits, inspections and control checks.

The resulting conclusive list or database of the agency’s services and functions constituted by far the most focal data entity used in the various research and analysis phases throughout the entire duration of this research. The database includes 146 services (or records) and covers the operation of the agency in its entirety. Each record consists of several data types and meta data, such as the service name, application regime, relevant transport mode(s), service type, source of financial transcription, service category, service description, service production volume, and where applicable, the financial data of the service (such as the allocated labor, other allocated resources, revenue from service charges, etc.).

Throughout the research progress, the database was constantly supplemented and replenished with various data types, including most importantly the service profile data where each of the services had been evaluated against 20 criteria. The database, and especially the service profile evaluations formed the basis for the decision matrix (DM). The evaluation data collection process is described in the next paragraph.

First, a pre-evaluation of the data records was performed, using a set scale. This in practice meant pre-filling most of the cell values based on all of the available information and the researcher’s expertise. The main reasoning behind having the data pre-filled was to make the data collection workshop much more manageable both in size and in duration. In cases where the researcher was even slightly unsure either of what the service content actually was or what (if any) cause-consequence mechanisms could exist in relation to any particular criteria, the cell was automatically and strictly left blank. A large proportion of the cells was given a definite null value. Namely, in most cases a service has been designed and directed to tackle a certain aspect of the agency’s main responsibilities (or policy targets) - and not all of them. Next, all pre-valuated data was completed, corrected and validated in two phases.

A workshop was arranged to tackle the main bulk of this workload and additionally, some elements of the data went through another checking, evaluation and scrutiny round within the agency. Fourteen Trafi employees participated the
workshop, and additionally two research group members were present in supporting roles. Supporting role here means that the researchers provided the guidance for the working methods, distributed the instructions and material, and operated as moderators and facilitators during the workshop – but did not take part in any way in actually valuating the services. The backgrounds of the participating agency personnel were intentionally selected so that all relevant aspects would be covered as comprehensively as possible. All transport modes were represented; the group included experts from service provision layer to top-level management; and experts for all of the agency’s areas of responsibilities were there (e.g., transport safety, environment, etc.). The workshop representation for both aviation-related service expertise and expertise on services covering more than one transport mode turned out to be slightly inadequate however - thus a secondary off-line scrutiny round was introduced. One service group was observed to include several strategically important functions (such as the international cooperation activities, functions dealing with issuing norms and regulations or supporting national legislation development), which deserved to be separately evaluated by representation from each transport mode.

The production costs of services were determined together with the agency’s financial department personnel, in cooperation with the mode specific experts and directors. The source for the cost data was the agency’s financial management system (FMS). Specifically, the selected cross-section represents financial year 2013, including the agency’s costs, income, service architecture, and organization and management structure. The monetary values portray the present annual values of 2013. The costs include the labor (i.e., person years and the associated direct and indirect salary costs) allocated to the production of the service or function, and where applicable, other related fixed or variable costs such as materials and consumables, travel, vehicle leases, etc. Additionally, the overhead costs from general administration and support functions of the agency were allocated (seeded) as a part of the service production costs. The seeding of overheads was based on the agency’s internal ERP and FMS, and their calculation coefficients, rules and principles.

The evaluation criteria definition, service impact evaluation, and the final stages of efficiency analysis were all based on general publicly available literature, such as scientific literature and articles along with reports, guidance, and studies published by global entities such as WHO, OECD, WEF, UN, ICAO, IMO and IATA.

Table 4 summarizes the data utilized in this research.
Table 4. Research tasks and the associated data.

<table>
<thead>
<tr>
<th>Research task</th>
<th>Data</th>
<th>Source</th>
<th>Public (P), Restricted (R)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Functional architecture</td>
<td>Legislation, literature on Finnish transport administration and its history, international regulation documentation, statistics on the operational environment</td>
<td>On-line &amp; off-line public sources, (MoTC, Trafi, EU, IMO, ICAO, IATA)</td>
<td>P</td>
</tr>
<tr>
<td>Performance management system</td>
<td>Performance management contracts (+ the associated documentation), legislated tasks, annual reports</td>
<td>On-line &amp; off-line public sources, MoTC, Trafi</td>
<td>P</td>
</tr>
<tr>
<td>Service description</td>
<td>Core process architecture descriptions, service descriptions, working guidance documentation,</td>
<td>On-line &amp; off-line public sources, ERP</td>
<td>P &amp; R</td>
</tr>
<tr>
<td>Evaluation criteria definition</td>
<td>Scientific literature, reports, guidance and studies</td>
<td>On-line &amp; off-line public sources, scholarly databases</td>
<td>P</td>
</tr>
<tr>
<td>Decision matrix pre-evaluation</td>
<td>Pre-evaluation scores</td>
<td>Researchers’ expertise</td>
<td>N/A</td>
</tr>
<tr>
<td>Decision matrix evaluation</td>
<td>Evaluation scores</td>
<td>Workshop participants</td>
<td>N/A</td>
</tr>
<tr>
<td>Costs of services</td>
<td>Annual reports, service production costs, service revenues</td>
<td>On-line public sources, ERP, FMS</td>
<td>P &amp; R</td>
</tr>
<tr>
<td>Service analysis</td>
<td>Service descriptions, service production volumes, anomalies &amp; deviations in service production, discovered discrepancies in audits, inspections &amp; control checks</td>
<td>On-line &amp; off-line public sources, ERP, scholarly databases</td>
<td>P &amp; R</td>
</tr>
<tr>
<td>Service impact evaluation</td>
<td>Scientific research on service impacts</td>
<td>Scholarly databases</td>
<td>P</td>
</tr>
<tr>
<td>Efficiency and performance analysis</td>
<td>Scientific literature on efficiency, performance and research methods</td>
<td>On-line &amp; off-line public sources, scholarly databases</td>
<td>P</td>
</tr>
</tbody>
</table>
2 Theoretical foundation

2.1 Theoretical framework

This research builds upon three main theoretical discussions; (i) public administration and public policy; (ii) new public management and performance management in the public domain, and (iii) impact evaluation of organizations, including the concept of cost benefit analysis. Figure 6 illustrates the relevant theoretical framework. In the figure, such theoretical discussion topics and/or areas that are not covered within this literature review (due to their minor relevance to the scope of this research work) are presented in smaller italic font.

Fig. 6. Theoretical framework of the research.

In the following sub-chapters, each of these main theoretical areas are discussed individually, followed by a literature review synthesis.
2.2 Public administration and public policy

This study focuses on research within the public domain and more precisely, on organizations within it. These organizations as a whole constitute ‘public administration’, which includes all agencies, ministries, administrations and other publicly funded and governed entities. The term ‘public’ refers to entities that have a clear public purpose as opposed to what is generally considered to be business or commerce. According to the United Nations Economic and Social council (2006), public administration has two meanings that are closely related: (i) the aggregate machinery funded by the state and in charge of the management of the executive government affairs together with its interaction with other stakeholders in the society and external environment, and (ii) the management and implementation of the activities in deploying governmental laws, regulations and decisions and the provision of public services.

As such, the theoretical discussion within the sub area of public administration and public policy has a wide basis, containing literature that could be labelled as ‘purely public administration’ or ‘purely policy’, but additionally bearing highly relevant connections to the theories within, for example, economics, law, political science, sociology, philosophy and psychology (Pierre 2003; Frederickson et al. 2011). Therefore, any one stand-alone theory in fact does not exist and the challenge in covering the area is not related to a lack of theoretical discussion but more to the surfeit of. Regardless, it is appropriate to cover public administration and public policy and the related theories to an extent, since they as a matter of fact constitute the backdrop for the research at hand. In the following, this theoretical discussion area is covered with an examination of administrative history and reform and of the design, implementation and analysis of public policy.

In addition, the theoretical discussion area of public administration and public policy includes issues such as sociology, organizational theory, institutions, political economy, legislation, public budgeting, political ethics, political entrepreneurship (lobbying), human resource management, etc. – but these have been largely cordoned out from this literature review due to their lesser relevance to the scope of this research work as well as to the abovementioned surfeit of theoretical discussion areas in and around this domain. (Political entrepreneurship is briefly covered though, as one of the potential factors in policy design.)
2.2.1 Public administration: reform and research

This sub-sub-chapter takes a concise look into how and why the public administration domain and the research around it have come to be what they are today.

Public organizational and managerial practices are as old as civilization. For example, the transition from feudal societies to modern nation-states, the characteristics of the colonial era and the developments in leading imperial China were all driven by centralization or decentralization of policy or other developments in the forms of administration (Fredericson et al. 2011). The latter two of the presented examples (colonial era and imperial China) show that public administration does not relate only to the concept of democracy.

The conscious examination of management principles and other aspects of public administration began in the US during the late 19th century when the administration in general was in a bad reputation. The administration was perceived as incompetent, corrupt and inefficient. The policy making and administration was also sporadic and characterised by discontinuity, mainly for incumbency reasons since after each election the winning party was practically free to reassign new persons to all administrative positions (Gruening 2001). Therefore the administrators changed frequently and their nominations were prone to be affected by personal and political ties and corruption rather than being based on proven competencies. The dissatisfaction of the people was thus the primary driver in moving towards the development of new ways to govern the public sector.

The definition of the age of the theories may vary by scholars, but formal study of public administration began in the US during the late 19th century when the administration in general was in a bad reputation. The administration was perceived as incompetent, corrupt and inefficient. The policy making and administration was also sporadic and characterised by discontinuity, mainly for incumbency reasons since after each election the winning party was practically free to reassign new persons to all administrative positions (Gruening 2001). Therefore the administrators changed frequently and their nominations were prone to be affected by personal and political ties and corruption rather than being based on proven competencies. The dissatisfaction of the people was thus the primary driver in moving towards the development of new ways to govern the public sector.

The definition of the age of the theories may vary by scholars, but formal study of public administration can be dated all the way back to an essay by Woodrow Wilson (1887) called “The Study of Administration”. There he argued that the study of public administration should be separated from political sciences into its own field of study or science, mainly because administrators are responsible and held accountable (see next paragraph) to political leaders, who in turn are accountable towards the citizens. Politics relate more to the principles of justice, democracy and equality, whereas administration searches for efficiency, as interpreted by Yeboah-Assiamah et al. (2015).

The term ‘accountability’ deserves a short specific discussion and definition here since it is of particular interest in this research. Within the public administration, public entities are held ‘accountable’ for their actions on many levels. These levels include democratic accountability which refers to the relationship between political leadership and citizens, and administrative
accountability referring to the relationship between administration and political leaders (Byrkjeflot et al. 2012). This research is mainly constrained to observing the vertical accountability as defined by Schillemans (2011) and, more precisely to the management dimension of accountability with regard to economic ‘bottom line’ and organizational performance targets, as constructed by Byrkjeflot et al. (2012). Also elements of service quality accountability, i.e. operational quality standards, are relevant to this research, to an extent (Byrkjeflot et al. 2012).

Weber (1952; Max Weber, a well-known German sociologist and political economist from late 19th to early 20th century) defined the elemental features of public administration including e.g. (i) formal authority that demands obedience; (ii) laws and statutes that concern all citizens, (iii) division of competences within the administration into differentiated tasks, expertise and professions; (iv) hierarchical coordination; (v) mechanisms to ensure some level of continuity, and (vi) the development of particular organizational practises. Douglas and Wildavsky (1982) claimed that these elemental features of public administration are in fact mandatory prerequisites for civilization in general.

Weber’s definitions of public administration aim at constructing an ideal of bureaucracy. According to Weber’s (1952) view, a fully developed bureaucratic organization is technically superior to any other form of organization. In today’s world the word bureaucracy seems also to carry some inherent negativity within it and is occasionally considered a synonym for the term ‘red tape’. It has been known to be used to describe inefficient, overly complex, unintelligible or intentionally non-transparent ways to handle public administration or other disciplines. The conceptions and misconceptions with regard to bureaucracy are also discussed in the literature (e.g. Yeboah-Assiamah et al. 2015; Hull 2012; Page and Jenkins 2011).

Since then public administrations have been more or less currently in reform or at least being both scrutinized and developed. The different models or paradigms adopted over the decades include the classical public administration, neoclassical public administration and adhering to the public choice theory (Gruening 2001). The most recent developments have related to the neoliberal thinking and new public management (NPM), which will be covered separately in the next theoretical sub-area. On the other hand, Gruening (2001) questions whether a single generally accepted paradigm for public administration model has ever really existed, or ever will, due to the volatility and constant development of the realm. This assertion is – at least to some extent – supported by the OECD (2010) view that there is in fact a need for continuous change in how the public domain is being managed.
Change and transition from one form of administration model to another one needs to originate from within a nation, as Featherstone (2014) observed when investigating an attempt to induce domestic reform via external conditionality. His research on EU’s intervention into the Greek debt crisis showed that trying to impose a shift towards e.g. more effective state performance management may face critical opposition and the success may therefore be modest. The external agendas may be perceived as alien and may induce national feelings of loss of own legitimacy and appearing weak in the face of foreign influence. This influence may even be labelled as foreign ‘blackmail’ (Featherstone 2014).

Even more recently, the refugee crisis and Brexit (the outcome of the United Kingdom’s referendum on leaving the European Union) have reincarnated some lively scientific discussion on many issues, including political reform and trust and confidence in the current political system. These issues are likely to induce some public administration reform within Europe in the near-future. De Angelis (2017) observes that Brexit has brought with it a need for the member states to continuously justify EU membership since it has raised some skepticism towards European integration. Brexit in itself could be seen as a reform and the factors that led to it transpiring can shed some light onto the drivers behind reforms taking place. Shaw et al. (2017) investigated the data used by the competing Brexit campaigns and presented as justification for their cause and themes. They found that the campaign with consistent and highly detailed data connected to a smaller set of campaign themes paved the road to the winning vote, as opposed to the broader range of themes that the losing group campaigned with (Shaw et al. 2017).

2.2.2 Public policy: design, implementation and analysis

Policy and its design

Policy design (or policy making) can be supported with various methods. ‘Governometrics’ applies quantitative and statistical methods to aid policy making (Sharna et al. 2014). Moreno-Jiménez et al. (2014) present a model and a method for the design and selection of public policies based on a cognitive democratic model called e-Cognocracy, that utilizes ICT and allows for co-decision making between citizens and politicians. Razu and Tagai (2012) study the impact of public policies and exogenous variables with regard to optimal vehicle design on greenhouse gas (GHG) emissions. Their optimization model facilitates finding the
optimum vehicle design from which the government can choose a public policy accordingly to produce a desired effect (Razu and Tagai 2012).

Policy design, by default, will always include some level of individual creativity and intuition, as Considine (2012) points out in his highly interesting work. He also links the more general design models to public policy design, thus presenting an approach that deviates from the common concerns of policy design failure and the related backdrop of constraints and structural imperatives. He asserts that by joining the public policy makers’ hard core expertise with their capacities of creating imaginative solutions opens up avenues in public policy design that include elements such as pattern recognition, anticipation, emotion engagement, even fabulation and playfulness, while still containing a level of risk protection (Considine 2012). The intuition element of decision making processes within the public domain is discussed to some depth in some of the original articles and in the discussion chapter of this dissertation. It is safe to assume that all direct and exogenous variables within as complex systems as governments and public administrations are, all quantitative and analytical evidence of policy decision impacts is never going to be available.

The direction and content of public policy can be affected by practically countless number of factors. In the following only some examples are provided of the variety of interdependencies discovered in earlier research. According to Hooghe and Oser (2015) the density of unions and social expenditure have a direct causal relationship and positive correlation, as pointed out in their research comparing OECD countries in this respect. On the other hand, the adopted levels of freedom in the state market economy (liberal or coordinated) play a role in how strong the effect of this relationship is (Hooghe and Oser 2015).

In addition to the national steering, policy design may often be influenced by supranational organizations or bureaucracies. In the case of EU member states, this organization first and foremost would be the European Commission (EC) and the European Parliament (EP). As Husted and Seyfries (2015) point out, the effect of the EC steering is a mixture of political influence and bureaucratic execution having a substantial weight on the latter, thus contributing to the functional efficiency of policy making and relieving the pressures from the political level. This should not be interpreted as a solely positive or an ideal arrangement, though: the more the decisions with regard to e.g. legislation are in the hands of bureaucracies, the more complex (i.e. less direct, slower, less transparent) the path from the peoples’ expressed will to the steering and adjustment of the direction of those decisions is. In addition to the EC and EP, other influential constellations
more or less exempt from the direct national rule include various commitments to international treaties and organisations, such as for example OECD, UN, WEF, etc. concerning general issues; and closer to the subject matter of this research, IMO, ICAO, IATA, etc., concerning the transportation sector.

One mechanism or factor in policy design sometimes can be the so called policy transfer, i.e. learning about policies and policy design best practises in other areas. According to Marsden and Stead (2011) policy transfer may lead to more sustainable policies.

Policy making can also be influenced by so-called policy entrepreneurs, referring to actors who use their knowledge of the process to advance their own policy ends (Caimey 2011). George Stigler (1971) defined the term ‘regulatory capture’, which is kin to political entrepreneurship. Therein, the claim was that any industry or group that gains enough political leverage, will inevitably try to use that leverage to their advantage (Stigler 1971). The commonly used term for this type of activity is lobbying. Entrepreneurs may be current or former politicians, members of interest groups or unofficial spokespersons for particular causes, such as supporting business or promoting particular societal or other goals. As Caimey (2011) states, entrepreneurs have the knowledge, power, tenacity and/or luck to be able to exploit windows of opportunity within policy problems to convey their solutions to policymakers. These solutions may be, for example, ideas or services that upon adoption further the entrepreneurs’ goals while at the same time solve recognized policy challenges.

Policy implementation, analysis and evaluation

The set and designed policy needs to be put into effect and implemented. Typically this is done by setting the legislation and building the institutions that control and convey the legislative framework into action.

Policy implementation theory is focused on conceptualizing and understanding by whom, how and why policy is put into effect (Schofield 2001). According to Hjern (1982) implementation research is the link between the political and economic analysis of policy and the institutional analysis of public administration. Analyses of policy implementation distinguish between the state’s policy promises and the state’s policy products (Brodkin 1990). There are three overarching themes within implementation research studies, namely the attempts to develop analytic models, the different approaches to studying public policy implementation (e.g.,
‘top-down’ or ‘bottom-up’ approaches), and trying to identify key implementation variables (Ryan 1995; Matland 1995).

According to Schofield (2001) the four key justifications for implementation research are (i) determining policy success and failure factors, (ii) increasing the predictability of policy results, (iii) making recommendations for policy design and (iv) unifying the approaches within multi-actor and inter-organizational operation research in administration and politics.

When searching for literature with regard to policy implementation, it soon becomes obvious that the heyday of that particular research was from the mid 1970’s to the mid 1990’s and since that period, direct implementation research has been less active. It looks as if the policy analysts’ research focus shifted more toward the study of other issues, such as institutional and organizational change within the public sector or policy analysis and evaluation.

Policy analysis can be aided with various kinds of tools such as analytical methods geared towards identifying the determinants of policy outcomes and simulation modelling methods based on systems science (Atkinson et al. 2015).

2.3 New public management and performance management regimes in public domain

Before entering the actual literature review on this theoretical discussion area (starting from 2.3.1 below), a short preface on the subject is presented here. Performance management is one of the key theoretical areas of this research. The main reason for this is the wide emergence of New Public Management (NPM) concept over the past few decades, and more importantly, its practical application into national public administration contexts in Finland, Europe and elsewhere. The main underlying assumption in NPM is that in the name of efficiency public entities should be steered and managed in a manner somewhat similar to the way businesses have been and are being run. The entire overarching topic of this research revolves largely around enhancing decision support for managing service oriented public bodies when NPM principles are being applied. This sub-chapter covers the theory and history of these principles and also – wherever meaningful – draws parallel comparisons between NPM and the practices applied within the private sector. This comparison is made to highlight both the potential contradictions and also the synergies between the two realms. This sub-chapter, however, omits a description of the theoretical discussions with regard to performance management and measurement systems/techniques that are generally applied within the private
sector. Just to name a few, these include EVITA, total cycle time, the Prism process, performance measurement questionnaires, IPDMS, ECOGRAI, the Fraunhofer approach (Bourne *et al.* 2003) and balanced scorecards (Haapasalo *et al.* 2006).

The first section (2.3.1) discusses the management of public bodies in general and NPM in particular. The remaining sections (2.3.2 and 2.3.3) focus on a discussion of the general NPM ambition of making the effectiveness of public sector transparently measureable. The progress of the discussion follows a down-stream path, starting from actually measuring (i.e. quantifying) the effectiveness all the way down to the strategies and missions that – more or less directly – define what ends up being measured. In NPM, effectiveness and efficiency are measured against a finite set of metrics and indicators. These indicators in turn, are (or at least logically should be) derived from the objectives and goals of the public entity being managed. And finally, the objectives of the public entities are dictated by the mission of the government (or the relevant sector of government) and by the strategies that have been selected to achieve the mission. In an attempt to illustrate, Figure 7 presents a simplistic view of the flow from public policy to performance management and vice versa. The upstream direction implies a connection where the evidence of what can actually be achieved with public activities helps to steer all upstream elements – at least in an ideal case.
2.3.1 New Public Management

Background

The birth of new public management (NPM) was connected to the emergence of the neoliberal ideology that is often associated to the governments of Thatcher (UK) and Reagan (USA) in the 1980s. Neoliberalism contained new ideas of generally aiming at decreasing the states’ role and responsibility in guaranteeing the welfare of the people. The role was to be decreased by means of deregulation, open markets, decentralization and privatization. Governments that have adopted the neoliberal thinking as the guiding principle in managing the public sector are commonly also referred to be applying new public management (Bessant et al. 2015). The management principles in NPM have been derived from the principles applied in the commercial for-profit sector (of businesses and companies) (Hood 1991, Hood and Peters 2004; Bessant et al. 2015).
The predecessors of NPM as the public management paradigm included classical public administration, neoclassical public administration and adhering to the public choice theory (Gruening 2001).

Hood (1991) recognizes several megatrends that had contributed towards the rise of NPM, namely: need to reverse (or slow down) the growth of government; the shift towards privatization of previously public institutions; emergence of information technology and globalization of policy making. These have been occurring in tandem with similar activities for government owned assets.

**Drivers, objectives and doctrinal content**

The emergence and adaptation of NPM has mainly been driven by the public agencies and their services having been under great scrutiny. The trend of NPM has taken rapidly advancing steps across the globe, in countries such as Canada, New Zealand, Australia, USA, but also in Europe (Hague 2001, Gruening 2001). The ideological shift towards more market-oriented economy also with regard to formerly public services is often dated to the earlier mentioned Margaret Thatcher’s governments in the UK. The ideas of new models of public governance brought along the metrics of private sector management, meaning that the outputs and effectiveness of public services needed to be assessed properly in order to facilitate efficient management control. These directions were adopted also by the Finnish governments since the late 1980’s, when the first waves of privatization and commercialization of former administrative and public functions took place. Since then Finland’s governments have persistently followed this path, and restructuring of public services have been carried out rather systematically (Finnish Ministry of Justice 2009, Finnish Ministry of Finance 2012, Finnish National Audit Office 2013). The reforms or re-structuring efforts have brought with them many performance management tools and principles, the applicability and effectiveness of which is still under debate (see e.g. Maugeri and Metzker 2013). The transport sector has been in the forefront of the renewal processes of different administration sectors. The Finnish Transport Agency is a fairly new entity in the transport sector governance architecture. The historical evolution of Trafi to its present form has been described in (Mononen et al. 2014).

Whenever a government has decided to adopt neoliberalism and NPM, the underlying aim and expectation has been to increase the efficiency, effectiveness and excellence of the public sector (Deem et al. 2007). This is expected to be yielded by reducing expenditure and the unit cost of the provided public products
The ‘public products’ can materialize in various forms, such as services, infrastructure or other types of publicly provided commodities. The doctrinal content of NPM includes (i) professional management of the public sector; (ii) measures of performance being implemented; (iii) resource allocation and rewards being linked to measured performance; (iv) disaggregation of public sector units; (v) shift to increased competition within the public sector; (vi) management practices being adopted from the private sector; and (vii) more efficient use of resources (Hood 1991).

Many of these contents, such as aspirations towards heightened efficiency and professionalism, share commonalities with the so-called Weberian governance model, named after Max Weber (Cho et al. 2013; Woods 2003). Interestingly, some foundations for the doctrinal contents of NPM were in fact laid down already in the earlier mentioned essay by Woodrow Wilson (1887). For example, it is extremely intriguing to compare the above points (i-vii) to a statement within the essay over a century earlier:

“It is the object of administrative study to discover, first, what government can properly and successfully do, and, secondly, how it can do these proper things with the utmost possible efficiency and at the least possible cost either of money or of energy” (Wilson 1887, page 1)

The penetration level (geographically and otherwise)

The development leading to a wide adoption of NPM originated in the latter part of 1970s and early 1980s. The first adopters were the Thatcher’s government in the UK and a few municipalities in the US. Soon after the governments of Australia and New Zealand followed (Gruening 2001); and eventually most of the OECD nations in addition to other countries (OECD 1995). Although, here to be remembered firstly the earlier mentioned Gruening’s (2001) doubt with regard to actual paradigms having ever existed, including NPM. And secondly, Hood’s (1995) observation, that even though allegations of new global paradigm in public management existence have been made (in Hood’s article namely NPM), there has been in fact considerable national variations in how the principles were being interpreted and to what extent they then had been adopted.

Unlike Sweden, New Zealand, Canada and the UK, Finland was not within the very first wave of NPM adopters, but in the second wave right after that together with e.g. Italy, Austria, Denmark and the Netherlands (Hood 1995).
emergence of neoliberalism started in the 1980s evidenced e.g. with many privatization and outsourcing exercises concerning several governmental sectors. The transport sector was one of the earliest practical adopters of this approach in Finland.

2.3.2 Management by Objective

Strategy, mission and objectives

Governments that apply NPM generally manage the ministries and the agencies through objectives. This can be referred to as ‘Management by Objective’ or MBO, which is a method of setting objectives in order to achieve the mission of an organization (Drucker 1974). For example, the entire Finnish state governance model and management system is currently built on performance management with MBO. The objectives can be annual and mid-term targets that relate to tasks defined in laws and decrees but they can also be more flexible objectives, such as current policy aims or pressures to tackle emergent topical issues, problems and challenges with regard to e.g. operational efficiency, quality of customer service or feedback from international commitments, etc.

The underlying principle in applying MBO in NPM is that the political agendas, i.e. the mission and strategy, of each government are filtered down via the ministries to central agencies and other units as mid-term and short-term objectives that take steps to the desired direction (Gruening 2001; Barberis 1998). Short-term objectives can be annual or multi-annual, hence covering one or more fiscal budget years, while mid-term objectives can cover the several-year planning cycle (i.e. electoral term) for each government. The mid-term planning process is therefore essential in the practical realization of each government’s agenda.

Typically ministries prepare annual and mid-term agreements with its agencies where the resources and objectives are balanced through a negotiation process. In the end, the ministries and the ministers usually have the final say. The process is usually not only about domestic endogenous objectives. In addition, the international processes taking place (in e.g. the EU or other international organizations) may affect the performance management negotiations. For example, the EU’s white papers and the regulations issued by the IMO or ICAO may have their impacts on national performance objectives.
The overall process of obtaining the objectives can be relatively complex. This complexity is further described and discussed later in sub-chapter 3.1 via an example from the Finnish transportation sector.

**Metrics and indicators derived from the objectives**

NPM has brought along the performance metrics with which the outputs and effectiveness of public services are supposed to be measured in more tangible ratios in order to gain efficient management control. Ideally these managerial controls should be separated from the political control. That is, in the day-to-day or annual management. When observing a longer temporal perspective, it emerges that it has to be in fact the political choices that have either directly or indirectly dictated what the managerial controls have become to be, whatever they are.

Depending on the government sector and also the type and the set mission of the public entity being managed, the shape and form of the indicators and performance metrics vary. Hood (2007) recognized three categories for managing government and public services by numbers: (i) target systems i.e. measuring actual performance against aspirational standards expressed as threshold numbers; (ii) ranking systems, i.e. comparing to the performance of another service unit; and (iii) intelligence systems, e.g. various activity logs, often combined with targets and rankings.

Recurring metrics within these systems include annual budgets, customer satisfaction and number of staff. For health care services (e.g. hospitals) these can include mortality rates, discharges, readmission rates, number of staff, number of beds, etc. (Alonso et al. 2013; Santiago et al. 2015). For education services the metrics can include the number of staff, graduation statistics, average study duration, or how fast and where the graduates get employed (Santiago et al. 2015). For transportation authorities, metrics such as traffic safety development, emissions, vehicle inspection statistics or fail-rates, infrastructure maintenance quantities and construction outputs, etc. are typical (Finnish Ministry of Transport and Communications 2012 and 2014).

In some cases dual and therefore potentially conflicting roles may easily appear, as pointed out by Vasikainen (2014). Namely, sometimes the overseeing organization (e.g. ministry) may in practice be the de facto manager of an agency, and at the same time a key client for the agency’s outputs. Examples of this can be found in abundance. For example, ministries of transport may manage several agencies under their jurisdiction. But in order to achieve the ministries’ own
missions and objectives, they may have to rely on a plethora of services produced by those same agencies. These services could include production and maintenance of transport infrastructure, support for regulation work, providing assessments of transportation system status, implementing measures to improve transportation system resilience, etc.

2.3.3 Evaluating success or failure of New Public Management

Evaluation of success and performance within the NPM concept

In NPM, and in other management models as well, public entities are held accountable for their actions and performance in various ways. The ontology of ‘accountability’ and as it is understood in this research was discussed earlier. A special characteristic of NPM is that the different accountabilities, or at least a subset of them, need to be translated into measurable performance indicators. (These indicators or metrics, and typical examples thereof, were discussed in the preceding section 2.3.2.) Controlling the success in achieving the indicators’ target levels then yields a quantitative or qualitative measure of the organization’s performance. The practical control is typically connected to a scheduled review of a performance contract (Binderkrantz et al. 2011) or performance agreement between the controlling body (e.g. ministry) and the public entity being managed (e.g. agency).

Bindercranz and Cristensen (2009) examined the practices with regard to controlling several Danish government sectors with performance contracts during years 1995, 2000 and 2005. During that time period (from 1995 to 2005), the number of public agencies in Denmark had increased from 50 to 59, whereas the number of agencies with performance contracts had increased from 9 to 54. That means that during that period the practice of managing via performance contracts had evolved from its test phase to becoming the standard practice. Bindercranz and Cristensen (2009) conclude that the performance contracts had been developing to a more mature state but also had generally become much more complex, at least in terms of the number of metrics (or ‘performance demands’, as they call them) involved. The complexity had increased either because of a learning process or a layer-on-layer effect where new metrics are being constantly added without removing any of the old ones (Bindercranz and Cristensen 2009).

Pollitt (2006) investigated the performance management systems in four European countries (Finland, UK, Sweden and the Netherlands) and he concluded
that there had been a tendency of the performance indicator systems growing more sophisticated: At the same time the governing ministries were feeble in applying performance based steering, apart from them being adamant with regard to the requirements of the steered agencies having to stick within their set annual budgets (Pollitt 2006).

As mentioned, in Finland the MBO process permeates through the entire state administration. Recent steps have been taken nationally to improve the efficiency and effectiveness of the process, driven e.g. by the Finnish Ministry of Finance (2012). The fundamental recommendations have concerned the interaction between administration units as well as with surrounding stakeholders, such as non-government organizations, private sector, etc. However, the most relevant recommendations with regard to the focus of this research have concerned the needs to enhance the performance metrics to a more quantitative direction (Finnish Ministry of Finance 2012; Finnish National Audit Office 2012).

Academic debate on goodness and badness of fit of NPM in general

The goodness of fit and the excellence of neoliberalism and NPM have been and are under vigorous debate within science. The discussion has involved both champions for NPM and critics against the approach.

Gruening (2001) recognizes several undisputed characteristics of NPM but also some attributes that remain debatable in terms of their goodness of fit. He finds that positive development with regard to e.g. budget cuts, decentralization, competition, management flexibility, financial management and accounting, utilization of information technology and customer satisfaction, have been reported by most academic observers. On the same token he questions whether there is academic consensus in e.g. regulation, administrative structures, citizen participation or policy analysis having moved to the same positive direction (Gruening 2001).

As early as 1936, Robert Merton introduced the ‘Mertonian approach’, which in the essence claims that any change in a society aiming at positive development always generates unintended (negative or positive) effects to some other parts of the society as a side product (Merton 1936). Later this was re-formulated by Moore and Tumin (1949) into a claim or a ‘law’, that Merton was universally right about the generation of unintended effects, but only partially right about the potential desirability of those effects:
“There is no exception to the rule that every time a culture works out an empirically valid answer to a problem, it thereby generates a host of derivative problems” (Moore and Tumin 1949, pages 794-795)

Some researchers argue that private sector operating philosophy and tools are not always, nor in all respects, a perfect fit for public sector management (Christensen 2007; Romzek 2000; Hood 1991; Hood and Peters 2004). According to Romzek (2000), the rhetoric of reform may pose consideration of accountability in terms of whether government is more accountable after the reform than it was before. Romzek (2000) points out that although it is possible to discuss accountability in quantitative terms, doing so still implies a linear concept that may not always fit perfectly together with the intricacy and the complex context of public management. A recent case study analysis (Alonso et al. 2015) could not find scientific evidence on Spanish hospitals run according to NPM principles being more efficient than traditionally managed ones are. Whereas the management itself does matter, the management model necessarily does not, according to Alonso et al. (2015).

Administrative reform has a potential of rendering various levels of public accountability more obscure and less transparent. The long-term success of reforms requires consideration of the performance dynamics and recognition of appropriate changes in the culture of measuring the performance. Some reforms may have been carried out with the presumption that once the reforms are in place, better performance will somehow emerge automatically (Romzek 2000). This is however not always so, as e.g. the above discussed (Alonso et al. 2015) case study had showed.

Alonso et al. (2013) suggest that although the outsourcing activities in NPM implementations have not reduced public expenditure in general, the administrative decentralization measures have in fact resulted in cost savings.

Hood and Peters (2004) conclude that although there are several paradoxes with regard to claims and promises affiliated to NPM, the concept itself and the inherent requirement for continuous improvement still encourages scholars to seek for ways to device more rationalistic reforms. Pollitt (2006) investigated the performance management in four European countries of Finland, the Netherlands, Sweden, and the United Kingdom. He showed that certain general tendencies embraced all countries and most functions. In terms of NPM, these included an aspiration towards more sophisticated performance indicator systems – but controversially, at the same time also the feebleness of governments in developing performance-based strategic steering (Pollitt 2006).
2.4 Measuring impacts, impact evaluation of organizations

Pressures to cut public expenditure and to reach the most value-adding use of scarce governmental resources are evident across the globe. Simultaneously decision support tools are lacking for pin-pointing whether public services are yielding net benefits. Accountability is called for but the ‘accounting systems’ that validate the right choices in service delivery are not yet thoroughly established.

There is a long history in research of evaluating the performance of businesses, projects (e.g. Proost et al. 2014) and programs or reforms (e.g. Fuller et al. 2013; Mulley and Ho 2013). However, evaluating the quantitative impacts of entire public bodies such as administrations, agencies, and ministries on a socio-economic benefit-cost (B/C) level is a territory not as thoroughly charted out. Some promising material on the subject seems to exist on the headline level (e.g. European Environment Agency 2013). However, a closer examination reveals that regardless of thorough analyses from several viewpoints, the evaluation has not taken place in quantitative terms. Instead, the evaluation has been focusing either on qualitative or descriptive analysis, although some quantitative attributes have been addressed. As such, these indicators often manifest themselves as numbers on various scales, but this still does not deem them quantitative in their core nature. In fact, these kinds of indicators and metrics do not easily – if at all – relate to or translate to quantitative socio-economic impacts. Consequently, some gaps in the research can be detected.

While the overall picture on what the problems are (and in many cases, what the solutions would be) is relatively clear, the critical task of governing, managing and overseeing the public domain in an effective and socio-economically cost efficient manner is yet a largely uncharted area.

Firstly, this is probably due to the fact that such research is demanding and there is a lack of tradition. By the same token it is important to acknowledge that this is not the case without exceptions. Namely, for public organizations that are responsible for investments, such as infrastructure maintenance or construction procurement, there is a long and established tradition of submitting investment options to benefit cost appraisal. Hence, the term ‘lack of tradition’, here specifically refers only to the regulatory and administrative functions and services of public organizations, and not to their practices for investment decision justification.

Secondly, it is not self-evident that administrations are always eager to see their effectiveness, manner of governance and management critically scrutinized by
researchers. Such efforts may be regarded as uncomfortable by the managers of transport system, who furthermore are more accustomed to monitoring and overseeing than to be the focal subject themselves. However, evidence-based policy, much called for but perhaps less exercised than one would think, requires critical assessments.

2.4.1 Measuring impacts and impact evaluation

Why to perform impact evaluation analysis?

An impact analysis and evaluation exercise collects researched, fact based information to be used in decision making, management, resource allocation and performance management (e.g. Kohta 2007; Hytti and Mäki 2008). Impact analysis is especially applicable in analyzing non-profit public organizations or governmental programs and projects, since their productivity cannot be measured directly with private sector principles such as profit margins (VATT 2008). Impact analysis is used to justify resource allocations and performed activities. Client oriented approach, emergence of service culture and the requirement for increased transparency in the public domain have increased the need for impact analysis. Political decisions have to be made on justifiable, need-based grounds that acknowledge the relation between costs and benefits (e.g. Kuitunen and Hyytinen 2004; Sirviö 2016). An impact analysis can support the remodeling of the welfare state by developing citizen involvement, renewing the activity and widening employee know-how and knowledge base (Karjalainen 2007). The motivation of impact analysis is to help the analysis subject in achieving its goals and invest the available resources in a way that truly yields benefits. The conclusions based on the evaluation data promote the performance of the organization internally, on the organizational achievement level and with regard to public benefits (Guerra 2007).

The evaluation of any public agency’s services would need to serve the purpose of providing administrations’ management the information on how well the tasks and missions of the agency are fulfilled in terms of increasing socio-economic well-being and adding to the value of different functions of the society. Without this the management is not in fact doing its job in making sure that agencies and other public bodies are functioning properly, because the management lacks relevant information on the levels of performance. Partly this management control is associated with the need of continuous change, as pointed out by the OECD (2010).
**Challenges within impact analysis and impact evaluation**

Impact analysis has been generally determined as a challenging field to develop (Guerra 2007). The generalizability of results is often weak, the activities under scrutiny are often complex in nature, peoples’ behavior is complicated and the evaluation often deals with challenging problems (Earl et al. 2001). Also the causalities within the activities under the lens may be complex. There are challenges both on the practical and on the theoretical levels. On practical level, typical challenges involve finding indefinite impacts or impacts that are minor. It is extra challenging to show that the change has been in fact induced by the impact evaluation subject and not by external causes (e.g. Karjalainen 2007; Baker 2000; OECD 2017). The evaluation may become challenging if the impact measurement indicators are not selected properly (Guerra 2007). Evaluation of socio-economic impacts is more complicated than evaluation of economic impacts (Rajahonka 2014). The limitations of the evaluation setup and framework need to be acknowledged when assessing how realistic and trustworthy the results are (Karjalainen 2007; Baker 2000; Guerra 2007). According to Kuitunen and Hyttinen (2004), the development challenges in impact evaluation mostly relate to the three sub-categories of challenges in (i) clarification of the strategic objectives of the evaluation, (ii) finding and usage of indicators that mirror the operative impact generation, and (iii) utilizing the evaluation results in organization level learning.

**Impact evaluation process (planning and frameworks)**

According to International Association for Impact Assessment (IAIA), impact evaluation is a process that identifies the consequences of current or future activity (Sirviö 2016; IAIA 2017). In impact evaluation, the phenomenon or subject is systematically described to define its value or benefits (Hytti and Mäki 2008; Kuitunen and Hyttinen 2004). The evaluation can take place in various stages of the process but often the focus is on observing the achieved results in relation to the objectives or international benchmarks. Then the objective is to evaluate whether the evaluation subject generated the changes that were sought after, whether the changes would have occurred even without the intervention by the evaluation subject or whether the changes were brought in by a third, external party (OECD 2017; Storey 2000, Baker 2000). The intervention can be for example a project, a program, a policy or implementation of a policy. However, only
observing the results is not impact evaluation but monitoring. In addition to observing results, impact evaluation seeks to reveal wanted and not-wanted changes, causalities, benefits and confounding impacts. In other words, impact evaluation also investigates the process that is yielding the changes (Rajahonka 2014). This, i.e. the search for not only results but also for not wanted changes, causalities, benefits and confounding impacts, was distinctive to the research at hand.

Impact evaluation also investigates the produced added-value and the wider impacts to the operational environment (Storey 2000; OECD 2017). Added value analysis includes the assessment of how well the needs of the stakeholders are being met (Rajahonka 2014). In an appraisal of public services, the added value can also be described as the real or total impacts of the project, program, etc. (Storey 2000; Tokila 2011). This should focus on impacts or changes that would not have happened without the intervention or activity. The part of the impacts that would have happened without the intervention is also called ‘dead weight’ (Robinson et al. 1987; Tokila 2011). This notion on dead weight (also called externalities in this work) was particularly imperative to this dissertation since the adopted approach strived to only focus on the impacts that would have not happened without public intervention.

Originally impact evaluation was utilized in attempts to improve the living conditions in developing countries. Payments were deemed conditional based on showing the improvement of social conditions (Guerra 2007). Impact evaluation has been widely used within health care services but is increasingly been adopted in other sectors also (e.g. Palomäki and Piirtola 2012; Rajahonka 2014; Kuitunen and Hyytinen 2014).

The impacts may emerge in various forms, both temporally and in terms of location. Short term impacts are visible immediately and long term impacts may emerge more indirectly, and be regarded as investments in the future. The impacts may be quantitative and perceived as e.g. sufficient quantity of available services. Qualitative impacts may be shown e.g. as perceptions of the service quality. Impacts may be either positive or negative, either expected or unexpected and either client oriented or socio economical. External impacts appear outside the evaluation subject itself (Meklin 2001; Karjalainen 2007; Kuitunen and Hyytinen 2004; Baker 2000).

The motivators to consider carrying out an impact evaluation can be either internal or external. External factors include the productivity and performance management requirements that have spread out to the public domain, as discussed
earlier. Ever scarcer economical resources together with the growth of the public sector creates a need to inspect the products and results of the activity (Kuitunen and Hyytinen 2004). Internal factors include needs for internal development and learning. Successful organizations aim to predict the effects of external factors to their entity, in order to best secure both client satisfaction and economical profitability (Guerra 2007).

Proper planning of an impact evaluation exercise is key to its success. The literature presents some frameworks for impact evaluation studies that can be utilized to some extent. However, due to the usual variation between various analysis cases, impact evaluations normally need to be planned on a case by case basis. Theory based approaches help better understand and recognize causalities and situations where an intervention might or might not work. To some extent this also facilitates making cautious generalizations with regard to similar contexts (OECD 2017; Baker 2000). Instead of trying to cover all imaginable impacts, a single evaluation exercise should focus on studying some pivotal impacts, due to the often wide and complex impact mechanisms. The impacts can be evaluated both with quantitative and qualitative metrics (Kuitunen and Hyytinen 2004).

Guerra (2007) listed a seven phase framework as follows: 1) recognize the relevant stakeholders and their expectations, 2) decide the objectives, 3) set measurable indicators, 4) identify the data sources, 5) plan the data collection and the tools for data collection; collect the data, 6) analyze the data, and 7) inform, i.e. report the required further actions and recommendations for improvement (Guerra 2007). Another point should maybe be added, or at least included into the seventh point, namely monitoring the impacts of the impact evaluation. If the recommendations from the evaluation work are not implemented, the original objective of inducing positive change has failed (Rajahonka 2014).

Impacts may be evaluated (i) in relation to achievement of objectives and goals, (ii) as a consequence from an action or intervention, (iii) as the ability of an organization or a process to induce impacts, (iv) as a reflection of clients’ or other stakeholders’ needs, or (v) as a realistic evaluation, when the impact is evaluated with the aid of impact mechanisms (Baker 2000; Guerra 2007).

Holden and Zimmerman (2009) emphasize the importance of careful planning and consideration of causal relationships. Earl et al. (2001) underline that the impacts are in fact changes in the behavior, relationships, actions and activities of the targets of interventions. Therefore, and somewhat contradicting Holden and Zimmerman (2009), the method is not based on causal relationships but instead, on linkages between actions and the resulting behavioral changes.
In the Kaufman Organization External Model (OEM also known as Mega-level –model), impacts are evaluated in five stages. The model builds upon identifying flaws in the performance that indicate the need to do things in a different way. The first two stages involve analyzing the organization’s means to induce impacts (i.e. inputs and processes). The needs are evaluated on mega, macro and micro levels. The micro level contains the products and services of the organization, the macro level contains the results that come from delivering those products outside the organization and finally, the mega level analyses the impacts to the external clients, stakeholders or the society (Guerra 2007).

Kuitunen and Hyytinen (2004) constructed an eight stage framework for evaluating the impacts of public entities. The stages are:

1. Objectives for the evaluation (why to evaluate, to whom are the results for, what are the research questions, etc.)
2. Methods (collection of data, analysis methods etc.)
3. Objectives of the organization being evaluated (what are these and who sets them, what is the organization’s reason for existence, etc.)
4. The expected impacts of the organization being evaluated
5. Contextual factors (which causalities promote the induction of impacts, what are the external factors)
6. The relationships between the objectives and impacts
7. The utilization of evaluation results
8. Future perspective

As a general note regarding the abovementioned evaluation frameworks from earlier literature and their generalizability and usability in this research: they provided some benchmarks for structuring the process and also highlighted some important aspects to be considered during the evaluation process. However, on many levels they are not fully analogous to this work. For example, neither Guerra’s (2007) framework nor Kuitunen and Hyytinen’s (2004) framework actually deal with evaluating socio economic impacts of public organizations. Furthermore e.g. Kuitunen and Hyytinen (2004) specifically discuss evaluation of public research institutes, not service oriented public agencies. Regardless, some analogies can be drawn and therefore this literature is relevant also here, to an extent. The best practices identified in previous literature were utilized in this work, such as the importance of careful planning and consideration of causal relationships, differentiating between the externalities and the direct impacts of interventions and not only looking at results, for example. With regard to the frameworks, this
research started the framework design practically from a clean slate (mainly because an earlier framework that would have been applied to evaluating impacts of public services could not be found). Still, it is very interesting and noteworthy, that the framework presented by Kuitunen and Hyytinen (2004) bears a strong resemblance to the structure this research ended up with.

**Metrics and indicators**

Impact evaluation utilizes various case specific indicators and metrics. It is pivotal to select indicators that express the fulfillment of the evaluation object’s strategy but at the same time also mirror the change drivers originating from the objects external operational environment (Lähteenmäki-Smith and Hyytinen 2006; Rajahonka 2014). Often both quantitative and qualitative indicators are deployed. Numerical values tend to simplify things which restricts a multifaceted use of quantitative indicators. The strength of qualitative indicators is that they can be used to answer questions of why, how, when and where (Lähteenmäki-Smith and Hyytinen 2006).

In ideal cases an indicator describes the object efficiently, precisely, adequately, reliably and repeatably. An ideal indicator serves many purposes and is based on systematical data collection from e.g. statistics or monitoring data (Aistrich 2006). An indicator indirectly describes the changes in the status or development building upon quantitative metrics and measurements or on qualitative narrations (Karjalainen 2007; Guerra 2007; Haapanen and Mustonen 2006). According to Rajahonka (2014) the employed terminology within impact evaluation is not fully established and sometimes metrics and indicators are used as synonyms. Rajahonka (2014) summarizes that the stakeholders define what is being measured, metrics define where the data can be acquired and the indicator produces the numerical (or other) value. Indicators should measure the results and impact rather than activity and objectives (Hyytinen and Konttinen 2006). Indicators can be either stable or volatile in nature (i.e. constant or changing). Prioritization of indicators may be used to facilitate looking at the bigger picture rather than observing the minute detail (Lähteenmäki-Smith and Hyytinen 2006).

See also discussion on metrics and indicators under section 2.3.2.
2.4.2 Cost Benefit Analysis

Cost-benefit analysis (CBA, also called benefit to cost analysis, BCA) is used as one pivotal impact measuring and quantification method within this research. CBA is an economic analysis concept that can be applied as a support tool within public decision making, often used when there are several competing alternatives being considered. CBA compares the monetary value of the benefits resulting from a specific project or policy with the costs accrued by it.

In the public domain, CBA has so far been used mainly for assessing projects that include investments, rather than for analyzing service intensive public activity. In some ways CBA is analogous to calculation of return of investments and kin to traditional management accounting. CBA as such is therefore yet another specimen of applying business world paradigms and practices in the public domain. However, CBA is different from traditional management accounting because of the way it allows inclusion of benefits and costs not priced by the market but identified through e.g. shadow pricing. In the following, CBA will be described starting from the historical background, following to the theory and concluding into ethical issues of CBA usage in analyzing public activities and criticism towards CBA. Due to their lesser relevance to this work, this section (2.4.2) omits detailed theoretical discussions with regard to investment criteria, handling risk and uncertainty and market rates of interest in calculating discounted present values (e.g. Mishan and Quah 2007; Layard and Glaister 1994).

Historical background of Cost Benefit Analysis

There are two underlying concepts behind CBA, namely consumer surplus and externality. In 1844 Jules Dupuit introduced the concept of consumer surplus, stating that the benefits to the users of roads and bridges in France exceeded the cost of the tolls they had to pay in order to use them. In the 1920s, Arthur Cecil Pigou introduced the concept of externality by claiming that there is a difference between private economic production and public economic product (Mishan and Quah 2007).

The next practical step towards CBA was the enactment of the US Flood Control Act in 1936, which said that any flood control project should be considered desirable if the estimated costs are less than the benefits to the people affected by the project. This led to the assessment of projects based on their social impact, and not only on financial aspects. However, there were no specific and concrete
guidelines at place, which resulted in inconsistent practices at the relevant agencies. Therefore, in 1946 a committee was formed to define consistent standards and guidelines and proposed practices for the agencies to apply in economic analyses (Mishan and Quah 2007).

The committee’s documents raised academic interest, and a robust theoretical framework for CBA was established in 1958 in publications by the economists Otto Eckstein, John Krutilla and Ronald McKean. In the 1960s, research on CBA kept further developing, and CBA started to become more widespread and standard. Governments of United States of America, Canada and the United Kingdom started to demand it for certain policies and projects. Also some international organizations, such as the OECD, the World Bank and the United Nations adopted CBA between years 1969 - 1975 (Mishan and Quah 2007).

In 1981, US President Ronald Reagan issued Executive Order 12291, which in practice mandated the use of CBA in the form of regulatory impact analyses for regulations that have an economic impact of exceeding $100 millions. In 1993 Executive Order 12291 was replaced by President Bill Clinton with Executive Order 12866, which was largely similar to the original one and continued the commitment to CBA (Mishan and Quah 2007; Boardman et al. 2006).

**Definition and general theory of Cost Benefit Analysis**

As stated by Boardman et al. (2006), “CBA is a policy assessment method that quantifies in monetary terms the value of all consequences of a policy to all members of society.” The function of CBA is to assist in making political and social decisions by providing a tool for utilizing society’s resources in an efficient manner. CBA is used to weigh the costs and benefits of a policy in order to determine its net social benefits and thus the desirability of the policy. Net social benefits as defined in the equation \( NB = B - C \), measure the value of a policy, where NB is the net social benefits, B is the social benefits, and C is the social costs (Boardman et al. 2006). In addition to analyzing the net social benefits, CBA often utilizes the benefit to cost ratio (B/C) as means of comparing competing policy (or other) alternatives between each other. Benefit to cost ratio can be defined as \( B/C = \text{benefit to cost ratio} \), where B is the social benefits, and C is the social costs. When B/C is greater than 1 the benefits exceed the costs and the investment is profitable (Hautala and Leviäkangas 2007). These rules apply to benefits and costs identified in a single year and can be discounted if necessary.
CBA has several points and aspects that need to be decided or been aware of when performing the analysis. These include: (i) Discount rate selection. For example, according to risk-income theory, high risk investments decisions demand a higher revenue expectation and therefore a higher discount rate value (Hautala and Leviäkangas 2007). Gramlich (1998) argues that this principle should not extend to CBA, but regardless, the discount rate needs to be set. (ii) Selection of the analysis perspective, and the relating distribution of internal and external costs. For example, from the viewpoint of an individual organization only the cash flows and other costs and benefits that are directly and exclusively connected to the organization are perceived as internal. Cost and benefits connected to second parties are external, at least partially. In the perspective of societal economy all costs and benefits should be evaluated, whether they would be internal or external. (iii) Distribution of the costs and benefits to different parties (also see the previously discussed internal and external costs and benefits) and the possibly differing valuation principles of those parties. In other words, the same benefit may be valued differently by e.g. young people and elderly people. (iv) The timing of the costs and benefits. (v) The different valuation and monetization techniques for the costs and benefits (Hautala and Leviäkangas 2007; Gramlich 1998; Layard and Glaister 1994; Brent 1996). The above points (ii) and (v) are relevant in this study. Points (i) and (iv) do not apply here since a cross-section of time was observed, and the distribution effects (iii) were analyzed only in a very limited manner.

Additionally CBA may try to separately analyze the marginal impacts, i.e. the difference between a decision (an investment, a policy choice, etc.) being made and not being made. Typically investment decisions present these kinds of situations. Another option is to analyze the costs and benefits in a static situation, where no investment or other decision is even assumed to take place and the analysis observes a cross-section of time (Gramlich 1998). This static perspective averts several challenges arising from the calculation of current value, namely the effects of selection of the discount rate and the determination of the observation period. Then the analysis deals with net benefits and net costs instead of marginal benefits and costs (or marginal utility and marginal cost) (Hautala and Leviäkangas 2007).

Generally CBA counts benefits on a specific viewpoint but does not limit the benefits spatially. Weisbrod et al. (2013) discuss the use of Economic Impact Assessment as an extension of CBA or rather, as a complementing approach that also includes the spatial boundaries for impacts.
From temporal perspective, there are two main types of CBA: *ex ante* (before decision or implementation) and *ex post* (after implementation). In *ex ante* CBA, policies are assessed while they are still under consideration and not implemented yet. *Ex post* CBA, on the other hand, examines projects after they have already been completed. *Ex ante* CBA aids policy makers in deciding whether or not to launch a project under consideration, while *ex post* CBA is useful in learning from the performance of implemented projects when considering similar projects or policy choices in the future. Since it is more difficult to estimate the impacts of a project before launching it, *ex post* analyses tend to be more accurate than *ex ante* analyses (Boardman *et al.* 2006; Layard and Glaister 1994).

Third alternative temporal form of CBA is *in medias res*, which examines projects or policy changes while they are in progress, but not yet finished. It can be used to determine whether a specific project should be continued or not, and to provide useful information for predicting costs and benefits in future *ex ante* analyses. It is far more common for service-orientated programs to be cancelled based on *in medias res* analysis compared to physical investment projects, such as roads, buildings or bridges. The fourth class of CBA is the *ex ante / ex post* or *ex ante / in medias res* comparison. These compare *ex ante* CBA with either *ex post* CBA or *in medias res* CBA of the same project. It can be used for learning about the success of CBA as decision-making tool (Boardman *et al.* 2006).

The rationale for CBA can be compiled from three different intellectual traditions: welfare economics, microeconomics and public choice. Welfare economics is at the core of public policy and therefore also at the heart of CBA (Stanford Encyclopedia of Philosophy 2016). Welfare economics recognises that even if any intervention in a market economy might improve the welfare of the society as a whole, it might at the same time put at least some members of the society to a disadvantage (Gramlich 1998). This issue is further discussed below in the ethical perspectives of CBA. Microeconomic tradition discusses the optimal quantity of public goods and services in relation to the general equilibrium of economics, whereas the public choice tradition extends from these two normative concepts to discussing how decisions are in practice made within democratic societies. All these three principles factually in the end lead to the fundamental principle of CBA: In choosing between programs or projects or policy alternatives, the best choice is the one that maximizes net social benefits (Gramlich 1998).

The various selections, decisions and the general application philosophy of CBA in this research was described earlier in sub-chapter 1.3.
Valuation and monetization of benefits and costs

This section describes the valuation and monetization methods that can be used within CBA for determining the value of impacts (i.e. costs and benefits). There are several techniques for doing this, as summarized in Table 5 below (Leviäkangas 2009; van Wee and Roeser 2013; van Wee et al. 2014; Hautala and Leviäkangas 2007).

<table>
<thead>
<tr>
<th>Method</th>
<th>Key features</th>
</tr>
</thead>
<tbody>
<tr>
<td>Market value</td>
<td>The object of valuation (service, information etc.) has a clearly established market price, that can be used as a reference point</td>
</tr>
<tr>
<td>Historical cost</td>
<td>Is based usually on e.g. cost accounting or acquisition price (a special case of market value)</td>
</tr>
<tr>
<td>Shadow price</td>
<td>Whenever the market values do not include all known costs, the valuation is corrected (based on expert assessment or research data) to better reflect the true value. In transport economic evaluation the valuation is typically based on shadow pricing (e.g. accident costs, cost of time, environmental costs)</td>
</tr>
<tr>
<td>Willingness-to-pay</td>
<td>Is based on expressed willingness to pay a certain price (for a service, product, information, etc.)</td>
</tr>
<tr>
<td>Willingness-to-accept</td>
<td>Is based on expressed willingness to accept a certain price (for a service, product, information, etc.)</td>
</tr>
<tr>
<td>Cost of damage</td>
<td>Refers to the expected cost of damage resulted if specific actions were not to take place. Or the other way around, the benefit generated by avoidance of damage, (Cost of avoidance)</td>
</tr>
<tr>
<td>Avoidance cost</td>
<td>The production cost of the actions that lead to (or aim to) avoiding damage or decreasing the risk of damage.</td>
</tr>
<tr>
<td>Expert assessment</td>
<td>A researcher or an expert analyzes the price data (regardless of the method used in determining them) and takes a judgement on the monetary values to be used.</td>
</tr>
</tbody>
</table>

Market prices (or market value) refers to a situation where the product (e.g. a service) has a commercial value determined by supply and demand. Several techniques and data can be used to determine market values. These include acquisition or repurchase price, accounting value, insurance value, etc. or just simply a direct observation of a market price. The latter is a special case of market valuation called historical cost. Also court orders on monetary compensations can be used as basis for market valuations (Hausman 1993). Market prices can often include elements hidden from the naked eye, such peaceful and pleasant
surroundings of a real estate, where the location and social status may affect the price. In such cases the value of ‘peaceful and pleasant surroundings’ can be determined by comparing the realized prices between the different locations (Pearce 1983).

Market valuation as a technique is very applicable when the market operates efficiently. In reality this however is seldom the case and often the prices do not include all costs or all benefits of the commodity. For example Mulley et al. (2013) stated that the investment appraisal mechanisms underestimate the health benefits of improving walking and cycling networks due to lacking parameter values. In such cases shadow prices can be defined (and applied) by agreeing on the costs and benefits to be included into the valuation (Mishan 1972; Pearce and Nash 1989; Mulley et al. 2013). Shadow price is a chosen calculation value that e.g. in this research is often more applicable than market valuation. For example the unit values determined for traffic accidents consist of shadow prices.

Willingness-to-pay (WTP) is another price assessment technique. In WTP the valuators are placed in a position to make a statement concerning how they see the value of the valuation object (e.g. product or service). As technique WTP can be determined with questionnaires or interviews when obviously the value can be regarded as perceived. WTP can also be empirically observed, but indirectly. For example, the differences in house prices in certain areas can be explained by the willingness to pay to be recognized as being part of certain social status group. The same house and of the same quality might have a different price depending on the housing area (Leviäkangas 2008). The same rule applies, e.g. to use of transportation mode. People with low-incomes are more prone to use public transport than people with high-incomes. People with high-income social status are more inclined to be willing to pay extra to be in their own comfortable cars. WTP is always an indirect method, since direct observation leads to market values. WTP may sometimes fail the market test, meaning that upon actually delivering the products to the marketplace, the WTP values might not be realizable, or can be much lower than WTP assessment had suggested (Leviäkangas 2008). The use of WTP has been demonstrated in a practical context widely, by e.g. Anaman et al. (1997); Rollins and Shaykewich (2003). WTP is widely used because its relatively straightforward nature but it is not without problems and is sometimes even regarded inappropriate for use within CBA (Hausman 1993). WTP measures can also be established through stated preference activities.

Willingness-to-accept (WTA) is another valuation method that to some extent resembles WTP. WTA is the minimum price that a person is willing to accept to...
abandon a product or to tolerate something negative. Normally WTA is greater than WTP, based on the different formulation of the question: “what are you willing to accept” vs. “what are willing to pay” (Leviäkangas 2008). As with WTP, also WTA is equally sensitive to the market test.

Cost of damage or cost of avoidance principles refer to the cost of damage resulted in if specific actions (and consuming money, i.e. investments) were not to take place. For example, in trying to decide whether or not build a dam in order to prevent flooding, the potential damage caused by flooding could justify the cost of the dam (Leviäkangas 2009). Cost of damage plays a considerable role in many unit values for traffic accidents (e.g. road and maritime) used in this research.

Avoidance costs are production costs of the actions that lead to - or aim at - avoiding damage or decreasing the risk of damage (Leviäkangas 2009). In terms of the research at hand, also avoidance costs provide a significant perspective. This is because the assessed impacts of services are often here compared against the production costs of those services. The avoidance cost does not necessarily realize immediately or in a fully straightforward manner. Let us use a hypothetical example to illustrate this, in light of this specific research: assume a transport safety authority regulates (or contributes to a regulation) so that in the future all imported new vehicles are compulsory to have some novel traffic safety technology installed in them. (E.g. a collision warning system or eCall.) From the authority’s point of perspective, the immediate and direct ‘avoidance costs’ in this scenario are the time resources allocated into the preparation work of the piece of regulation. However, the indirect avoidance costs will be in this case probably more significant – and by several orders of magnitude. These costs relate to the price increase of any new vehicle caused by the extra technology. The payers for these extra costs are primarily the consumers who are buying new cars after the regulation enters force.

Expert assessment need to be resorted to in cases where none of the aforementioned valuation methods are possible. A researcher or an expert analyzes the price data (regardless of the method used in determining it) and takes a judgement on the monetary values to be used. Expert assessment is commonly used for example in valuation of real estates, indemnity insurance cases and in settling dispute situations. Expert assessment is always a subjective method where the accuracy depends on the qualification and competence of the evaluator and the quality and the quantity of available data (Hautala and Leviäkangas 2007). In this research expert assessments were conducted either in expert group work or in expert workshops. The expert assessments were based on data collected from the
literature. Some benefits assessed in this research had established valuations available that have been accepted into use within the sector.

In addition to this there exists a group of qualitative methods that can be used for indirect judgement of price or value. These include Analytical Hierarchy Process (AHP) and Multiple Criteria Decision Analysis (MCDA) (Mononen and Leviäkangas 2014).

Usually the methods discussed above need to be applied in various different ways depending on the case at hand. In other words, it is not reasonable to assume that a simple standard all-purpose process or a toolkit could ever be compiled for the valuation. It is also challenging to put these mentioned methods and techniques in any sort of rank or other order in an objective manner. Figure 8 illustrates the differences and applicability areas of different methods (modified from Leviäkangas 2008). The private sector tends to favor the methods grounded on economics simply because the operation of the private sector more often focuses on attributes that are measurable on monetary terms. In the public sector often the more indirect methods are applicable, since the variables within the decision making process commonly tend to be more numerous and heterogenic (Leviäkangas 2008).

Fig. 8. Valuation and monetization approaches, methods and techniques (Modified from Leviäkangas 2008).
As to this study, except for AHP and MCDA, all of the valuation and monetization approaches in Figure 8 were either used directly here or were an inherent part of some of the predefined valuations that were utilized in the monetization. For example, the accident costs used (such as Tervonen and Ristikartano 2010 and Finnish Maritime Administration 2008) include elements of e.g. cost of damage and shadow pricing. (MCDA was also utilized in this study, but its application did not relate to valuation or monetization, but instead to the sampling process.)

Ethical perspectives and criticism of CBA

As mentioned earlier, welfare economics discusses interventions that may improve welfare as a whole but at the same time may make some part of the population worse off. This ethical issue has raised interest in social scientists for more than a century and moral philosophers even longer. One logical criterion or a rule for showing that a change (a policy change, for example) is net beneficial was worded by an Italian nineteenth century social scientist Vilfredo Pareto: “Project X improves the welfare of society if it makes at least one person better off and no one worse off” (Gramlich 1998). The applicability of this rule could be manifested for example so that the winners of a policy fully compensate the loss to the losers. If after the compensation the winners are still better off than before the policy change, then the policy has shown net benefit.

From an economic efficiency viewpoint, another rule that includes these compensations or side payments was worded by two British economists Nicholas Kaldor and John Hicks into the so called Kaldor–Hicks rule: “Program X has positive net benefits if the gainers could compensate the losers and still be better off” (Gramlich 1998). Notably, the Kaldor-Hicks rule leaves the compensation in an open-ended status without actually stating whether the side payment should or should not take place - but could.

The abovementioned rules (Pareto optimum and Kaldor–Hicks) are in the core of this study, since regardless of compensations taking place or not, the adopted viewpoint here was focusing on socio-economic net benefits. In other words, the agency operation could be evaluated as overall net beneficial only if the costs of operating the agency could be compensated back to the society, and the society as a whole would still be better off.

Some critics argue that it is not moral to place a monetary value on human life (Ackermann and Heinzerling 2002; Frank 2000). Another criticism has to do with incommensurability. In CBA the pros and cons of an action have to be placed on a
common canvas by setting them a monetary value, even for things that are not
normally considered in terms of money (Frank 2000; Kelman 1981). The argument
is that, for example, if a power plant pollutes the air, the gains from the cheap power
can’t be compared in monetary terms to the loss of a pristine landscape. Supporters
of cost-benefit analysis admit that comparing disparate categories is immensely
difficult in practice, but critics insist that such comparisons are impossible even in
principle (Frank 2000). These ethical issues are highly relevant in this study. For
example, the nationally predefined accident unit costs include valuations of loss of
life. However, this study omits discussing the ethical implications in detail and
assumes the nationally agreed valuations as the best available data i.e. the state of
the art.

Others say that CBA is too rooted in utilitarianism (Frank 2000; Kelman 1981).
Utilitarianism is a theory in a category of theories called consequentialism.
Consequentialism is ‘the view that normative properties depend only on
consequences’. Utilitarianism ‘is the claim that an act is morally right if and only
if that act maximizes the good, that is, if and only if the total amount of good for
all minus the total amount of bad for all is greater than this net amount for any
incompatible act available to the agent on that occasion’ Benefits and costs should
be summed over all elements of society. The preferred option should be the one
that results in the greatest utility. Utility here serves as a proxy indicator for
‘happiness’ or ‘welfare’. Utilitarianism was developed by Jeremy Bentham and

One ethical criticism of utilitarianism is that it is not sensitive to the inequality
of utility (Roemer, 1996). Roemer uses an example of two utility pairs of two
persons as (1, 99) and (50, 50) where the figures represent the utility, that could be
for example their incomes. Summed up, these pairs have equal outcomes and from
a utilitarian perspective they are socially indifferent, but in reality they are in fact
very unequal. Egalitarian theories provide and alternative for utilitarianism.
Egalitarians claim that all people should be treated equally (Sen, 1992 and Sen,
2009) and focus on differences, not on absolute levels of well-being.

In terms of this study, the theoretical discussion on utilitarianism is valid,
although this work omits further discussion on its ethical implications. The adopted
approach here is basically utilitarian since the benefits and costs are here summed
over all elements of society.

The practice of using prevailing interest rates in discounting future amounts for
costs and benefits and converting them to present value may also be seen as
problematic. For example, in environmental protection projects the benefits may
often realize much later than the costs. The costs today and the benefits tomorrow can affect different generations when dealing with large time spans. A low discount rate increases the relative importance of the future compared to the present, and vice versa. Climate damages, radioactive waste and other environmental problems may become less visible at higher discount rates, but appear to have a higher impact at low rates (Ackermann 2008). For this study however, the interest rate discussion could be omitted, since the adopted approach observed a cross-section of time and did not include any discounting.

In addition, distributional issues have been a target for critics of cost-benefit analysis. The objection is that since willingness to pay is based on income, cost-benefit analysis might assign too large a decision weight to high-income citizens (Frank 2000).

However, since there is only a limited amount of resources for public officials to spend, it is sensible to use those resources efficiently by choosing the alternatives that have the highest benefits, e.g. prevent the most deaths or injuries. Where hard choices have to be made, CBA can sometimes help in finding the best decision in these situations (Frank 2000).

2.5 Synthesis of literature review

The literature review on the theoretical discussion presented above provides a solid foundation for studying the socio-economic impacts of a public agency and also facilitates the study of interactions and connections between the key components and concepts therein.

The sub-chapter on public administration and public policy provides the backdrop and the operational environment for the entire study. The discussion on performance management and new public management explains how the current public management system has evolved into what it is today, what are the drivers are, the justifications and pros and cons of the currently adopted approach and how the principles are or could be used according to the underlying theories. Combined, these two theoretical discussion areas are especially relevant in addressing research question 1 of this study.

Finally, the sub-chapter on impact evaluation details the thinking in science and applied practices around studying the performance and socio-economic or other impacts of organizations, especially with regard to public bodies. This theoretical discussion is relevant in addressing research questions 2, 3 and 4 of this study.
When looking at the approach adopted in this study, the relevance and importance of the earlier mentioned main theoretical discussion areas are also reflected. The operational environment of the study is public administration, the direction, mission and strategy of which is being formulated and dictated by public policies. Policies in turn, are set and adjusted by the prevailing government, adhering to the signals from the electoral term cycle and other recurring parliaments’ planning cycles. The agency, that is, the main unit of analysis was the source for the vast majority of the employed empirical data. The agency operates within the public domain, yields its output mainly via service provision that is supposed to materialize as positive socio-economic and other impacts and applies new public management principles in its performance management. Finally, the impact evaluation research entity, with all its planning and implementation and analysis activities, operated in the middle by collecting data from the agency and analyzing them. In addition, the research reformulated and developed the methodology and evaluation process, and provided the observations and recommendations as results both to the agency itself and also to its governing bodies. In other words, the covered theoretical discussion areas can be seen as gradually progressing toward the core subject matter of this research, moving from the more general to the more specific. For example, NPM can be perceived as a sub-area of public administration and similarly, impact evaluation is one of the ways of assessing the success of NPM. (Figure 9.)
The practical implications of this research – in other words, the pragmatic utilization of the research results – are envisaged to materialize along two potential avenues: (i) the agency learning, and the internal changes induced by that learning into its every day operation, and (ii) the recommendations and the required actions thereof being channeled from the governing body to the agency (and alike) via the steering mechanics currently in place. All this, that is, the research work and its connection to the theoretical foundation, is illustrated in Figure 10, where the black arrows represent the flow of main practical implications and change induced by this research. The practical and other implications of this research are discussed in more detail in chapter 4.

![Fig. 10. The interconnections between the main theoretical discussion areas in the scope of this study.](image)

The main theoretical discussion areas, the key concepts for each, and also the most focal references for each theory area are summarized in Table 6 below. These form the basis for analyzing the contributions of this research work to the original studies.
Table 6. Main theoretical discussion areas and concepts.

<table>
<thead>
<tr>
<th>Area</th>
<th>Key concepts</th>
<th>Main references</th>
</tr>
</thead>
<tbody>
<tr>
<td>Public administration and public policy</td>
<td>Accountability; (constant) reform; ideal of bureaucracy; policy design; intuition element; policy implementation;</td>
<td>Byrkjeflot et al. 2012; Gruening 2001; OECD 2010; Considine 2012; Schofield 2001</td>
</tr>
</tbody>
</table>

Earlier research of this kind

In terms of the specific scope of this research work, from intergovernmental organizations’ publication list only the UN-related Road Safety Fund’s “Decade of Action” (Road Safety Fund 2014) acknowledges governance and management as one of the key areas to answer the challenges of improving transport safety. All other approaches, by and large, maintain the managerial status quo meaning that seldom are institutional or high-level managerial practices and modus operandi viewed critically. Mostly, when management and governance are addressed, the questions relate to indicators, data gathering, statistics building, and other technical issues (Chapelon and Lassarre 2010).

In earlier literature, perhaps the most analogous and therefore the most comparable individual piece of research relating to this work is the evaluation of the Finnish Meteorological Institute (FMI) (Hautala and Leviäkangas 2007; Leviäkangas and Hautala 2009). The FMI research and the present study share some commonalities, although there are big differences as well.
The most remarkable common feature is that both researches were trying to determine the socio-economic impacts of a public organization. Both analysis units produce their tangible outputs predominantly as various kinds of services, through which the impacts are (assumedly) created. FMI’s services include weather services (e.g., forecasts), research and development services (meteorological research, space research, air quality, etc.) and technical services (e.g., measurement technologies and information systems). In terms of resource usage the focus of FMI’s activity is heavily within the research and development area, covering one third of the annual budget (Hautala and Leviäkangas 2007). Additionally, FMI is controlled and managed by the Finnish Ministry of Transport and Communications, that is, the same ministry that manages the traffic safety agency.

The most notable difference between FMI and the main unit of analysis of this research is the fact that FMI is not an administration at all, but a research institute. Therefore the practical activities and modus operandi within these two organizations differ substantially from each other. FMI is not involved in administrative duties such as regulatory work, performing control, carrying out inspections or permit granting activities. FMI services rely largely on information and data, in other words, they are information services. The research on information and the theories of information value as well as the entire theoretical framework around information is more mature than the research on public agencies’ socio-economic impacts. Some (but not most) of Trafi’s services are also information services, so there the two studies share some commonality. Another big difference is that as a government owned enterprise, FMI is just one competitor among others within the open market of weather service provision. A transportation safety administration does not have competitors in its field, per se.

To some extent, but not excessively, some other earlier research could be found that at least on the headline or title level appears comparable and relevant to the focus of this research. On closer examination, however, it turned out that either the applied lens was not on socio-economic impacts or that the applied methods were not transparent enough. For example, the European Environment Agency (EEA) was evaluated in a study completed in 2013 (EEA 2013). The focus of the study however, was not on the socio-economic impacts but on the administrative productivity (such as amount of reports produced, document handling latencies, etc.) and also on the stakeholders’ and clients’ perception of the quality and overall usefulness of the agency. Therefore, the applied approach and the methods used were neither applicable to nor comparable with the work presented in this research. As another example, a government-owned enterprise named Finnpiilotage Ltd.
which provides harbor pilot services for national and international vessels entering Finnish harbors, published an assessment of its socio-economic impacts as part of one of its annual reports (Finnpilot 2014). The average costs of maritime accidents were based on the same basic data used in this research in studying maritime services (e.g. Finnish Maritime Administration 2008). The quantification of the impacts was based on the pilots having themselves reported to have prevented a total of 57 serious accidents overall during that year. (By comparison, annually in Finland there are somewhere around 20 serious accidents. This figure also includes incidents or close-calls or near-misses that may have had a potential to escalate into serious maritime disasters (Trafi 2014).) This figure, 57, was then converted into direct socio-economic impacts by simply multiplying it with the average cost of a serious accident. The reporting or any other publicly available data did not discuss the validity of the method to make the deduction sequence transparent enough. Therefore it was not possible to compare with or utilize that evaluation in the work presented in this research.

In light of the literature review while also considering the earlier research of this kind, it can be concluded that the research problem of this study and the research questions that were derived from that problem, are valid. While there exists an abundance of relevant discussion on and around the issue at hand, the understanding with regard to the methods for deducing the overall performance of a public agency is still insufficient. Sound and transparent socio-economic performance metrics are needed to support better informed decision making. Therefore this research is necessary because of its aspiration to improve the current understanding and the available methods for determining these metrics.
3 Research contribution

3.1 Performance management system architecture and its coverage

Article I investigated the performance management system of a national transportation safety agency with qualitative methods. First, it introduced the evolution history and the surrounding institutional architecture of the agency. Next, the goal-setting, steering and management control mechanisms were described, followed by a cross-check of mandated tasks and objectives and the associated performance indicators.

Management by objectives (MBO) is a well-defined method of setting objectives to achieve the mission of an organization (Drucker 1974). The mission of Finland’s transport safety agency is based on national legislation, EU directives and the overall objectives of the Finnish state administration. The state applies MBO throughout the administration in order to operationalize and specify the mission of each ministry, agency and state institution. The main objective in article I was to test how successfully this managerial approach is applied to the transport safety administration.

The analysis focused on the governance and management architecture of the agency and specifically addressed the following points: What is the structure of the management system – including the surrounding governance and management structures – in terms of power, managerial performance indicators and managerial processes? How do the aforementioned indicators and targets correspond to the policy objectives stated by the national government and the EU? What are the potential areas of policies where the agency’s actions and services can potentially best benefit the society and its citizens and organizations (such as private sector companies)?

Article I first describes a brief history of Trafi (the agency) and its current governance and management system as part of the state administration and analyzes how the set targets and policies are met with the current performance management systems. The article presents a cross-check of the main tasks, objectives and indicators and shows that there are significant gaps between stated policy objectives, operational annual performance targets and available indicators, which became visible during the cross-analysis. Finally, the article discusses the implications of the discovered blind spots and sums up the recommended next steps to be taken to
fill in those gaps, in order to facilitate discussion between the agency and the overseeing ministry (Finnish Ministry of Transport and Communications) in the collaborative design of sustainable and balanced operational and strategic targets. Furthermore, the article aspired to initiate a discussion on how transport safety could be best and most effectively managed within a country and as an integral part of any intergovernmental effort.

As research material, official published documents and discussions with the agency’s management were used. The discussions played a minor role compared to the objective material since the main objective of the discussions was to identify and get access to the relevant internal documentation in addition to the publicly available data. The documents that were analyzed included the founding legislation for the agency and the performance management documentation. By far, the most focal document within this list was the performance agreement between the agency and its governing ministry. These documents comprehensively described the agency and its tasks, objectives and performance indicators, therefore representing the population (i.e. not a sample) of the performance management documentation and data.

The methodological approach in article I was a descriptive qualitative analysis. The data were analyzed by cross-tabulating the agency’s legislated tasks against the performance agreement’s indicator system and cross-tabulating societal impact objectives against the performance agreement’s indicators for i) transport system safety, ii) the environment and iii) service capability and quality. This was done in order to verify whether the objectives and their fulfilment were actually being measured.

As mentioned section 2.3.2, the entire Finnish state governance model and management system is based on performance management with MBO. Also the agency is managed through annual and mid-term targets that relate to tasks based on laws and decrees. Section 2.3.3 also referred to the relative complexity of overall process of obtaining the objectives for public bodies. Here an example from the Finnish transportation sector. (Figure 11 modified from Article I.) There the entire managerial process, including its international ties, is shown. The figure describes the “planning circle” or Deming’s circle (Deming 1982) of a transport agency as part of the national and international government planning system.
In figure 11 the actors are presented with italic font, the rest represent elements that feed from and into the process of defining the performance measurement metrics for an agency. Rather that attempting to introduce a construct of ‘going-around-in-circles’, the Deming’s circle is used here to depict the interwoven development cycles that aim at increasing the benefits to the citizens and society with each cycle.

Article I concluded that the indicators within the performance management agreement cover the direct legal mandate of the agency reasonably well. However, there exists an imbalance within the handling of different transport modes, especially for the environmental objectives and the societal objectives’ missing indicators constitute a blind spot in the performance management system. In the following, the implications of these findings are discussed in detail.

Trafi’s tasks, objectives and corresponding indicators to the latter cover the mandate of Trafi in the performance management agreement between the agency and overseeing ministry. This makes the management function relatively straightforward. Although the agency, as it now stands, is a relatively young organization, for the other segment of objectives that is equally present in the
performance agreement, the logic is not as straightforward. The societal objectives’ missing indicators are an ambiguous element in the management of the agency and a debatable issue when assessing the success and impacts of Trafi’s services. This is not to argue that when societal objectives and corresponding indicators are there, management problems do not exist. Part of the management task is to carry out subjective assessments as well, but the absence of indicators has several implications.

It is vital to understand why agencies behave and perform as they do. According to article I, poor performance can be associated with unrealistic impact expectations with regard to power and resource allocation. As pointed out in the early works of Peter Drucker (1954), performance and motivation are improved when results are measured. And without a measurement tool (i.e. indicator), part of the performance and motivation potential is lost. The rewarding or sanctioning of management must be based on a mutual understanding of the level of achievement. Without any yardstick, this task becomes a quest of negotiation and argumentation rather than real assessment.

Objectives with missing indicators are of high relevance. Table 7 (modified from article I) illustrates the cross-check for societal impact objectives, that are listed in the performance agreement as follows: (S1) travel and transport chains work seamlessly and safely enhancing well-being and competitiveness; (S2) the reliability and resilience of transport system is improved; (S3) the service level and quality of public transport is improved; the private car use is decreased and public transport volumes are increased; specifically, rail public transport services should be enhanced; (S4) Finland belongs to the top of EU with regard to traffic safety; (S5) greenhouse gas emissions are reduced according to international agreements; harmful impacts on environment and health are minimized; (S6) the efficiency of the transport system management is improved and enhanced through the use of modern ICT; and (S7) the climate change adaptation is supported in collaboration with other actors and research entities. (SIA, Safety Investigation Authority.) The crosses in Table 7 show where indicators relating to the societal objective can be found in the performance agreement.
Table 7. Cross-check of societal impact objectives and indicators (Modified from article I).

<table>
<thead>
<tr>
<th>Indicators</th>
<th>Societal objectives</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>S1</td>
</tr>
<tr>
<td>Transport system safety</td>
<td></td>
</tr>
<tr>
<td>Statistical definition of “injured” / “injury”</td>
<td></td>
</tr>
<tr>
<td>Fatalities and accidents in</td>
<td></td>
</tr>
<tr>
<td>commercial transport</td>
<td></td>
</tr>
<tr>
<td>Fatalities in non-commercial</td>
<td></td>
</tr>
<tr>
<td>transport</td>
<td></td>
</tr>
<tr>
<td>Environment</td>
<td></td>
</tr>
<tr>
<td>CO₂ emissions</td>
<td></td>
</tr>
<tr>
<td>New energy sources</td>
<td></td>
</tr>
<tr>
<td>Companies in energy efficiency</td>
<td></td>
</tr>
<tr>
<td>program</td>
<td></td>
</tr>
<tr>
<td>Service capability and quality</td>
<td></td>
</tr>
<tr>
<td>Vessels halted abroad</td>
<td></td>
</tr>
<tr>
<td>Inspection sites’ statistics</td>
<td></td>
</tr>
<tr>
<td>SIA recommendation processing</td>
<td></td>
</tr>
</tbody>
</table>

For example, indicator (S6)“use of ICT to enhance the management of the transport system and make it more efficient”, lacks an indicator although the theme is directly associated with the MoTC’s strategy alignment (Ministry of Transport and Communications 2013). ICT or Intelligent Transport Systems (ITS) have a prominent role, and several measures for how to enhance its use are recommended. The same applies to European transport policies. However, the objectives for Trafi in this respect lack direction and concreteness – and what you don’t measure, you don’t get.

One of the conclusions was that indicators for important societal goals need to be developed. This exercise must be done with care. Poor or wrong indicators could well take both attention as well as the results in undesired directions. A more efficient use of the agency’s extensive registers could be a good springboard for new ITS applications, both for public transport and private car traffic. Exploring those prospects could easily be measured, and the effort would most likely not be in vain. Lacking a significant automotive industry but still having a fairly strong industrial basis for mobile technologies and networks, the potential for more advanced utilization of registers and databases is visible.

It is not possible to avoid the impression that transport safety management is still relying on the ex post analysis of accident statistics. Simultaneously it can be
seen is that both the total and the relative road accident fatality numbers in Europe have been and are going down radically. The radical development in accident statistics raises some questions. How much of this development is due to all the traffic safety measures that are taking place in areas other than automotive safety technology? If the technology push is now doing the job, how should safety policies and strategies realign to provide maximal cost efficiency?

Gaps were also discovered within the environmental objectives’ indicators. When looking at greenhouse gas emissions, the trend is not at all as that with traffic safety. This means that with regard to the agency, the role of an environmental agency calls for more serious effort. Although this is already witnessed in the performance agreement, still all indicators regarding environmental objectives were observed to relate only to road transport. The efficiency of some of the steps taken toward more environmentally sustainable road transport remain under debate (e.g. Nygrén \textit{et al.} 2012), still indicators for other modes were entirely absent. One major reason for this may lie in the recent development history of the agency. The agency is still partially under the process of adapting to the shift from transport mode-specific thinking to a transportation system perspective. In the 2010 restructuring merger, most of the agency employees were relocated to Trafi from one of its mode-specific predecessors. The individuals’ long-term background and expertise is more than likely to be more substantial with regard to one transport mode than the others. This holds true in all employee categories.

The absence of indicators for other modes calls for a more balanced modal split in terms of administrative and managerial effort. Admittedly, some modes are regulated with international agreements more than others. But nevertheless, Finland is represented in the international fora that discuss the future changes to these agreements. In Article I, one practical recommendation was to involve rail operators in energy efficiency programs and to also take a closer look at other modes’ environmental impacts, even if road transport is still the major challenge. Article I claims that such a more balanced approach would probably be regarded as a motivating signal that all emissions and energy consumption matter.

Relating to balance between modes, another alarming observation was made: the absence of objectives with regard to non-motorized traffic i.e. pedestrians and cyclists. It is neither part of the agency’s legal task nor its performance objectives. Non-motorized traffic safety is on the agenda of MoTC and some associations, but it does not appear in the performance agreement objectives of any sector authority (Trafi, FTA), although some indicators – like “road fatalities”, for example – do
include some non-motorized traffic. This observation has urgent practical implications, as discussed later.

Some performance management indicators may include metrics that are affected by numerous phenomena external to the object that is being managed (Mononen and Leviäkangas 2015). The challenge then is how to isolate the impacts induced solely by an agency’s actions from other changes in its field of operations. For example, a typical metric in any transport safety agency’s performance management contract might include elements from statistical trends in transport safety, e.g. casualties on roads. But those trends are in fact affected by numerous factors such as developments in vehicle’s technical safety, macro-economic fluctuations’ effects on traffic volumes, enforcement, infrastructure – just to name a few. Many of these are external to the jurisdiction of the agency. This issue is further discussed later under practical implications sub-chapter (4.2) of this dissertation.

Full 360 degree coverage of metrics is not a practical, nor an advisable target but at the minimum, exogenous factors that are affecting from outside the agency’s mandate should be excluded from the equation when attempting to measure the sovereign impacts of an agency. In other words, it should be ensured that performance related metrics are under the control of the entity being measured.

To sum up the contribution of article I, it covered the agency’s evolution history and the surrounding institutional architecture; national and agency specific performance management architecture; steering and management control mechanisms; cross-check of mandated tasks and objectives and the associated performance indicators; and the deduction of the coverage and gaps of the of performance management system. The main finding was that significant gaps between stated policy objectives, operational annual performance targets and available indicators could be identified. Especially with regard to societal objectives, the steering framework turned out to provide unsatisfactory coverage. Performance indicators for some major objectives were missing and vice versa, and some measurement metrics did not seem to link clearly to set objectives. Not all the set objectives need or even could necessarily be measured, but certain shortcomings in the performance control system may be critical. The findings implied that there is a risk of sub-optimal use of public resources if the targets and indicators of agencies are not thoroughly considered so that they logically cover agencies’ mandates. The implications of the discovered gaps were outlined, together with recommendations for a more balanced approach. The analysis
concluded with some recommended steps in order to cover the blind spots. With the aid of these steps, performance management systems can be improved to better meet policy and societal objectives.

Article I mainly addressed research question 1 and its sub-questions.

3.2 Service architecture and the first steps towards investigating socio-economic impacts

Article II presents the first steps in a process of prioritizing services of a public organization via a case study of a Finnish agency. The prioritization is made in order to enable later analysis of the effectiveness of the agency and the evaluation of the impact of its services. The main contribution of this article is in laying out a process for identification of the services that are most relevant in this respect. The article presents one viable approach and process to solve one challenge of the evaluation process, namely the selection of what to evaluate and what to leave out – i.e. what is reasonable and possible to be evaluated and what not. Article II contains a presentation of a viable method to select an appropriate sample for impact evaluation research on a public agency. Article II does not argue that the applied method is the only viable process of performing the prioritizing. The prioritizing phase covers only the first early steps of the entire impact evaluation research entity. Article II did not yet go into the actual impact evaluation process of the prioritized subset of services.

As mentioned earlier, in Finland the transport sector has been in the forefront of the renewal processes of different administration sectors. Among many others the Finnish Transport Safety Agency in its current form is a fairly new entity in the transport sector governance architecture. The historical evolution of the agency to its present form has been described in (Mononen et al. 2014).

While the restructuring of administrative architecture has been systematic, it has also brought along several challenges. These include difficulties to assess efficiency and effectiveness of public services as the structures – including the cost structures – have been changed and there is no historical track record upon which to build management control of costs and effectiveness. It is also challenging to draw clear-cut lines between fully public, semi-public and private (or to-be-private) services in order to maximize the value for money of services provided to citizens and other customers, such as private sector companies and other sectors of public administration. Further, it is not easy or straightforward to avoid historical build up and accumulation of duties by preventing de facto non-value adding services to
exist, i.e. services that have been a part of the former public administration service architecture but for which in the current world the demand may be low and which could be supplied via alternative channels or suppliers, or even cancelled entirely. Also setting effective and fair performance targets for the operational management of the agency presents another challenge.

The need to evaluate benefit-cost ratios or cost-effectiveness of Trafi’s services stem from wider evolution and change processes of administration. The change processes were started already in the 1980’s as part of global transition towards new public governance and administration models that adopt managerial tools and practices merging from the business sector. The evolution history of Trafi shows that these processes of institutional redesigns are ongoing.

The prioritizing of the agency’s services was built on the need to exercise responsible management control. The services that were after the prioritization analysis regarded as most relevant, were later exposed to full benefit-cost and cost-effectiveness analysis. Prior the research at hand, some traffic safety administration evaluation has been carried out by the Finnish National Audit Office (2001) and Finnish Ministry of Transport and Communications (2013). These evaluation reports dealt mainly with roles and responsibilities of country’s different traffic safety actors and with the agency’s internal efficiency. While the reports excluded in-depth evaluation of the agency’s services they recognized the need to assess the benefits and costs of different functions of the agency.

The applied process entailed 4 iterative steps: 1) service catalogue mapping, 2) design of evaluation criteria, 3) compilation of decision matrix, 4) Multiple Criteria Decision Analysis (MCDA) - and recommendations thereof. Steps 1-3 were discussed in this article in detail, the end result being a decision matrix for MCDA. In the following, a summary of this process is presented.

First, a conclusive list of the agency’s services and functions was accumulated in a multi-phase process. A classification of services was created with a top to bottom approach, where first the list of potential targets/functions of various services was created. The targets then were classified by transport mode and finally each of the agency’s services and functions were allocated to these classifications. (See Figure 12.) Next, a service type was attached to each service as meta-data for later use.
Fig. 12. Service classification structure (Modified from article II).

After the service listing and classification, the criteria for the prioritization exercise was designed. The design process was based on the general transportation knowledge and impact evaluation research expertise within the research group and literature survey. Based on the agency’s main responsibility areas, the main criterion sets were formulated and divided into three main groups of 1) transportation safety and security 2) transportation environmental effects and 3) other criteria. The final set of criteria is presented in Table 8 below. Note that the term ‘individual’ includes all relevant groups, i.e. passengers, travelers and other personnel – e.g. in aviation the cabin personnel, the ground personnel at or around airports etc. Similar logic was applied across all transport modes.
Table 8. Evaluation criteria for the agency’s mandate areas.

<table>
<thead>
<tr>
<th>Traffic safety and security criteria (CS)</th>
<th>Environmental criteria (CE)</th>
<th>Other criteria (CO)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Individual skills</td>
<td>Equipment maintenance</td>
<td>Resiliency, normal circumstances</td>
</tr>
<tr>
<td>Individual condition</td>
<td>Equipment, propulsion fuel</td>
<td>Emergency circumstances</td>
</tr>
<tr>
<td>Individual education/attitudes</td>
<td>Equipment, emissions</td>
<td>Transportation system status</td>
</tr>
<tr>
<td>Equipment maintenance</td>
<td>Demand, output</td>
<td>Transportation markets</td>
</tr>
<tr>
<td>Infrastructure condition</td>
<td>Infrastructure condition</td>
<td></td>
</tr>
<tr>
<td>Infrastructure design</td>
<td>Infrastructure design</td>
<td></td>
</tr>
<tr>
<td>Infrastructure capacity usage</td>
<td>Control / Overseeing</td>
<td></td>
</tr>
<tr>
<td>Actor</td>
<td>Individual mobility choices</td>
<td></td>
</tr>
</tbody>
</table>

Next the evaluation scales were designed together with the decision matrix structure and the final decision matrix was valuated and compiled. The list was valuated with regard to 20 different impact criteria using an interactive workshop method (e.g. European Commission 2010). These iterative phases of service catalogue mapping, criteria design, pre-filling and evaluation, yielded the completed decision matrix. This was the main data for the next research phase of decision analysis to prioritize services for further analysis. After the process, the decision matrix at this point included 149 services valuated on a scale from 0 to 4, each in regard to 20 criteria. Inherently, all transport modes resided in the matrix with indifferent weighting in comparison to any other transport mode group. This means that no mode is deemed more important than the other, regardless of the manner of how generously or sparingly the process ended up giving the valuation ratings on the 0-4 scale. In other words, this study did not try to define any ‘importance levels’ of various modes, hence the word ‘indifference’ here relates not to importance, but to the method for avoiding the risk of bias. The strength of this approach is that the most promising and potential services to be evaluated in full-scale in the following research will be highlighted separately for each mode category. Failing to do so would have risked ending up in a lopsided and hence unsatisfactory sample of the agency’s service and responsibility portfolio.

It should be noted, that the above description of the process is a simplification of reality. As a matter of fact, the process was not linear, but more iterative in nature. The choice of analysis method had to be considered at all times during the design of the research data format, structure and collection method design; and vice-versa, the capabilities and restrictions of all of the available analysis methods needed to be kept in mind at all times in the data structure design and collection phases. This
iterative procedure and open-endedness had to be in place to avoid further mismatch between the collected data and the analysis method of choice. A flowchart illustrating the iterative nature of the prioritization process and research design phasing is presented in Figure 13.

Fig. 13. Iterative research process (Modified from article II).
Article II showed the first steps in a research process for prioritizing the services that were to be evaluated in more detail. Not all services are even possible to be evaluated and some other services, mostly based on existing legislation, are obliged to exist in an administration’s service architecture. On the other hand, excessive metrics in management control could lead to controversial results (see e.g. Vakkuri 2013 or Thomson and Martin et al. 2010) and reduce agility. Furthermore, it could lead to reduced accountability, meaning that accountability may be reduced to those issues covered by performance metrics only.

The results achieved by this stage of research indicated that collection of data for prioritization of services is possible in a structured manner. The selected approach suited well for the data collection, and in later stages it was observed that the data was well suited for use within an MCDA method. Furthermore it was observed that the process presented was highly normative, so similar approaches can be implemented elsewhere. Repeating the approach in other settings, such as for corresponding agencies in other countries, would be highly interesting. In that way further information on the versatility and generalizability of the presented approach could be acquired. The developed process proved successful, hence clearing path for similar analyses on other agencies, administrations and services elsewhere. Even more sophisticated methods are hopefully being considered, but the approach described here is one potential foundation for future development. It should be noted, though, that any prioritization is dependent on reference groups’ preferences and perceptions and by definition are colored according to the interests of the group. Therefore, prioritization could also be considered to involve several interest groups in order to gain multiple angles and engage a wider community to assess which services are essential, relevant and meaningful to be included in the impact evaluation. For this particular study (and its perspective and needs) though, both the selected interest group and the approach were appropriate.

Another challenge to be dealt with in following research is the chronological preference. In other words, some services are bound to be more relevant in the long run and hence carry a strategic loading. For example, the services or tasks related to international co-operation with European Commission and EU central agencies are time consuming but the prospective impacts of lobbying and influence work could be substantial in the long run. Examples of directives that have radical impact on transport sector’s performance are known.

To summarize article II contribution, it covered the processes of service architecture mapping, service cataloguing, service classification, evaluation criteria design, and decision matrix design as well as decision matrix compilation. The next
step in the research – as described in the following sub-chapter 3.3 – was the application of the acquired decision matrix within applicable MCDA methods.

Article II mainly addressed research question 2 and its sub-questions.

### 3.3 Methods for determining cost efficiency of public services

Article III presents a process for validating vertical economic accountability of a public agency. The main contributions of the article are presentation of a novel and repeatable approach to select a sample of functions or services for full-scale impact evaluation; description of applicable methods to chart out the impact mechanisms of various types of services; presenting examples of quantification of the socio-economic impacts and the associated production costs for the services; demonstrating cost/benefit appraisal of individual services, and finally; taking a first look into an agency’s overall accountability. The article also discusses the distribution effects of benefits from socio-economic impacts, since the final scope of evaluation cannot be limited to looking at the research subject internal productivity. It additionally must encompass the impacts from the perspectives of public economy, national economy and societal economy. Article III had an additional ambition of initiating discussion on how impact evaluation could be seen as an integral part of any intergovernmental effort.

Article III starts by presenting methods and processes for selecting a sample for an impact evaluation study, together with the reasoning behind such a sample selection. Via examples of performed analysis the article then describes ways of charting out logical impact mechanisms and methods for quantification of impacts. Finally, article III discusses the implications of the findings and sums up the limitations of the study and possible further use of the findings.

Article III provides new information on the current state of accountability. From the various definitions and categorizations for accountability in literature, article III constrains to vertical accountability as defined by e.g. Schillemans (2011) and, more precisely to the management dimension of accountability with regard to economic ‘bottom line’ and organizational performance targets, as constructed by Byrkjeofto et al. (2012).

The analysis within article III focuses on the process of quantifying impacts of public services. The aim is to support the governance and management architecture of the agency and specifically address the following points. What types of services’ impacts are quantifiable on a socio-economic benefit level? (i.e. How can we recognize such services?) What is the connection between a public service delivery
and a tangible impact on the field? (i.e. What are the impact mechanisms?) How can the impacts be recognized, isolated and quantified? What are the areas of policies where the agency’s services can best benefit the society as a whole, thus responding to the accountability requirement?

The methodological approach of article III includes descriptive qualitative analysis and quantitative methods. As research material the article relies on published official documents and discussions with the agency’s top level management. The documents and data that were analyzed covered a wide array of legal documents comprehensively describing the agency, its tasks, objectives and performance indicators. Furthermore, all available documentation and statistics with regard to planning, production and delivery of the agency’s services were utilized. Targeted literature reviews were performed to search evidence of quantifiable impacts of services or service types. The data were analyzed by the following methods:

Full list of the case study subject service portfolio (described in 3.2 earlier) was used in the selection of sample with Multiple Criteria Decision Analysis (MCDA) method ELECTRE III. Service descriptions were utilized in constructing causal relation maps to illustrate viable and logical impact mechanisms for the services within the sample. Collection and synthesis of evidence with regard to quantitative impacts of a service or a service type was carried out. Quantification of (i) production costs including cost recoveries and revenue returns from priced services and (ii) impacts, on socio-economic cost level and (iii) benefit/cost ratio by combining (i) and (ii), was carried out.

The applied quantification and cost/benefit calculation values related to induced benefits in traffic safety (e.g. reduced casualties, injuries, material damage); environmental benefits (e.g. reduced CO2, pollutants, noise), time savings (e.g. traveler choices, market control) and other benefits (e.g. data hand-out and value of information to data end-users’ competitiveness or efficiency). In article III examples are shown both for the impact mechanism construction and for the quantification of impacts.

To illustrate the research continuum, article III begins with a short summary of the agency’s basic characteristics and the preceding research phases (i.e. service catalogue accumulation, evaluation criteria and scales design and decision matrix compilation processes). Next, the properties, strengths and limitations of MCDA methods are discussed, followed by a description of how the MCDA method ELECTRE III (e.g. Bernard 1973; Clemen 1996; Belton and Stewart 2002; Jablonsky and Dlouhy 2003) was applied in this research in condensation of data,
weighting calculations, MCDA analysis runs, definition of the sample, and the sensitivity analysis cycles.

After the sample definition phase, based on the full chain of events discussed earlier, all of the 149 agency services were classified into three baskets in the following manner: Basket 1: services that would undergo full-scale quantitative impact evaluation process; Basket 2: services that would be evaluated with qualitative methods, and Basket 3: services that at this point were excluded from further analysis. Sensitivity analysis was performed to verify the validity of selecting a subset of services to the MCDA analysis and to verify the robustness of the data and its ELECTRE III analysis. As an overall result from the sensitivity tests, it was concluded that the selected approach had excellent stability and as such was compatible and robust enough to be used for ELECTRE III analysis.

Article III continues by presenting two examples (out of the total of 19 that were carried out) of constructing the impact mechanisms for services, quantification of impacts and CBA. The selected examples are 1) the regulation and control of driver education and the associated permits and qualifications and 2)
Port State Control inspections. Figure 14 shows an example of an impact mechanism construct.

![Logical construct of impact mechanisms – case 'inspections' (Modified from article III).](image)

One of the main contributions of article III lies in facilitating efficient use of research resources. Inevitably the resources committed to accountability and impact evaluation research are and will remain within a finite boundary. Therefore, it is of core essence to be able to limit the scope and volume of a study (i) to a manageable magnitude and (ii) so that the research focus is set on the right and most efficient place. This can be efficiently achieved by a sample selection process, as article III presented. As article III also showed, there are viable ways to chart out logical impact mechanisms for public agency’s service portfolio. The key points to be looked at are tangible changes at some element of the transport system, whether it would be a change in behavior, a change in equipment maintenance levels, or an improvement in skills. Furthermore, it is possible to quantify these impacts on socio-economic monetary terms, as article III presented. Overall, the presented result examples promised positive outcomes with regard to general accountability of the case study subject, e.g. when comparing to state budget input that was showed. In an ideal case, the basis and evidence for the quantification can be acquired from scientific literature, either as analytical models (e.g. causation...
mechanics within a road accident and death/injury probability correlation to usage of seat belt) or as validated empirical findings (e.g. quantitative analysis based on an empirical data set). Other viable models for the quantification include un-validated empirical analysis (e.g. based on expert interviews or case study analysis) and deductive or heuristic models.

To summarize article III contribution, it covered the application of MCDA, selection of sample, logical impact constructs, quantification methods, two examples of service CBA and an initial discussion on impact and benefit distribution effects. The next phase of research – as discussed in the following sub-chapter 3.4 – included analyzing (not only some examples but) the full set of services that had been evaluated for their socio-economic impact and deducing the agency’s overall efficiency by applying the findings to system level.

Article III mainly addressed research questions 2 and 3 and their sub-questions.

### 3.4 Cost efficiency of a public agency

Since there is a limited amount of prior tradition in this kind of research, the main contribution of article IV is methodological, including the benefit to cost (B/C) appraisal of services and service bundles, and; evaluating an agency’s overall B/C ratio by applying the findings to systems level. As in articles I, II and III, the unit of analysis in article IV was Trafi. Article IV commences with shortly summarizing the preceding research phases, as discussed in 3.2 and 3.3 above. (E.g. full list of the agency service portfolio, valuating the list with regard to 20 different impact criteria using an interactive workshop method; the valuated data being used in the selection of sample with MCDA method ELECTRE III and the several methods that were applied to constructs of impact mechanisms, impact evaluation, benefit to cost ratio determination and benefit quantification/monetizing.)

In article IV, the data were further analyzed by the following methods: the sample validity was analyzed with service clustering and cluster impact profiles, which were tested for uniformity (i.e. for mathematical congruency) with Spearman’s rank correlation coefficients (e.g. Marascuilo and Serlin 1988), supported by visual inspection (i.e. for visual congruency). The validity of the results from sample services’ impact evaluations were analyzed with non-parametric tests (e.g. Marascuilo and Serlin 1988). The sample results were scaled-up to reflect the agency as a whole by first looking at the public funder’s return for their investment and second, by analyzing conservative approximations of the observed benefit to cost ratios of the evaluated sample services.
All 20 criteria were included in the calculations, aggregated into eight logical sub-groups. The aggregation was performed so that each group simultaneously and logically highlights both the impact type (safety, environment, etc.) and the target group for the impact (individuals, equipment, etc.). The aggregation also facilitated better visualization of the impacts. Across all eight sub-groups, in the following this set of eight values is referred to as an ‘impact profile’. The validity of the sample was tested by comparing the full sample impact profile to the full set of the agency’s services, and compiling service clusters around a subset of sample services and comparing the cluster impact profiles to the sample impact profiles.

The conformity – that is, the similarity or dissimilarity of the all of the pairs of impact profiles – was tested with Spearman’s rank correlation coefficient $\rho$ (Marascuilo and Serlin 1988). This effectively yields a mathematical representation of profile congruence. In cases with perfect positive or negative correlation ($\rho = 1$ or -1) the function of rank orders (on the right in Figures 15 and 16) would be fully monotonic. The values on the axes represent arithmetical sums of valuation scores, for the specific criterion sub-group. In the test, Spearman’s rank correlation calculation yielded a correlation coefficient of $\rho = 0.9762$, which indicated a very strong positive correlation between the two impact profiles (Marascuilo and Serlin 1988). The critical value for the test’s statistical significance was $|\rho| \geq 0.88$, when the desired risk of error was $\alpha \leq 0.005$.

Fig. 15. Impact profiles for sample and full service portfolio (Article IV, reprinted with permission from Elsevier).

Next the service clusters were compiled. The sample services acted as a ‘stem services’ for each service cluster. The clustering was based on a generic description
of the stem service; all services allocated into a cluster needed to conform to the corresponding generic description to the letter. Seven clusters were compiled (see Table 9), and impact profiles were calculated for all of them, based on the decision matrix data. Spearman’s rank correlation coefficients were calculated for single stem services and the corresponding cluster profiles. An example is shown in Figure 16 and Table 9 shows the summary of the analysis for all clusters.

Fig. 16. Impact profiles for stem service and the corresponding cluster – an example ‘driver training organization permits’ (Article IV, reprinted with permission from Elsevier).

Table 9. Summary of impact profile analysis.

<table>
<thead>
<tr>
<th>Service cluster</th>
<th>n</th>
<th>$\rho$</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Personnel training organisation permits</td>
<td>12</td>
<td>0.92*</td>
<td>&lt;0.005</td>
</tr>
<tr>
<td>Personnel permits and licenses</td>
<td>20</td>
<td>0.49</td>
<td>&lt;0.2</td>
</tr>
<tr>
<td>Inspections of equipment</td>
<td>14</td>
<td>1.00*</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Equipment permits and approvals</td>
<td>6</td>
<td>0.88*</td>
<td>&lt;0.005</td>
</tr>
<tr>
<td>Information services, data hand-out</td>
<td>7</td>
<td>0.96*</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Infrastructure responsibilities</td>
<td>9</td>
<td>0.99*</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Regulation and international co-operation</td>
<td>5</td>
<td>0.98*</td>
<td>&lt;0.001</td>
</tr>
</tbody>
</table>

* = statistically significant; n = number of services allocated into each cluster; $\rho$ = Spearman’s rank correlation coefficient according to Marascuilo and Serlin (1988); p = the critical p value according to Zar (1984)

Based on the above analysis it could be concluded that both the sample selection process and the service clustering approach had performed well. The sample is an encompassing representation of the agency’s service repertoire, and according to
the cluster analysis, the impact profiles of the sample do not bias towards any impact areas. Six out of seven clusters’ impact profiles conformed to that of the specimen with statistically significant positive correlation and the remaining exception also indicated positive correlation. The author’s intuition on explaining why permits and licenses might have showed lesser significance is the following. Out of the 7 clusters, the permits and licenses sample may have been the most heterogenic one, in terms of comparing it to the corresponding cluster. The two services that were within the sample represented the permits and licenses in aviation and maritime. Especially the maritime service was somewhat peculiar in the sense that the particular sample service covered licenses issued to personnel that are responsible for maintenance of both infrastructure and vessels. Most (11/20) of the services within this service cluster relate to granting various licenses for drivers (common car driving licenses, for pilots, for captains, for engineers, etc.) and very few services relate to maintenance issues. Therefore the traffic safety implications (to infrastructure and to vehicles or vessels) may have been overrepresented in the sample, when comparing it to the service cluster.

The statistical significance of the findings was tested in three phases with binomial distribution analysis. First, the distribution of services into the service clusters was tested with two fractals of ‘the cluster stem service B/C ratio was observed to be under one’ and ‘the cluster stem service B/C ratio was observed to be equal to, or greater than one’. Second, between two fractals of ‘the stem service B/C ratio was observed to be equal to, or greater than one’ and ‘the stem service B/C ratio was observed to be considerably greater than one’, was performed. Finally, the sample observations were tested. There were those services within the sample of which B/C value could not be calculated (i.e. they were moved to basket 3) represented one fractal, and the rest of the services the other fractal. The results from tests indicated that the probability to reject the null hypothesis when it is really true (error type I, false positive) is less than 0.0001. The same applies to false negative (error type II), i.e. probability of failing to reject a false null hypothesis is less than 0.0001. For sample observation testing, $p = 0.0192$. The tests indicated statistical significance (assuming a significance level threshold of 5%).

To illustrate the overall implications of the findings with regard to the sample evaluation, the perspective of the public funder was first considered. As was observed, all of the sample services that could be evaluated yield positive net impacts, i.e. all of the observed B/C ratios were at least equal to, or greater than one. By generalizing this minimum observation to cover the entire agency a descriptive viewpoint to the return of investment for the public funding was
acquired. Assuming that the agency as a whole operates with an overall B/C ratio of exactly one. The public funding covers 23% of the costs of the sample (by comparison, including all administrational functions etc. the public funding covers 29% of the agency’s annual operational costs) (Trafi 2014). Therefore, by looking only at the public funding the total annual incremental benefit (or return of investment) is 1/0.23 (or 1/0.29) and B/C ratio is 4.3 (or 3.4, respectively).

Up to a certain limit, this above example was illustrative and also somewhat meaningful from the point of view of the public funder, i.e. the state. But in socio-economic perspective the comparison was not fully rational. Specifically, it only showed that the overall operation of the agency is not yielding net loss and hence the state funding is not in unproductive use, at least. In reality however, the operational costs to maintain and run the agency are collected from all across the entire society, not just from the state. And similarly, also the impacts (benefits or dis-benefits) are distributed across all imaginable stakeholders within the nation, including other public organizations, commercial enterprises – all the way down to single individual citizens. Therefore the sample needed to be analyzed further.

To analyze the B/C ratios that were deduced for the sample services, some conservative and cautious simplifying approximations were made that represented the minimum of what was observed in the evaluation phase. (In fact, the approximations either represented the minimum observation, or were cautious estimates or in some cases were true values and did not include any presumptions.) A weighted minimum B/C ratio for the sample was defined and with this a summary of service evaluations was compiled. The weighted minimum B/C ratio of the sample as a whole was calculated to be 2.6. Based on the used approximations, this could be regarded as the probable minimum benefit-cost ratio of the agency’s service portfolio. The theoretical absolute minimum is B/C=1 and the earlier discussed public funding B/C ratio could be regarded as the probable maximum (and at the same time the lower threshold for potential impacts). Figure 17 illustrates the socio-economic impact according to the above.
The agency’s socio-economic impact most likely lies in between the probable minimum estimate and the potential benefits’ lower threshold. Calculating based on the agency’s annual overall costs, the probable minimum of socio-economic impacts and the probable maximum of socio-economic impacts, the agency’s annual socio-economic gross benefit was calculated to be between 387–641 million euros. By subtracting the annual operational costs from these ends of the range, the agency’s annual socio-economic net benefits were calculated to be between 238–492 million euros. The socio-economic balance sheet for the agency for year 2013 was illustrated with a graph, Figure 18.
The overall operation of the analysis unit turned out to be socio-economically beneficial. Annual return on investment (i.e. on total cost budget of the agency) is around three to fourfold. This is the answer to one of the main research questions of article IV (What is the overall B/C ratio of the organization?). By funding an agency that has a positive benefit to cost ratio, a society is investing into positive impacts.

In the light of article IV, traffic safety issues are and will remain in the key focus of such agencies, but the future potential of enhanced/increased positive impacts may lie in other responsibility areas, such as the environment and transport market efficiency. Traffic safety does need continued attention and control; otherwise the currently positive safety trends might turn less positive or even deteriorate. The management of environmental impacts and risks will likely be emphasized in the future, and these aspects will require heightened attention.

In article IV, an additional area of potential benefit generation was discovered, namely the segment of information-intensive services. Willingness to pay for information and data is being witnessed e.g. by commercial companies buying data from the agency. For the users the real value of the data is at least worth its price. Added value and commercially sustainable services are being produced based on the data. Opening and freeing of the data even further – within the constraints of
data privacy and agency’s operational budgetary considerations – would be justifiable on socio-economic grounds.

To summarize article IV contribution, it covered the sample validity analysis, the analysis of validity of the results from sample services’ impact evaluations; scaling-up of the sample results to reflect the agency as a whole; and the deduction of the agency’s efficiency and socio-economic balance sheet.

Article IV mainly addressed research questions 3 and 4 and their sub-questions, additionally supporting research question 2 to some degree.

### 3.5 Results synthesis

Based on the above, Figure 19 summarizes the individual contributions of the original publications and in Table 10, the responses to the research questions and the research problem (see sub-chapter 1.2 and Table 1) are presented.
### Table 10. Research questions and responses to them.

<table>
<thead>
<tr>
<th>RQ#</th>
<th>Research Questions (RQ)</th>
<th>Article</th>
<th>Main results</th>
</tr>
</thead>
<tbody>
<tr>
<td>RQ1</td>
<td>What is the performance management architecture?</td>
<td>Al</td>
<td>Studied and described in article I</td>
</tr>
<tr>
<td></td>
<td>…the operational environment and the national performance management system?</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>How is the study subject’s performance managed as part of the national system?</td>
<td>Al</td>
<td>Studied and described in article I</td>
</tr>
<tr>
<td></td>
<td>How effective and comprehensive is the study subject’s performance management?</td>
<td>Al</td>
<td>Overall reasonably good coverage. Severe and potentially compromising/risky blind spots were discovered however. The coverage can be improved with the help of this analysis and the presented recommendations. Similar findings are estimated to be highly typical for equivalent organizations elsewhere.</td>
</tr>
<tr>
<td>RQ2</td>
<td>With which methods can socio-economic impacts of a service oriented public organization be investigated and quantified?</td>
<td>A II, III, IV</td>
<td>The research design and methodology, as described in articles II, III and, IV, and as summarized in this dissertation.</td>
</tr>
<tr>
<td>RQ3</td>
<td>What are the B/C ratios for individual services?</td>
<td>A III, IV</td>
<td>Information intensive services were discovered to be exceptionally promising in this respect. Also e.g. driver education systems and control regimes have a fair B/C ratio. Ratios between ≥1x to 20 were recorded.</td>
</tr>
<tr>
<td></td>
<td>Which services or which type of services yield positive B/C ratios, and what are these ratios?</td>
<td>A III, IV</td>
<td>None were detected (but for many services deducing an unequivocal B/C ratio turned out impossible)</td>
</tr>
<tr>
<td>RQ4</td>
<td>What is the overall performance?</td>
<td>A IV, this thesis</td>
<td>Yes it is, as presented in this research, including the original publications.</td>
</tr>
<tr>
<td></td>
<td>Is it possible to construct a research process for deducing a B/C ratio for a public agency?</td>
<td>A IV, this thesis</td>
<td>The overall B/C ratio (or annual incremental net benefit ratio) was calculated to be from 2.6 (current probable minimum) to 4.3 (current probable maximum).</td>
</tr>
<tr>
<td></td>
<td>…What is the overall B/C ratio of the agency?</td>
<td>A IV, this thesis</td>
<td></td>
</tr>
</tbody>
</table>
As a response to the main research problem of this work (“The current understanding and methods for deducing the overall performance of a public agency are insufficient. Access to sound and transparent socio-economic performance metrics is a prerequisite for efficient performance management. Further research is needed in order to improve the understanding of and the methods for identifying these metrics.”), the results presented above, and especially the results for research questions RQ4.1 and RQ4.2, clearly imply that the overall performance of a public agency can be assessed.

To the author’s knowledge, an assessment of socio-economic impacts and efficiency of a service oriented public agency has not previously been performed this comprehensively and with this level of depth and detail. Although much of the present research has relied on earlier studies and literature, the designed and presented process for the evaluation entity is novel and unique. The presented set of methods proved functional in determining the impacts and the cost efficiency of individual services and of the entire service portfolio. Furthermore, the process as a whole turned out to be capable of quantifying these impacts on socio-economic monetary terms, as was presented. Therefore, and as another principal response to the targeted research problem, it can be claimed that this research has contributed to the understanding and methodology of socio-economic evaluation.

On a more tangible level this research has shown that the operation of a public agency can be socio-economically beneficial. Return per annum on investment (on the total cost budget of the agency) with regard to the inspected agency was observed to be around three to fourfold. By funding such an agency, the society is paying for positive impacts that are worth more than the investments. The impacts may occur in a multitude of forms. With regard to the transportation sector, for the citizens the impacts may manifest as increased well-being and health, increased free time and prolonged life expectancy. For businesses and other communities they may emerge as enhanced competitiveness, more resilient logistic chains and increased transport market efficiency due to common rules of conduct and equality in market regulation.

An agency that is well managed and therefore performs as intended is essentially an apparatus that efficiently produces wellbeing and prosperity. Further pursuits of constantly fine tuning the performance of any agency in order to achieve maximum efficiency are clearly to be encouraged. According to these findings this can be done, for example, by introducing or enhancing the application of NPM and MBO. But also in light of these results, increasing efficiency by trying to run such agencies with fewer and fewer resources is in fact a scenario that may risk wasting
those resources rather than saving them. A limited amount of public resources
seems a poor justification – or at least a very short-sighted justification – for budget
cuts whenever those cuts are targeting de facto investment multipliers.

However, a singular agency assessment, such as presented here, provides only
limited data and support for the governing body and for the entirety of government
decision making. The government can only really fully compare the overall socio-
economic efficiency of public administration in a situation where all public entities
are assessed. Partial optimization can be carried out when all units within the same
jurisdiction (e.g. all agencies under MoTC) are assessed. Even then the question
would remain, whether or not the distribution of funding between the ministries is
optimal. Therefore, extending similar assessments to cover singular jurisdiction
areas (e.g., transportation) and other areas of jurisdiction would provide more
perspective and decision support in the quest for seeking the overall socio-
economic optimum.
4 Discussion

4.1 Theoretical implications

This sub-chapter summarizes the key contributions of this research work to the relevant literature. In addition to addressing the set research questions as presented in sub-chapter 3.5, the dissertation contributes to eight themes in the impact evaluation research stream within the fields of (i) the tradition and maturity of impact evaluation research, (ii) policy design and implementation, (iii) the coverage of performance management indicators, (iv) the comparison between performance management indicators and impact evaluation indicators, (v) the impact of enforcement, (vi) the goodness of fit of NPM, (vii) impacts’ spill-over effects, and (viii) contribution to the impact evaluation methodology.

The tradition and maturity of impact evaluation research

The first key observation is not specific to this study but is more general in nature and therefore applies universally across this field of research. In light of this study it seems that within scientific research, the evaluation and scrutiny that cover public entities’ full spectrum of socio-economic impacts are still very limited both in volume and with regard to methodology. The pressures to deepen the knowledge and practices therein have been and are evident (Papadimitriou and Yannis 2014; Howlett et al. 2014; United Nations 2010; National Audit Office 2001), and advances are being made (Saetren 2014; Leviäkangas and Hautala 2009). Regardless of the present findings and contribution, we are still a long way from a comprehensive set of methods, tools and data for the analysis. Furthermore, the impact evaluation of more recently emerged responsibilities (such as environmental impacts or transportation system resiliency) proved to be less mature and also less in quantity and tradition.

Building on earlier research within the field of general impact analysis (e.g. Kohta 2007; Hytti and Mäki 2008; Kuitunen and Hyytinen 2004; Guerra 2007), the general literature on impact evaluation processes (OECD 2017; Storey 2000; Baker 2000; Guerra 2007; Holden and Zimmerman 2009; Kuitunen and Hyytinen 2004) the literature on indicators and metrics (e.g. Lähteenmäki-Smith and Hyytinen 2006, Rajahonka 2014; Aistrich 2006) and the literature on CBA (e.g. Mishan and Quah 2007; Boardman et al. 2006; Hautala and Leviäkangas 2007; Gramlich 1998;
Layard and Glaister 1994; Brent 1996), the present research adds to the earlier knowledge and methods on impact evaluation research. This has been shown mainly by presenting one viable approach and a set of methods to carry out evaluation that covers the full spectrum of socio-economic impacts of a public entity, hence addressing the knowledge gap.

In all, the other implication to the field of science is that plenty of research has yet to be carried out and that the existing data still is more or less sporadic. Since this is the case, and the steering and control of administration cannot always be justified with established research, the following question emerges: how is the steering and control currently being determined and justified in the absence of data? This is discussed in the following section ‘Policy design and implementation with limited data’.

**Policy design and implementation (with limited data)**

In terms of background, history, reasoning and justification, each single piece of regulation and public service always has a distinct and often complex story behind it.

On one hand, some of the services may have been imposed on national agencies by means of, for example, international regulation and agreements. As shown here, the national government may often be represented in some way within the regulating bodies and thus the nation has a voice in whatever is being decided upon, at any given time. But more often than not – especially with relatively minor nations (i.e., in world-economic terms less substantial and otherwise minor nations, such as, e.g., Finland) – the power of this voice is not the ruling one. In any case, whatever the voting power and origins, upon joining the consensus and signing the convention or regulation, the nation is de facto agreeing and therefore obliged to conform by implementing the imposed services, control regimes and frequently also the associated adaptation measures of national legislation.

On the other hand, the regulation or services may have more national or internal origins. Within the transportation sector, it was observed that some transport modes are more heavily internationally regulated than others. From the viewpoint of national levels of freedom to govern, aviation and maritime transport are the most ‘set in stone’ – and rail transport is not far behind. Whereas in aviation there is practically no leeway at all for differences between countries, at the same time road transport is more in the hands of national regulation and legislation, even if some principles (especially with regard to vehicle technology and rules of the road) are
being defined internationally, as shown in this research and also elsewhere (e.g. Buigues and Sapir 1993a and 1993b). And at the farthest end of the spectrum, light traffic (vulnerable road users; walking and cycling) has no international regulation to speak of. This was also witnessed in this research.

Why is this so? Why are there such big differences in levels of regulation? To aid in understanding (without going into any detail or backing up the argumentation with comprehensive scientific evidence), some conceptual logics may be constructed here to at least partially explain the reasons for this being the case historically. Firstly, the transport modes within the strictest global regulatory grasp (Buigues and Sapir 1993a and 1993b) are also the most ‘international’ ones. One single long-haul flight can involve a dozen nations within a matter of hours, whereas the most moderately regulated transport mode of light traffic is best described as being not even national, but local. Secondly, the more technologically complex the transport mode (i.e., the related vehicles and infrastructure), the more stringent the international control. These characteristics hopefully go some way in accounting for the diverse levels of regulation in various transport modes. In this study, the relevance of these characteristics relates to the degrees of freedom that national agencies have in influencing the future development of regulation. That is, with the modes under the strictest international grasp, the freedom is limited to having one voice among other countries, whereas with, for example, non-motorized transport there is much more national leeway.

In any case, regardless of the mode and history and origin and also regardless of the amount of coherent and transparent scientific research to back up the decisions involved, all decision making processes with regard to implementing and deploying new administrative control and services always have an element of intuitive decision making in them. ‘Intuitive decision making’ here means that in addition to the available hard facts, the decision making mechanisms are always affected by, for example, the decision makers’ experience, perceptions, associations, preferences, expertise and idiosyncrasies (see e.g. Sedlmeier 2015; Belton and Stewart 2002; Considine 2012). This intuition may be more or less collective in nature. In other words, some decisions are made by smaller and more national groups (e.g., typically national regulations) whereas others are made via the consensus of a larger and/or more international faction.

Whenever the facts and figures to back up the decision are scarce, the role of this intuition element is prone to come more into play. Several observation units of this research (i.e., services that were investigated) provide valid examples of this phenomenon. In more than one instance the literature search revealed, that even
when the core object/target of the analysis and its impacts may have been studied comprehensively, the sovereign impact of the administrative control of conformity may have received little or no attention per se. An even more tangible example may be in order here. According to the literature, the technical maintenance of railway rolling stock (including locomotives) and infrastructure and its relevance to the railway transport safety level seems to be reasonably thoroughly researched. The same applies to the transport of dangerous goods on rail, in relation to both safety and the environment (e.g. Saat and Barkam 2011; Cafiso et al. 2006). The transport safety agencies are mandated and obliged to enforce that the operative and maintenance organizations’ practices, expertise, technological readiness, human resources etc. conform to the relevant international and national regulations. But according to the findings of this study, the sovereign impact of that control with regard to the transport safety or to the environment remains to be researched in depth. (The particular literature survey for this part covered, but was not limited to, e.g., all of the articles of the Journal of Rail Transport Planning and Management from years 2011-2014.) Despite this, the public control services are there – and there is no reason to doubt that they would also remain operational in the foreseeable future. Clearly the decision paths that originally led to the introduction of such services have had to include some intuition. New public control regimes are – out of necessity – sometimes set based more on collective intuition than on acknowledged and direct research evidence on their impacts.

Therefore the empirical findings presented here resonate with earlier notions (e.g. Considine 2014) of the role of intuition in policy design and implementation. As a logical consequence from the absence of hard research facts, new public control regimes are sometimes set based on collective intuition. The present observations suggest that this collective intuition usually proves to be right, and that the input generally generates net benefits. But identifying the most productive use of public resources requires relentless further research – or the chart and compass are not there. This in turn has implications into how the public agencies are managed.

The coverage of performance management indicators

The present research was able to identify that an agency’s performance management systems may comprise valid indicators per se, but mostly as a consequence of a limited amount of research evidence the coverage may not be
ideal. The selected metrics may need further consideration whenever new research data on the actual impacts emerge.

Furthermore, this research showed that some performance management indicators may include metrics that are affected by numerous phenomena external to the object that is being managed (Mononen and Leviäkangas 2015). The challenge in such a case is how to isolate the impacts induced solely by an agency’s actions from other changes in its field of operations. For example, a typical metric in any transport safety agency’s performance management contract might include elements from statistical trends in transport safety, such as casualties on roads. But as shown here, those trends are in fact affected by numerous factors (developments in vehicles’ technical safety, macro-economic fluctuations’ effects on traffic volumes, enforcement, infrastructure – just to name a few), many of which are bound to be external to the jurisdiction of the agency being managed. A full 360 degree coverage of metrics is neither a practical, nor an advisable target but at the very least, exogenous factors (i.e., affecting from outside the agency’s mandate) should be excluded from the equation when attempting to measure the sovereign impacts of an agency’s functions. (This holds true, regardless whether the measurement attempt relates to performance management or impact evaluation.)

This thinking was applied throughout the research described in this study, and is presented as one of the main theoretical implications for future similar studies. Similar cross-checks of coverage of performance management indicators’ coverage could not be found elsewhere in the earlier research so the finding is a novel contribution.

**Performance management indicators vs. impact evaluation indicators**

From the onset, the most recent Finnish national government program was observed to highlight the importance of both setting performance targets and following the success with regard to the targets (Finnish Government 2015). Each main chapter within the program document commences with a sub-chapter titled ‘The objectives and how to measure them’. This in fact, is the essential concept of Management by Objective, and is not a novel concept within the nation, as discussed earlier. It has been applied as the main management principle in governing the entire national public sector in several preceding governments and their programs. Within MBO, another more recently emerged aspiration has been to complement the qualitative objectives and move more towards measuring the socio-economic impacts of public agencies’ activity (Finnish National Audit Office
This in fact was one of the most focal motivators of this research. But as discussed in this study, concrete metrics and indicators to facilitate quantifying the impacts have been lacking. They will unfortunately remain so also after this research, but some steps in the desired direction have been taken.

On first glance, the performance management contracts in various sectors of government generally contain valid sets of metrics and indicators per se, that is, the metrics relate directly to the sector being managed. For example, the general trends of traffic safety undeniably and obviously relate to the management of a traffic safety agency. On closer inspection, however, it may turn out that the indicators include some fuzziness and ambiguity, as was also shown in this research. For instance, the mentioned safety trends (with regard to any given time frame of a performance contract) as such actually tell very little about the real impact generation, in other words the sovereign contribution of the agency under scrutiny to the changes in the trends. In addition to the agencies’ contribution, in reality these kinds of trends may be influenced by many externalities such as changes in the economy or technology development, typically. It would be safe to even argue that such trends are always dictated by a complex set of factors, and never by the actions of one single entity.

This finding also implies that the activity of an agency may be beneficial (i.e., producing positive net impacts) even though the relevant general trends would be negative. And vice versa: inefficient public services may be able to ‘fly under the radar’ and go unnoticed and unadjusted if the sector-specific trends happen to show overall positive development. Because of this, it is of paramount importance to be able to isolate the sovereign impacts that are induced directly by the public actors’ activities and separate them from the other changes occurring within the sector or in the surrounding operational environment. Otherwise the applied performance management metrics may be yielded less meaningful, at least as tools for the correct steering and control of the agency being managed. All this is assuming that true MBO remains the desired and chosen management model to be applied.

The dilemma discussed above that was discovered, is not specific only to the management of transport agencies or to governing public agencies in any sector. It is generalizable to all levels of government, all the way up to the government policy choices within the programs for successive electoral terms. Many of today’s problems that programs are intended to tackle may be fully or partially solved – or worsened – over time, depending, for example, on the fluctuations of the world economy. In the most extreme cases, some changes in the operational environment
may even transfer today’s challenges from one sector of the government to another one in the future. For instance, in an imaginary but possible future scenario where a full-scale electrification of transportation – on the road or beyond – has taken place, many of the current environmental challenges will also have gradually shifted from the transportation sector to elsewhere. In that new scenario, the need to optimize and minimize overall energy consumption would still be relevant to the transportation sector, especially including the automotive industry, and further down the road maybe even also the vessel, aircraft and locomotive industries. But the challenges with regard to GHG and particle emissions would then have transferred to the energy sector. The focus there would probably then be (as it is already now) to determine with which technologies the bulk of electricity is being produced with – including, but not limited to the electricity used by transportation. (Another similar example could be a transport agency promoting active travel to reduce obesity where the benefits accrue to the health ministry because of better outcomes for health.)

In fact, from the points of perspectives of optimizing the societal economy and maximizing the net positive impacts, this particular example of a potential shift of paradigm is actually irrelevant. The challenges would still be there although the responsibilities would have been reallocated between sectors. But when the lens is on the steering and control of the various elements of the public economy, such as transportation agencies, then these kinds of subversive and disruptive evolutions are in fact highly relevant – and ideally they should be accounted for in a proactive manner. One solution to this (or recommendation) would be to move from performance management to the management of impacts. Granted that this is an ambitious objective, the author recommends highlighting performance metrics that have shown their connection to a capability of generating impacts, such as in studies like this one (this instead of including metrics that are admittedly measureable but do not necessarily tell very much about the impacts of the agency being managed.) For this, open-minded development of public services and processes would be required to succeed in starting to produce positive impacts instead of outputs and indicators. This calls for responsible and skilled governance, since the more strategic the management of agencies becomes, the more carefully the operational management and service production has to be performed.

The mechanisms described above bear relevance to the performance management of public bodies – but, very importantly, also to evaluating the impacts of public bodies. As mentioned earlier in sub-chapter 1.4, in this research the underlying principle was at all times to try to evade the bias and distortion caused
by the changes in the surrounding operational environment. The adopted approach
does restrict the levels of freedom in the research but more importantly, at the same
time it isolates the impacts induced by the public actor and separates them from
other changes within the field of observation. The selected lens therefore was that
of microscopic measurement and evaluation. The current or emergent trends in
traffic safety, energy consumption, the economy, and emissions — and the plethora
of external variables affecting them — played a diminutive role, if any, in this
research. Instead, the adopted approach focused on evaluating the observable
sovereign impacts induced by individual public services. The final conclusions and
the deduction of the macro-level implications rely on the detailed analysis of
individual services that are clustered into service families covering the full range
of the case study subject’s service architecture. The present findings and the
approach highlight the importance of identifying and selecting the impact
evaluation’s indicators carefully. The indicators should describe the object
efficiently, precisely, adequately, reliably and repeatably, and they should measure
the results and impact rather than activity and objectives. Thereby this research
mirrors and complements the current knowledge on indicator selection and
definition (e.g. Aistrich 2006; Karjalainen 2007; Guerra 2007; Haapanen and
Mustonen 2006). The approach itself and its justification also constitute a
methodological contribution that similar future studies should consider in their
research design.

The impact of enforcement itself

In reality the processes leading to new regulations and enforcement are highly
complex and the abovementioned intuition element alone is not enough to explain
them comprehensively. Nevertheless, some analogies might be drawn upon to
explain why the decisions to introduce new services and public control mechanisms
may have been logical, meaningful and acceptable after all — even without
established direct research to back them up. Namely, if conformity to regulation is
not being systematically controlled, the entire purpose of regulation is defeated and
it ceases to yield positive impacts. This is admittedly a relatively bold claim that
deserves some discussion to open up the reasoning behind it, albeit it aligns with
findings from earlier research (e.g. Montella et al. 2014; Ferguson 2012).

In the analysis of the impact mechanisms of the units of observation (i.e.
services) in this research, often the primary direct impacts were induced by agency
interventions that led to tangible changes on the field. This was especially so with
regard to services that involved control, auditing and inspection routines. The changes could involve practically any part of the transportation system, including the infrastructure, vehicles, other equipment, safety management systems, personnel, practices, documentation, etc. In addition to this direct mechanism, a less direct secondary impact mechanism was also recognized, one that stems from the existence of the inspection systems itself. Namely, the awareness of a chance of any of the above listed parts of the transportation system undergoing inspections at any time presents a de facto deterrent for operators and other actors. In the most severe cases (upon discovering severely hazardous discrepancies) a vessel, aircraft, etc. may be put into detention, meaning it will not be allowed to depart before the discovered shortcomings have been rectified. Depending on the faulty item and the severity of the fault, it may sometimes take a long time to lift the detention. Therefore, to avoid time losses and the associated financial loss, it is clearly within the operators’ interests to make sure that their equipment (and staff, documentation, etc.) always conforms to existing regulations as well as possible. In addition to saving quantifiable accident or environment related direct costs, avoiding accidents in any transport mode naturally has ethical implications as well. A similar logic for indirect impacts applies, and is generalizable, across all modes, such as road vehicle inspections, aviation inspections, etc. For example, entire airlines – and even all airlines from designated nations – have been temporarily quarantined entering European airspace, on the grounds of agency inspections having revealed reoccurring and serious non-compliance and discrepancies in international aviation inspections (Wall, 2011). This secondary impact mechanism underscores one major generalizable observed characteristic of agencies – and particularly the impacts they induce via means of enforcement and control. That is, the purpose and outputs of any regulation are diluted in the absence of systematic enforcement.

Direct evidence of this was uncovered with regard to some of the units of observation of this research – but not all of them. However, analogous evidence for this exists in abundance elsewhere – from across practically any sector of regulation and government, including the transportation sector. How well would the speed or intoxication limits be obeyed if they were not being controlled systematically? Not very well at all, as has been shown in a wealth of established research, a considerable bulk of it quantitative in nature. As presented, evidence on this can be found to some extent within the realm of public service provision in the transport sector (e.g. Knapp and Franses 2006 and 2010) and more from within the general transport sector (e.g. Montella et. al 2014; Ferguson 2012) – and even more when looking at many other sectors of society.
Unfortunately, the linkage between these earlier studies (and alike) and that of the type of public enforcement discussed in this research is hard to establish – at least in quantitative terms. On a conceptual level, the control regimes are highly similar. In all cases, a generic description can be formulated: a public actor (the police, an agency, etc.) controls how well the real world conforms to the regulations, and acts upon the findings thereof by issuing warnings, fines, penalties, detentions, orders to fix, and so forth. Observed through a qualitative lens the writing is on the wall. Effectively, any public agency that provides conformity control services is acting as the enforcer (or as the ‘police’) of the regulation where non-compliance is perhaps considered legally less criminal or punishable when compared to what the police have been assigned to enforce. When the ‘criminality’ or the relative importance of a regulation is considered to be on the higher end of the spectrum, then it normally is the police (or customs, border control, etc.) that are put in charge of the enforcement. When the importance or severity is perceived as less critical it is then not the law enforcement sector but some other public agency that takes care of the control duties. But nonetheless, there are no grounds to directly generalize or transfer such findings, such as those of Montella et al. (2014) or Ferguson (2012), quantitatively to the research at hand.

The theoretical implications of the occasional absence of direct quantitative research on the sovereign impacts of agency enforcement’s impacts, and also the lack of linkage to the abovementioned analogues, are twofold. (i) The unequivocal positive benefit to cost ratio of such a service cannot be determined analytically. But simultaneously, (ii) demonstrating that such a service produces neutral or negative net impacts is similarly impossible. For example, although the impacts of controlling the conformity of hazardous material transport on railways have not been studied directly, there are no data to support an argument that operating the control service would be unnecessary, redundant or pointless.

The findings of this research suggest that the current policy design and policy implementation mechanisms work adequately, in the sense that the public resource input is not where it would be generating net losses. But as to whether or not the public resources are being used as productively as possible, that remains anyone’s guess without thorough impact analysis and valid data to support it.
Goodness of fit of NPM

In regard to the theoretical discussion around goodness or badness of fit of business-related approach to running public actors, one could ask: “Could public bodies be run as businesses are? And if they could, should they be?”

In light of this study, some analogies between businesses and public administrations can be constructed. This is in a way one prerequisite for adopting management principles originating from the private sector and being able to apply them successfully within the public domain, for example, in the way that is being attempted in exercising NPM. But nevertheless, some discussion on the underlying analogies and divergences between these two sectors could shed some further light onto the challenges and thinking processes that were faced during this research work.

At first it might appear somewhat far-fetched to apply any business-related approach for investigating or describing an administration. One might argue that from the very onset, a business exists for doing business, whereas administration exists for the purpose of ‘the general good’ or because of some type of generic societal need. But the horizon could be broadened by looking at the analogy in the following way. (See Table 11.) Instead of the more simplistic differentiation between businesses’ and public agencies’ reasons for existence, the use of the word business could be observed as a broader concept and looked at in a wider scope, in comparison to the context the word is more generally used in.
Table 11. Some analogues and commonalities between businesses and public bodies.

<table>
<thead>
<tr>
<th>Perspective</th>
<th>Businesses</th>
<th>Public bodies</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reason to exist is...</td>
<td>to ‘make business’</td>
<td>for ‘the general good’</td>
</tr>
<tr>
<td>Objective is...</td>
<td>financial gain for owners</td>
<td>added value to the society</td>
</tr>
<tr>
<td>Output / products are...</td>
<td>things, services, IT...</td>
<td>services &amp; infrastructure (education, health care, safety audits, roads, schools,...)</td>
</tr>
<tr>
<td>Accountability is towards...</td>
<td>the owners</td>
<td>the ‘owners’ = the people, the nation</td>
</tr>
<tr>
<td>Oversight (representative of owner) taken care by...</td>
<td>the board</td>
<td>ministry, the government</td>
</tr>
<tr>
<td>Vertical accountability shown with...</td>
<td>annual reports, finances</td>
<td>performance contract audits / impact evaluations</td>
</tr>
</tbody>
</table>

The more or less deep down underlying business of a commercial enterprise is in the end always to generate financial gain for its owners. Similarly, the business of a public administration is to produce added value to some part of the society or to the society at large. This added value should emerge as improved economic, social and/or humanitarian well-being created for the society (the ‘people’) – and the fruits thereof should be catered for the benefit of the masses that are choosing to support and fund the agency’s activities. Some types of added value would quite clearly be easier to quantify in monetary terms than others – this has also been witnessed in this research work overall. Added value to economic well-being is as such already on a monetary or financial canvas on a conceptual level, whereas both social and humanitarian well-being are generally more challenging to put a price-tag on.

Analogues between businesses and public bodies can be observed on other levels as well. For example whereas a business always has products (physical, IT or service products), similarly a public organization has products as well. These can manifest themselves as physical deliveries (e.g., infrastructure like core electric network, water and sewage networks, airports, harbors, roads, etc.) or as service provision (e.g., health care, education, transportation safety inspection/audit, etc.). Or they can materialize as less tangible products such as generally providing a functioning societal framework with the supporting legislation, enforcement, security, etc. so that the citizens – and businesses alike – have ample opportunity to live and prosper. Similarly, whereas businesses have owners, in fact so do...
agencies – the de facto owner is the society, in other words, an agency is held accountable to the state including its citizens, government and businesses. The evaluation result of vertical accountability, as presented in this research, could therefore be seen as the public realm equivalent of profit announcements (or warnings) from a company to its stockholders.

As implied, the owners of conventional businesses are often easily recognizable by looking at the jurisdiction documentation, whether dealing with sole proprietorships, partnerships, corporations or other types of ownerships. The owners of public entities are the people or the society in general. For efficient day-to-day management of the company’s financial resources, capital or tangible resources, and human resources, instead of all owners taking part into each individual decision there usually needs to be some type of representation structure in place. (Very small companies are an exception to this.) In companies the owners’ interests are typically looked after by a management board, board of directors, or the equivalent. With public entities such as agencies, the owners (the people) are represented by, for example, a ministry, in addition to the agencies’ internal management mechanisms.

Typically market-driven businesses have fixed (and to some extent legislated) cycles and established processes in place for evaluating their performance levels, namely the annual (and for some companies quarterly) financial performance reporting and financial statement cycles. In addition to this, separate evaluations can be and are performed for businesses as a whole or to some selected elements of enterprises, with the objective of enhancing the performance locally or globally. On an analogue level, somewhat corresponding annual reviews also take place in some instances for public bodies in the form of control of the fulfilment level of performance contract objectives – that is, performance contracts between an agency and its overseer, such as the relevant ministry or some other governmental body.

However, some of the abovementioned analogues between businesses and public bodies could be looked at from an opposite perspective, namely from that of the differences between these two realms. Table 12 illustrates this perspective.
Table 12. Some divergences between businesses and public bodies.

<table>
<thead>
<tr>
<th>Perspective</th>
<th>Businesses</th>
<th>Public bodies</th>
</tr>
</thead>
<tbody>
<tr>
<td>Impact of products</td>
<td>sales revenue minus production, delivery and other logistic costs</td>
<td>socio-economic impacts minus production and delivery costs</td>
</tr>
<tr>
<td>Impact distribution…</td>
<td>goes according to the ownership</td>
<td>should be distributed ‘fairly’</td>
</tr>
<tr>
<td>Product portfolio renewal…</td>
<td>based on demand, unprofitable</td>
<td>largely comes as a given (international commitments etc.), outdated products die slowly</td>
</tr>
<tr>
<td>Product development speed is…</td>
<td>flexible and agile</td>
<td>slow, a lot of inertia in renewing legislation and regulation</td>
</tr>
<tr>
<td>Stability of objective sets is…</td>
<td>dictated by the ability to generate profits</td>
<td>dictated by the political agendas of the most recently elected government</td>
</tr>
</tbody>
</table>

From this perspective the key observation is the flexibility of the business sector in changing its operations as opposed to the public sector’s rigidity in the same respect. Public bodies should exercise some societal fairness in distributing the gains of the operations whereas in business, the ownership dictates to whom the spoils go. The development of the offering is free and therefore agile in commercial settings, whereas the public sector suffers from some inertia in the renewal process – including, but not limited to the historical build-up of duties and services, as discussed for example in article II. The development direction of, for example, public service production depends not just on the impact or revenue generation potential but also on the political agendas of the current government.

All this said, it can be concluded that judging by the observable commonalities, private sector management principles can be applied to the public sector – but by the same token, the observable divergences between the two realms illustrate that these principles cannot be applied in a fully straightforward manner, as this research has also shown.

In summary, based on the divergences between the two realms, public bodies cannot and should not be run exactly as businesses are. The missions are different and, for example, the steering mechanisms and the product portfolio renewal are prone to remain much less agile in the public sector. Some of the products will even remain ‘untouchable’, at least as long as the legislation is as it is and the nation is committed to the international agreements and directives dictating it. These observations as such therefore confirm and support the views of the sceptics and/or
critics (such as Christensen 2007; Romzek 2000; Hood 1991; Hood and Peters 2004) of administrative reform toward a more business oriented management approach.

But by the same token, adopting some practices from the private sector (as has been done), is a development to be further encouraged, since the public sector organizations need to be increasingly accountable for their actions to the owners and they need to show the most efficient use of public resources as possible. This observation, in turn, speaks partially against the claims of the abovementioned sceptics and simultaneously supports and reinforces the views of the champions of NPM and also the OECD (2010) recommendations for continuous change.

As a whole, building on preceding studies, the present research adds to the earlier knowledge on the objectives and goodness of fit of neoliberalism and NPM (e.g. Bessant et al. 2015; Gruening 2001; Christensen 2007; Romzek 2000; Hood 1991; Hood and Peters 2004; Maugeri and Metzker 2013), performance management in public domain (e.g. Bessant 2015, Hood 1991, Hood and Peters 2004; Binderkrantz 2011; Bindercranz and Cristensen 2009), and management by objective (e.g. Hood 2007; Gruening 2001; Barberis 1998).

**The spill-over effects**

To be complete the scope of an impact evaluation study should not be constrained to the research subject’s internal productivity. As presented in this research, ideally it also should encompass the perspectives of the public economy (which includes the research subject + other public entities), national economy (which includes the public economy + citizens + companies) and societal economy (which includes the national economy + external/international economic ties). (See Figure 20.)

The observed impacts of services are distributed throughout societal economy. Such impacts residing within the agency relate mostly to internal productivity, such as the more efficient production of services. Ideally, individual services or the entire service portfolio could be produced and supplied with smaller costs without compromising the quality. Public economy is positively affected whenever the agency can generate cost savings elsewhere within the government.

Impacts to the national economy and societal economy are of the most paramount importance. These include the cost savings with regard to accidents and environmental cost savings, which is where this study was focused.
From the perspective of the entire government, all of the abovementioned levels of impacts are highly meaningful. But in particular, peoples’ well-being is largely measured by socio-economic impacts. An increase of internal efficiency may reward, for example, the agency management but at the same time in the worst case even decrease the level of service (quantity and/or quality) catered for the general public. This is in fact why this research focused on evaluating socio-economic impacts – without playing down the other impact type categories, however.

The key rule to be kept in mind when evaluating socio-economic impacts, is that a service is socio-economically beneficial only if it benefits someone without at the same time putting others at a greater disadvantage (Pareto optimum; see e.g. Brent 1996; Gramlich 1998). According to the strictest possible interpretation, if a service is mandatory and does not have optional or competing suppliers, covering its production costs fully with user charges is not as such an indicator of socio-economically efficient policy, although it may be fiscally efficient from the state’s point of view. But if the users benefit from using the service more than they pay for it and, what the service production costs are, then the service is socio-economically sound and beneficial. The same applies even if the user is free to select where he

Fig. 20. The impacts’ spill-over effects – from internal productivity to societal economy (Modified from Article III).
or she buys the service from. The purchase choice would then be based more on perceived benefits, rather than on compulsion.

When benefits, savings or cost coverages are merely redistributions of income from one party to another, the socio-economical profitability becomes a question mark – but notably, the related services are not necessarily unprofitable either. In this research the distribution of benefits were not analyzed quantitatively, that is, no valid data were available to enable calculating what proportion of the observed benefits is the share of each beneficiary type or layer.

Similarly, since a cross-section of time approach was assumed, the impact distribution over time was not addressed. Namely, some benefits are not necessarily realized during the year, or only during the year in which the costs occur. Here the impacts were observed and reported on an annual level, although they might keep reproducing during the following years as well. This presents one limitation to the study.

All of the presented distribution effects however, have a bearing on the re-assessment of an agency’s goal setting and performance management system.

Contribution to the methodology

The final, and perhaps the most important theoretical implication of this research is its methodological contribution, that is, its added-value to the impact evaluation methodology itself. As such, this research builds upon, and adds to the earlier literature on impact evaluation processes and practices (e.g. OECD 2017; Storey 2000; Baker 2000; Guerra 2007; Holden and Zimmerman 2009; Kuitunen and Hyytinen 2004; Leviäkangas and Hautala 2007). The work and the processes presented here can potentially be applied elsewhere as they stand, or as modified versions, or by using isolated sections of this research as support in similar studies in their various phases and processes. The transferability of the process and the methodology are discussed in more detail in sub-chapter 4.3, under section ‘Limitations of the study and transferability’.

Table 13 below summarizes the theoretical implications of this study.
Table 13. Theoretical implications.

<table>
<thead>
<tr>
<th>Implication</th>
<th>Required or possible actions/outcomes, comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>Evaluation and scrutiny of public socio-economic impacts is limited; maturity of research tradition</td>
<td>Further research required to gain a more comprehensive set of methods, tools and data for the analysis</td>
</tr>
<tr>
<td>The validity of indicators and metrics in public sector performance management</td>
<td>Selectable metrics need further consideration and research to increase justifiability and effectiveness</td>
</tr>
<tr>
<td>Indicators include metrics affected by external phenomena</td>
<td>Exogenous factors should be excluded from the performance management systems, and their study</td>
</tr>
<tr>
<td>Each piece of regulation has a distinct background, history, reasoning and justification behind it</td>
<td>The background should be acknowledged, since it may have numerous implications (e.g. to the power of being able to alter or abandon the regulation)</td>
</tr>
<tr>
<td>Intuitive decision making to be acknowledged</td>
<td>Public control regimes are sometimes set more based on intuition rather than on research evidence on their impacts</td>
</tr>
<tr>
<td>Controlling conformity to regulation directly and indirectly ensures the regulations’ potential to yield positive impacts,</td>
<td>However, more data and research often needed to show whether or not the resources are in the most productive use possible</td>
</tr>
<tr>
<td>Public bodies cannot be managed exactly like businesses are managed</td>
<td>However, adopting some practices from the private sector is to be encouraged, in order for the public administration to maintain and strengthen their accountability towards the citizens</td>
</tr>
<tr>
<td>Shift of responsibilities and external variables</td>
<td>As discussed in 4.1.4: The re-design or fine tuning efforts of performance management systems should duly note the changes in the surrounding operational environment - but an evaluation of the performance must exclude these external variables.</td>
</tr>
<tr>
<td>Spill-over effects</td>
<td>Insufficient data to calculate the proportion of the benefits for each beneficiary.</td>
</tr>
<tr>
<td>Methodological contribution</td>
<td>Further use in similar studies - and beyond</td>
</tr>
</tbody>
</table>

4.2 Practical implications

As the case study subject’s name implies, traffic safety issues are in central focus in the agency’s operations. This will also be the case in the foreseeable future, even though the organization of the Finnish transport administration sector is soon about to change – the tasks and responsibilities will undoubtedly be transferred to new or restructured actors. In any case, some adaptation measures are called for in the total balance of operational focus. Traffic safety needs continued attention and control
on the level that it is at today; otherwise the currently positive safety trends might
turn less positive. This research has implied that the actual drivers for the positive
trends reside mostly within technology development. This was most evident in road
transport, but is probably the case with other transport modes as well. Therefore,
the future potential of increasing the positive impact generation (within the public
domain and with public resources) actually resides in other responsibility areas, not
so much in traffic safety. These other responsibility areas include the environment
in general, resiliency issues and transportation market efficiency. The management
of environmental impacts and risks is likely to be emphasized in the future due to
political and economic pressures in regard to carbon emissions, other greenhouse
gas emissions, and transportation system resilience issues – and as a practical
implication, these aspects deserve more attention as soon as feasible, preferably
immediately.

Information intensive services constitute a considerable part of the case study
subject’s operations. According to the analysis, these services operate with a good
benefit to cost ratio. This segment also presents an area of potential extra benefit
generation. The data are produced (and maintained, updated, distributed, etc.) with
public funding. At the same time, the willingness to pay for the information and
data is being witnessed indisputably. For the users the real value of the data is at
least worth its price, probably often manifold. Added value and commercially
sustainable services are being produced and sold based on the data. Therefore, the
users actually appear on several layers, starting from the primary buyer of the data
(typically a service provider adding value to the data or re-packaging it), all the
way up to the end-users (a private individual, another agency or a company,
typically).

Digitalization is constantly bringing forth new opportunities and also
challenges for information intensive public services. Staying up-to-speed with the
development will require constant cultivation of the services, or even totally new
thinking and approaches. For example, opening and freeing the data even further –
or at least some of it, within the constraints of data privacy – would be justifiable
on socio-economic grounds. This would among other things lead to better
transparency and equality from the perspective of the market and the competition
therein. By default, free data would also be utilized more extensively than the same
data with a price tag on it. On the other hand, when looking from the public
economy perspective or from the agency’s internal economy point of view,
providing data free of cost would be a disadvantage, mirroring against the status
quo. The production costs would still be there but not the incomes. (This relates to
the earlier discussion on the distribution effects of the economic impacts.) Such a shift should however be observed just as an issue, not as a problem: the issue can be easily resolved with compensation from elsewhere, such as within the state budget.

In general, the findings of this study imply that robust, transparent and high-quality administration goes hand in hand with a high quality national transportation system. The quality of the transportation system is in this respect defined by the level of traffic safety, sustainable choices supporting the environment and the transportation system resilience, for example. On the other hand, the quality of government emerges as positive socio-economic impacts via the abovementioned transportation system’s quality indicators. The findings of this study thus support the claim, that there is a connection or a positive correlation between quality of life and quality of government.

International and national regulation (and the increase thereof) on the other hand inevitably creates extra administrative burden on all of the actors, irrespective of the sector, including transportation (OECD 2010). But on the other hand, the net effects of regulation in the end seem to be positive. Finland is among the most regulation-intensive nations within Europe (Litmanen 2004). Nations that economically rely heavily on international trade and that pursue transparency in their administration and economy, seem to benefit from international regulation the most (OECD 2013). Regardless of these abovementioned characteristics, this is however not to imply ’the more regulation, the better’. The direction and control of new and existing regulation need to be done with careful consideration. Also, in the best case scenario, the deregulation or transfer of administrational duties from the public sector to private organizations may result in win-win situations where ideally the public resources can be conserved while the avoidance and/or accident costs diminish and the level of service of the transportation system increases (National Audit Office 2001; Trafi 2014).

As pointed out by this research, transport safety agencies performance management systems may comprise valid indicators per se, but the coverage may not be ideal and the selected metrics may need further consideration. The measurement of more recently emerged responsibilities may be underrepresented and some indicators may include metrics that are affected by numerous phenomena external to the agency that is being managed. Full 360 degree coverage of metrics is neither a practical, nor an advisable target but at the very least, the aspiration to, for example, equally cover all transport modes should be declared. External impacts should be excluded from the equation when attempting to measure the sovereign
impacts of an agency’s functions. This observation has both theoretical and practical implications.

The primarily desired final result and outcome of any impact or efficiency or accountability evaluation exercise – whether it entails evaluating businesses, projects, programs or public agencies – should always in the end be discovering better, or more economical, or more efficient ways of reaching set goals, whatever those might be. The goals depend case by case on the subject of evaluation. This means that the evaluation focus should be at identifying the best practices - and importantly also the ‘not-so-good-practices’ - in regard to the performance of the project, business, program, or whichever other type of subject is being evaluated. The results may be used to improve the performance of the research subject, or optionally the following or neighboring ones. Typically such evaluation initiatives would not generally be justifiable on the grounds of the research subjects later being able to feel better about themselves, but instead rather on the grounds of them being able to later perform better. This also holds true for assessing the vertical accountability of public bodies. As a practical implication, the results of this research are foreseen to facilitate discussion between agencies and overseeing ministries in the joint design of sustainable and balanced operational and strategic targets.

Further, with regard to traffic safety, as discussed earlier, the safety of non-motorized traffic cannot be pinpointed as the responsibility of any one single agency. Instead, the responsibilities are distributed across many actors, such as the FTA, cities, and municipalities. This in fact presents a potential blind spot, which in this study became evident in an indirect manner. Namely, the agency does not have services that directly address non-motorized traffic – and thus such services were also absent from the agency’s service catalogue and the sample that was compiled within this study. However, it is a fact that a substantial proportion of all road traffic deaths are among pedestrians and cyclists (World Health Organisation 2013). Regardless, the governance still does not highlight this fact to the extent it deserves and the decentralized responsibility makes it even more challenging to apply coherent control. As the World Health Organisation (2013) put it, the national and sub-national transport policies still largely neglect pedestrians and cyclists. The observations from this research do not challenge this claim – unfortunately, a shared responsibility seems to be no-one’s responsibility. As a practical implication, a thorough evaluation and restructuring of these responsibilities and their measurable management is therefore recommended. The Ministry of Transport and
Communications is the most apparent and also the best positioned candidate as the organization to take the initiative and lead in this restructuring effort.

A full socio-economic assessment of a single agency, such as presented here, can improve the performance management process of that agency. Practical examples of this are the recommended improvements to the performance agreement. But, still a singular assessment provides only limited data and decision making support to the governing body, and to the entire government. The government can only really fully compare the overall socio-economic efficiency of public administration and adjust it toward the optimum in a situation where all public entities have undergone similar assessments as presented in this dissertation. Partial optimization or sub optimization can be carried out when all units within the same jurisdiction (e.g., ministry, in the case of this study, all units governed by the Ministry of Transport and Communications) have been assessed. Even then the question would remain, as to whether or not the distribution of resources (funding) between the ministries is optimal. Therefore, extending these kinds of assessments, first to cover singular jurisdiction areas (e.g., transportation), and second, other areas of jurisdiction, is recommended. However, since a full socio-economic assessment exercise covering one agency, such as presented here, is both resource and time consuming, this process (or even meaningfully updating it) cannot realistically be integrated into the annual management cycle. Instead, the process could be integrated into a longer cycle, for example into that of the electoral period with lighter mid-term updates.

Table 14 below summarizes the practical implications of this study.
Table 14. Practical implications.

<table>
<thead>
<tr>
<th>Implication</th>
<th>Required or possible actions/outcomes, comments, recommendations</th>
</tr>
</thead>
<tbody>
<tr>
<td>Key drivers for the positive trends in traffic safety reside within technology development</td>
<td>Re-direction and allocation of resources to other, more potential fields of responsibility (more potential in terms of positive impact creation)</td>
</tr>
<tr>
<td>Environmental aspects will need heightened attention in the future and also have potential to increased impacts</td>
<td>Same as above</td>
</tr>
<tr>
<td>Information intensive services have a good benefit to cost ratio and promising impact creation potential</td>
<td>Same as above, additionally consider the practical and economic feasibility of opening and freeing the data further</td>
</tr>
<tr>
<td>Administrative burden from international and national regulation</td>
<td>Control of new regulation done with careful consideration, where applicable also use deregulation or transfer of administrative duties</td>
</tr>
<tr>
<td>Less than ideal performance management indicators</td>
<td>Review and redesign of the performance agreements, based on the recommendations (e.g. remove or reconsider indicators that are largely affected by externalities, include a more balanced coverage of transport modes, etc.)</td>
</tr>
<tr>
<td>Non-motorized traffic neglected nationally (as performance targets)</td>
<td>Same as above, additionally redesign the responsibility allocations and roles in collaboration with other actors within the field of transportation</td>
</tr>
<tr>
<td>Discussion on the need for further evaluation of impacts has been initiated</td>
<td>Continue the discussion between agencies and overseeing ministries in the joint design of sustainable and balanced operational and strategic targets, extend to other fields of administration</td>
</tr>
</tbody>
</table>

The above sub-chapters 4.1 and 4.2 contain some discussion that on the outset seems to have neither direct theoretical nor direct practical implications. Nor does each piece of this discussion directly relate to answering the set research questions of this work. Rather, the above is an account of the author’s observations that surfaced during the research process. These observations do all relate to the surrounding operational environment, including, but not limited to, the public administration, its performance management and to the transportation sector. Therefore the author sees that recording all of these thoughts here may have some added value for further similar research or to general research in other fields dealing with similar challenges. The transferability of the process and the methodology are...
discussed in more detail in sub-chapter 4.3, under the section ‘Limitations of the study and transferability’.

4.3 Reliability, validity and transferability

The research presented here is based on an investigation of an analysis unit via observing a finite sample of the analysis unit’s service repertoire (the observation unit). In the final conclusions these observations were scaled up to give indication of the performance of the entire analysis unit (the agency). Considering the available temporal, human and financial resources this was the way this research was set up. In an ideal situation all of the services of the analysis unit (i.e., the population) would have undergone a full-scale evaluation, instead of a sample thereof. Unfortunately, in science ideal research set-ups are seldom possible but are more often than not constrained by boundaries emerging from the surrounding real world. These constraints also inevitably mirror as constraints with regard to the research results’ validity, transferability and/or generalizability. In the following the account of the concerns of the researcher (e.g., with regard to things that could be done in a superior way in ideal or closer-to-ideal scenarios – in terms of available data and/or resources, for example) are presented and discussed in some detail.

**Measures to ensure reliability and validity of this research**

Reliability and validity are the two main aspects to be controlled while designing research of this kind (Creswell 2009). Construct validity means establishing appropriate research settings for the concepts being studied. Construct validity can be enhanced through three ways: using multiple sources of evidence in the data collection, establishing a chain of evidence, and letting key informants and the research group members review report and article drafts (Yin 2009). External validity is the ability to apply the research results to other contexts. External validity mainly deals with the generalization of the results. However, in research of this kind, generalization may be limited and challenging, and hence the focus is not on generalizing the results but rather on discussing and exploring the studied phenomenon in the selected context. External validity can be increased by defining the research scope and boundaries carefully and comparing the findings and results to previous literature and research. Internal validity considers the research results as less important than the process of discovering them. The credibility of the researcher is evaluated by the people who participate in the research. Thus, internal
validity is achieved when the research participants see the results as believable (Yin 2009; Eisenhardt 1989). Table 15 summarizes the actions and practices that were adopted to increase and ensure the reliability and validity in this research.

Table 15. Measures and practices to ensure reliability and validity.

<table>
<thead>
<tr>
<th>Aspect</th>
<th>Measure</th>
<th>Research phase for the measure</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reliability</td>
<td>The research methodology and design and the data collection methods were carried out and decided by several researchers, including the supervisors. The data, data analysis, and drafts were reviewed and commented on by the case study subject employees, informants and other researchers.</td>
<td>Research design and data collection</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Data analysis</td>
</tr>
<tr>
<td>Construct validity</td>
<td>Used data triangulation. The data, data analysis, and drafts were reviewed and commented on by the case study employees, interviewees and other researchers.</td>
<td>Data collection</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Data analysis</td>
</tr>
<tr>
<td>External validity</td>
<td>Defined the scope and boundaries of the research clearly (case company, data collection methods, the inclusion and exclusion of services to be investigated, etc.). Compared the results with the current literature and research.</td>
<td>Research design</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Data analysis</td>
</tr>
<tr>
<td>Internal validity</td>
<td>Different efforts to encourage the honesty and clarify the role of informants. Let the informants check the collected and summarized data. Discussed the research approaches and results with the supervisor and other researchers. Used double-blind review process in journal articles.</td>
<td>Data collection</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Data analysis</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Research design and data analysis</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Dissemination</td>
</tr>
</tbody>
</table>

In the following, some specific issues relating to the reliability and validity of this research are discussed in more detail.
Benefits were discovered but what about the dis-benefits?

Upon charting out the individual impact mechanisms for the sample services, in some places potential dis-benefits (or costs) were observed, in addition to the benefits. So some impacts may well be negative – or at least they may be perceived as such initially. For example, the previously discussed case of detaining a vehicle (that has proven non-conformant to regulations in an inspection), would at first seem to yield immediate time losses for the operator, cargo clients and/or passengers. However, evaluating the net impacts of such a detention did not prove possible, mostly because the full impact mechanism is highly complex, as explained in the next paragraph.

In normal circumstances the inspections are resourced and carried out in such a way that they cause minimal disturbance to the day-to-day traffic operations. For example, with Safety Assessment of Foreign Aircraft (SAFA) inspections (EASA 2013), the full inspection can normally be performed during the standard aircraft turnaround time. Turnaround refers to the process of conducting all the required actions and preparations on the aircraft that are needed to get it back into operation (i.e., into the air en route to the next destination) safely and within the planned schedule. These actions include disembarking, fueling, catering refills, waste disposal, clean-up, cargo handling, re-boarding, etc. (Toivonen 2008). A detention of a vehicle or vessel is enacted only in cases of discovering serious safety-critical discrepancies. Regardless of the transport mode, rectifying and fixing such discrepancies have potential positive effects on traffic safety either by decreasing the accident probability of the detained vehicle or by yielding it better equipped to manage and survive in the case of an accident. Therefore, the logic relates to analyzing risks and making immediate decisions based on that risk analysis. The prioritization (by an inspector or a group of inspectors) in the case of a vehicle detention favors the chances of the vehicle actually being able to arrive at its destination safely rather than allowing it to try to make the trip within the originally planned and advertised time table—regardless of the safety risks. In other words, it is in such a case de facto considered socio-economically more beneficial that the vehicle arrives at its destination late than that in the worst case scenario it does not arrive at all but is lost in an accident. Therefore, in the most extreme cases the decisions (i.e., to detain or impose other corrective measures) may yield highly meaningful savings of time costs and accident costs by preventing disasters or by decreasing the risks of catastrophes. The complexity of the above-described
partially contradicting impact mechanisms made it impossible to evaluate the full net impacts relating to the used example and also on quantitative terms.

On the other hand, some at first glance seemingly positive impacts may well generate dis-benefits also – or at least confounding impacts – as a sort of a side-product. For example, increasing the resilience or fluency or safety of traffic in general, simultaneously makes increasing the ‘consumption of travel and transport’ more appealing to the consumers such as travelers, cargo clients, etc. This in turn may yield an increase in traffic and/or freight volumes, therefore generating an increase in carbon emissions and other negative effects to the environment and also to traffic safety.

The examples of dissonance presented above underscore the fact that the focus of this research approach was mainly on trying to discover benefits. Potential dis-benefits were observed and reported, but closer evaluation of these proved challenging because of the abovementioned reasons. The selected viewpoint is logical and justifiable based on the underlying basic assumption of public services having been set and regulated with at least the intention of producing positive impacts while trying to avoid negative ones. By the same token, there are no grounds to assert that public services could not have or do not have any negative impacts. In addition to the above examples, such negative impacts could relate either to overly excessive regulation or to public activities forming a barrier/hindrance to the efficient operation of markets. This connects directly to the theoretical discussion and the “iron law” formulated by Merton and Tumin (1949). Some further elaboration on these aspects follows.

*Excessive regulation?*

As became apparent during the course of this research, the amount of existing regulation is vast and furthermore, it is increasing all the time. It may well be that selected elements of the society or the transportation system are suffering from excessive regulation. Any regulation and its preparation, implementation and enforcement cause additional administrative load that emerges as increasing cost and inertia within the government. Or optionally the increased load may manifest as decreased quality of the regulation due to an inability to efficiently manage and enforce the growing mass of regulatory burden with resources that do not increase on the same slope – or might even decrease. And it is not enough to look only at the regulators’ side of the table because similarly, all new pieces of regulation also impose actions, adjustments and/or costs within the society among those parties.
affected by the regulation. According to the Finnish institute of juristic studies, Finland is among the most legislation intensive countries in the Western world (Litmanen 2004). Over a decade ago, the institute highlighted the following challenges with regard to regulation: (i) the increasing amount of regulation, (ii) finding the ideal level of regulation, and (iii) assuring the acceptability of regulation.

An additional special feature of the regulation concerning transport is that it often originates from outside of the nation; that is, it is largely regulated in the international fora. This is especially so within the aviation and maritime sectors but concerns all other transport modes as well, at least to some extent. According to the OECD (2014) the amount of intergovernmental and/or international regulation has been increasing exponentially since the 90s. The international activities of the case study subject are well established, but the emergent pressures and challenges within environmental concerns, energy consumption, and climate change surely will not decrease the need for more international regulation—rather the opposite. On the other hand, according to the OECD analysis (2014), the net impacts of regulation are by and large positive and especially those nations that are open and deal with a lot of international trade benefit the most from such international regulation. It also needs to be kept in mind that whereas small nations, such as Finland, do not hold a commanding voice in making decisions on new international regulation and commitments, they usually are not silent or voiceless partners, either. The national interests of, for example, Finland are represented in many round tables dealing with international regulation, and quite often the representative organization is in fact the main case study subject of this research. In the case of Finland, this national presence covers all transport modes.

The OECD (2014) recognized and classified some benefits and costs of international regulation in the following way:
Table 16. The perceived costs and benefits of intergovernmental regulation.

<table>
<thead>
<tr>
<th>Benefits</th>
<th>Costs / dis-benefits / problems / challenges</th>
</tr>
</thead>
<tbody>
<tr>
<td>Economic benefits</td>
<td>Increased administrative work and burden</td>
</tr>
<tr>
<td>Better risk management</td>
<td>Challenges in the implementation</td>
</tr>
<tr>
<td>Enhanced efficiency of regulation</td>
<td>Decreased national autonomy with regard to setting regulations</td>
</tr>
<tr>
<td>Decrease of overlapping work</td>
<td>Political challenges</td>
</tr>
<tr>
<td>Increased information and communication</td>
<td></td>
</tr>
</tbody>
</table>

Regardless of the origin of the regulation being domestic or international, any excess of it is clearly and by definition to be perceived as a dis-benefit or an undesirable set-up. The increase of regulation adds to the workload of ministries and agencies exponentially. One reason for this is that the impacts and effects of new regulation on various actors and citizen segments need to be charted out well in advance. Furthermore, every new piece of regulation calls for its practical implementation, setting the enforcement regime and auditing the performance of the regulation and assuring general conformance to the regulation.

There are some potential means to ease the burden and make the regulatory framework lighter and less strenuous. These include information guidance and flexible regulation. In information guidance, for example, the service providers and product manufacturers are obliged to inform the consumers about various aspects of the product in question, such as its fabrication and other characteristics. In an ideal case, the de facto market control can then be generated via consumer demand and self-regulation. Flexible regulation focuses on generating rules on a more general level. The aim there is to provide steering to the subject of regulation without diving into too much detail.

Useful or not, the amount of regulation is on the increase, and therefore the planning and setting of it needs to be done with care. The quality of regulation depends on, for example, how skillfully the different stakeholders can be included into the process so that the impacts of the regulation can be estimated and shown in a transparent and unambiguous manner.

*The efficiency of the markets is not ideal – partly due to regulation?*

All regulatory and other activities of public bodies undeniably have direct effects on many sectors of the society—and consequentially also on how efficiently and openly the markets are allowed to operate. The underlying goal of regulation is not
to hinder or restrict the markets. But on the other hand, it is not at all safe to assume
that all of the impacts of regulation regimes are solely positive for the markets and
their efficiency.

There is a reasonably established body of earlier research that has addressed
the connection between regulation and the efficient operation of the markets. For
example, the OECD (Gönenç and Nicoletti 2000) studied the air travel markets in
relation to the associated regulatory framework and constraints. It concluded that
the provision of air travel services could be made more efficient and thus the
consumer price of air travel could be decreased by developing the regulatory
framework in a more market-friendly direction. Pelkmans and Renda (2014) on the
other hand investigated the impacts of regulation on innovation activities. Several
of their case studies were from the transportation sector. In their synthesis they
estimated that the overall impacts are positive—but notably, potential dis-benefits
were recognized as well.

Another perspective that is highly relevant to market efficiency (and
additionally to the abovementioned issue of excessive regulation) in any nation is
the deregulation or the delegation approach. Some pieces of regulation may not
always require the direct and all-encompassing presence and oversight of a public
authority. In such cases it may sometimes be possible and also sensible to delegate
some of the regulatory burden down from the central government to the industry or
research sectors or even down to associations, without jeopardizing the quality or
integrity of the intended regulatory impact. The OECD Expert Groups provide one
valid example of this approach (OECD 2017b). These groups of experts investigate
various challenges within transportation in general in an informal setting. As
outputs from their analyses and investigations, they produce various kinds of
recommendations for the member countries.

Additional evidence of the successful delegation of regulatory burden was
witnessed within this research (Mononen and Leviäkangas 2015b) in analyzing the
non-commercial aviation related services and the particular example from the
United Kingdom with regard to the safety of ultra-light aviation. In United
Kingdom, the risk of casualty within ultra-light aviation has decreased dramatically
over the past three decades. Thirty years ago the risk of dying was four deaths per
10000 flight hours whereas it now is 0.28 deaths per 10000 flight hours.

1 This claim holds true only when operating within a market area. Outside and between market areas
(e.g., between European Union and other parts of the world) some pieces of regulation may well be
gear more toward protecting the market rather than toward opening up the market.
Simultaneously the regulatory and enforcement responsibilities have gradually been decentralized and delegated to a certain extent from the central national civil aviation administration down to the flight clubs and associations, step by step. Currently in Finland the risk of fatal accident while flying an ultra-light aircraft is 1.23 deaths per 10000 flight hours (Trafi 2014b), while all of the regulatory and enforcement responsibilities are still in the administrators’ hands. It is to be acknowledged that such advances as those witnessed in the UK do take a lot of effort, development and time – and also plenty of committed cooperation and communication between various actors. But on the other hand, it is hard to pinpoint undisputable reasoning for why ultra-light aviation safety should not or could not be as safe in Finland as it is in the United Kingdom.

To conclude based on the above, in ideal cases trimming down, streamlining and delegating the regulatory framework or responsibilities can lead to win-win situations. And since this is the case, we can also deduce that all regulation is not to be assumed to always be on an ideal level. However, the development, direction and control of regulation need to be carried out with great consideration and discrete care.

Some services of a transport administration may be directed (via legislation and policy) straight at conducting market oversight and control. This control aims at securing that all of the vehicles (e.g., cars, boats, vessels, aircrafts, mopeds, bicycles, motorbikes, snow mobiles, etc.) and other transportation equipment being produced, imported, and sold are safe, provided that they are used and installed properly and according to the associated instructions. The term “safe” here covers not just the primary user’s health in terms of transport safety but also the safety of the surrounding people, environment, and property. The quantitative impacts of the market control regimes and services were not directly studied in this research, but by definition this control is in place with the intention to increase both transport safety and transport environment friendliness directly. By the same token there are valid grounds to assert that the market control regimes at the same time inevitably increase the administrative burden for the businesses operating within the regulated economy.

*The constructed impact mechanisms are generalizable, but are the impacts being reproduced elsewhere in the same quantity?*

In this research, direct benefits and costs of services were evaluated for the services within the sample only. Leaning on these findings, an attempt was made to paint a
picture of the benefits and costs of the entire agency—a picture as transparent and as coherent as possible. The assumed chains of deduction aim at applying consistent logic throughout, but it is not safe nor realistic to argue that no holes or blind spots could exist therein.

The impact mechanism maps constructed for individual services are generalizable to an extent. This is to say that even though all maps and the underlying logical chains of yielding impacts were always constructed with a particular single service under inspection in mind, often the basic impact construct could be applied to other services with minor modifications, regardless of, for example, transport modes. Let us look at an example by comparing two separate services within the sample representing two modes of transport. The services in this example are maritime vessel inspections and aviation inspections for foreign aircraft. The impact mechanisms for these two services are practically identical, except for the setting and the vocabulary/terminology used. Whereas in the maritime inspections the inspection objects are vessels and shipping companies, in aviation they are aircrafts and airlines.

The same applies to the process of clustering the services into service families. The clustering was based on the generic descriptions of the stem services constituting the cores of all families. Other services could be included into a family only by them directly and fully adhering to the generic description of that family’s stem service. Therefore it can be concluded that the reasoning and methods behind the clustering as well as the similarity and commonality of all services within a family have been defined as carefully and transparently as possible.

Regardless, certain blind spots will inevitably remain. The final conclusions were not based on an assumption that all the services within a family would yield impacts exactly the same (quantitatively and otherwise) as the stem service has been observed to yield them. But on the other hand, the presented analyses have catered for recognizing and showing some of the connectivity and relationships of these kinds. The blind spot and the potential source of error herein is that regardless of fully valid impact mechanisms, it is possible that the other services within the family differ from the stem service in terms of the magnitude of the impacts yielded.

This error may occur in either direction. That is, it is possible that the impact generation of the stem service is greater than the other services within the family. But it is as likely that the stem service in fact yields impacts to a lesser extent than the other members of the respective family. Because of this characteristic, it is reasonably safe to assume that the potential errors largely cancel each other out, and therefore a significant systematic overall total error is highly unlikely.
Gaps in the available research data

The analysis and evaluations of this research were largely based on earlier scientific research and data. That was the basic objective and approach adopted early on, even at the onset of the research. Regardless of the prior understanding, it was still somewhat surprising to observe how limited the previous research is with regard to the sovereign and direct socio-economic impacts of public functions and services.

As an example of the research sometimes residing in silos and lacking a true systems approach, it may be useful to look at some observations with regard to an assumedly relatively mature sector of research, namely that of road transport safety. The research gap observed there covers an entire transport mode and as such extends beyond the public sector—in fact the field includes all impacts of public activity and regulation, among other things. As said, this research sector and the required research data have an established and long research tradition and history. The abundance of valid examples with regard to road safety research and its maturity is staggering, so much so that quantifying or illustrating this in a meaningful way is somewhat challenging. One display of this could be that globally there are hundreds of research institutes, universities, faculties, and consultancies (or divisions of these) dealing with road transportation safety research. Scholarly literature database searches on this general level yield a great deal of information, not to mention that online searches with the terms “road safety research” and the like yield millions of hits.

Regardless of the maturity of the research, it was somewhat surprising to find very little evidence of a systems approach - that is, of the roles of different components of road safety and their contribution to the status of road safety. “Components of road safety” here refers to the parallel developments within, for example, vehicle technology, enforcement, infrastructure, and education. These all have an undisputed and clear connection to road safety, and each of these sectors has been researched extensively—but largely within their own silos. What is missing is either a national, European-wide, or global understanding of the market shares of each of these sectors with regard to the observable trends within road safety. “Market share” here refers to the individual contribution of each sector to the highly positive development of road safety over the past few decades.

So, the various sectors of improving road safety have been studied a great deal, but a macro-level synthesis of the total picture is nowhere to be found; it seems that it has yet to be collated. During this research an answer to this was first sought via a literature survey. After failing there, the personal professional cooperation
networks were approached—in particular by contacting many experts within national and international research organizations, transport administrations, and the vehicle industry—unfortunately with similarly unimpressive findings.

In an attempt to further visualize the emerged dilemma, Figure 21 was created by including only the previously mentioned categories of vehicle technology (65%), enforcement (10%), infrastructure (20%), and education (5%). In the absence of more solid data, these quoted percentages used in producing Figure 21 are the more or less educated estimates of the researcher for the market shares of each sector.

With regard vehicle technology, four remarkable “quantum leap”-level interventions can be recognized. From the earliest to the latest, these leaps are seatbelt, the crash-worthiness of the vehicle chassis, airbags, and active & passive driver assistance systems. Another attribute relating to vehicles is the vehicle condition that is affected by things such as the interval between and quality of vehicle inspections, vehicle fleet age, etc. Infrastructure-related attributes include the condition of the infrastructure, the relative number of roads (in various road classes), planning guidance and control, infrastructure capacity in relation to traffic volumes, penetration levels of intelligent transportation systems & data (a quantum leap-level change), etc. In addition to the mentioned four sectors, at least one important factor has been omitted here, namely the fluctuations and development of the economy in general. Changes in the GDP and other economic activity indicators are mirrored more or less directly within transport via changes in transport volumes (and therefore via changes in capacity usage and exposure), changes in motorization and median vehicle fleet age, infrastructure investment activity, and so on. With regard to enforcement, the introduction and implementation of various automated enforcement systems can be considered a quantum leap-level change. In comparison, the driver training and education regimes have remained more stable.

The x-axis of Figure 21 represents a time-line from 2001 to 2012, and the figures on the y-axis are annual number of lives saved on European roads in comparison to 2001 (European Union 2013). A cumulative total number of approximately 153,000 lives saved on European roads during the past decade have been allocated to each sector with the quoted percentages. During 2012 there were circa 27,000 fewer casualties on European roads than during 2001. This means that the annual number of deaths on European roads has nearly halved within a decade, and therefore the trend could be called extremely positive without much fear of correction.
Fig. 21. Number of lives saved on European roads 2001-2012 (the total number of lives saved from European Union 2013).

In addition to these being non-validated researcher estimates, it is also highly unlikely that the relative market shares of these components would have remained unchanged during the depicted period of time. But the real question still is, “What have these market shares actually been in reality?” This research argues that these historical trends and especially the reasons behind them would provide valuable insight into where the improvement potential has been, and therefore most likely also where it will be in the future—at least in the short- and mid-term time perspectives. This would also give an indication of where the foci of both research and other public or private development activities should be today and in the near future.

Referring to the previously mentioned wealth of research in transportation safety, another observation was made: the situation is somewhat different with regard to research on the environmental impacts of transport. The available prior research there is less plentiful and mature, which is at least partially explained by the international political pressures and ethical concerns for the environmental aspects being of more recent origin. However, this situation is continuously
changing for the better since environmental research is currently a hot topic in many sectors of life and research and not just within transport.

These examples bring forward some of the research gaps observed during the research at hand. Others were also observed, especially during the detailed analyses of the sample services within the main analysis unit. As such, it did not come as a total surprise that some of the data—especially quantitative data—might be hard to find or be missing altogether. As for mitigating the implications of missing data with regard to the validity of this research, some precautions were constructed. These included preparedness to involve several impact analysis methods, as described earlier in sub-chapters 1.3 and 2.4 and within articles II, III, and IV. In the absence of direct scientific quantitative evidence, these methods were utilized.

**Limited amount of directly comparable prior research**

Evaluations of companies, projects, and programs have been conducted widely, and in that sense evaluation research has long established scientific traditions. The preceding evaluations of public organizations have usually focused on qualitatively analyzing the organizational effectiveness, governance structure, operational success, or administrative productivity. Also, qualitative studies on the external stakeholders’ perceptions of agencies’ brand value and acceptability can be found in the literature. As discussed earlier in sub-chapter 2.4, evaluating the quantitative socio-economic impacts of public authorities has presented a research gap that this research helps to address and contribute to. The need for such evaluations is being expressed and witnessed in many ways, nationally and globally. The topic is challenging, not least due to the fact that the scientific methodology around it is not yet thoroughly established.

**Limitations of the study and transferability**

Some limitations inherently apply with regard to the further use of the findings and the approach presented in this research. The direct findings presented here are observations and discoveries from the employed material. Strictly speaking, the findings cannot be as such directly generalized to a larger population. The analysis describes the phenomena in detail but only as they are embedded in the local context. Therefore, the findings apply directly only to the case study subject.

The findings and the approach may, however, be quite transferable to another setting. One of the strengths of the adopted approach is that, unlike often in
qualitative research, the main bulk of the data used in this research does not rely on interviews but instead relies on more objective documentation. This adds robustness and transparency to the analysis since, for example, the interviewees’ personal beliefs and idiosyncrasies regarding the research topic are excluded, and issues of confidentiality or anonymity do not come into play when presenting the findings.

As a multimodal agency, the Trafi case provides a benchmark for other governments or agencies considering how a sustainable transport system is governed and managed and, more importantly, how the performance level of the performance management system could be evaluated and improved. Therefore, this research can be seen as normative and value-adding: similar analyses could be repeated in other contexts, be they regional, national, or international. If an agency’s performance management system is lacking the relevant aspects of transport safety or environment, we can expect the performance of such an agency to be inadequate regarding the missing areas, therefore implying a suboptimal use of resources. Should such findings prove to be common when repeating similar research, an assumption would follow that some public resources are in less than optimal use and that the accountability of the agencies is incomplete in this regard. This makes the described case of Trafi an observation that has wider implications across transport safety administrations in different countries.

In trying to critically evaluate the entire research process presented here, and its transferability, the question to be asked could be: what would need to be done in a different way if one would need to start this process again with the knowledge we now have? In all honesty, to some degree the author may have become blind to the process because it is now very hard to pinpoint such cases. In other words, during the process the author did not enter regrets, referring to situations where this or that earlier choice would have blocked further progress. By the same token it should be noted that (as described earlier in this dissertation) not all of the steps in the methodology design were linear but were rather iterative. In practice this is to say that smaller blocks were in fact encountered, and as a consequence, steps back needed to be taken along the journey. One example of this was the need to scrap the first version of the agency’s service catalogue and start a-new, as described in article II. Another example is the criteria design, where the original criteria sets that were defined based on the literature were not adequately covering the agency’s legislated responsibilities. But regardless, the final process and set of methods, as presented here, proved functional and could therefore be transferred as such.
This is not to say that everything would or should be done in the same exact manner, and that no alternative approaches should be considered. On the contrary, there are elements within the process that would be highly interesting to be tried in another way. One prime example is the sampling process. During the method design, several methods were considered when making this choice. The finalists were MCDA and AHP, out of which MCDA ended up as the method of choice. The final deciding factor was that AHP as a group decision making process was considered more prone to be affected by the informants’ beliefs and idiosyncrasies than MCDA. Regardless, an attempt of applying AHP in similar processes would be highly interesting. Another potential avenue for determining the sample would be some form of random sampling, which might be feasible in other settings.

To conclude, the process and the methodology, including the evaluation criteria, could be transferable as they are, to corresponding agencies internationally. The author estimates that in such settings the entire process could be replicated with a fraction of the resources that were dedicated to this research. With moderate modifications the process could be transferable to other types of agencies within transportation; however the evaluation criteria at least would need to be re-thought, mainly based on the legislated responsibilities of the given agency. (When the intended purposes of services and the impacts sought after change, so must the evaluation criteria.) Transferring regardless of the field of administration seems also feasible to an extent – but the further away from transportation or civil engineering the intended application area shifts, the more significant adjustment and redesign needs are envisaged.

4.4 Recommendations for further research

As for future development ideas resulting from the work presented, one rises above all others. Namely, in the literature review for this work, corresponding cross-checks of tasks, objectives, and the associated indicators in regard to transport safety authorities were not easily found. It would not be a great surprise to the author if similar blind spots also exist elsewhere—actually, the author dares presume the presence of gaps in performance management systems broadly across equivalent organizations everywhere in the world. The deduction goes as follows: according to World Economic Forum (2014), when it comes to the global competitiveness index, Finland ranks fourth overall, and, moreover, with regard to the soundness and fairness of the institutional framework, Finland ranks as number two—the second best country in the world. Therefore, if flaws and gaps can be
discovered in Finnish performance management systems, it is not unimaginable that they might also exist elsewhere.

It would be highly recommendable to perform similar analysis in other countries and in varying fiscal, political, cultural, administrative, and other settings and surroundings. Among the world’s many nations, Finland can in many ways be considered a “model student” in terms of, for example, the transparency, robustness, and soundness of handling national administration. It would be extremely intriguing to find out more about, for example, (i) how much diversity exists in the models and legislation regarding organizing the national traffic safety and other administrative duties; (ii) how different the structures of service portfolios may appear in other settings, and (iii) how available and accessible the needed basic data would be in less developed countries for this kind of research. Unfortunately, because of resource constraints, it is not possible to include this kind of analysis within this research framework and, correspondingly, within this manuscript. This kind of very interesting and value-adding comparison therefore needs to be a separate exercise, but, as such, it presents a valid addition to the conclusions of the study—namely the recognition of the emerged research gaps and recommendations for further research.

Therefore, repeating the approach in other settings, such as for corresponding agencies in other countries, would be highly interesting. In that way further information on the validity and generalizability of the presented approach could be acquired. The developed process proved successful, hence clearing a path for similar analyses on other agencies, administrations, and services, provided the necessary data are available.

Even more sophisticated methods are hopefully being considered, but the approach described here is one potential foundation for future development. A specific recommendation for further development of the process is designing more comprehensive ways to address the spill-over effects of impacts and benefits both across the society and across time.
5 Summary

This research showed that the overall performance and impact of a service oriented public agency can be assessed. Relying on previous research and literature, this work presented a novel and functional process for such evaluation. The presented methods are capable of determining the impacts of a service portfolio, quantifying impacts on socio-economic terms, and presenting them on a monetary canvas. Therefore this research contributed to increasing the understanding within evaluation research. The findings also provided fresh access to more sound and more transparent socio-economic performance metrics. These can be used as tools in efforts to achieve new and more efficient performance management systems.

This research indicated that the overall operation of a public agency can be socio-economically net beneficial. The impacts may emerge in a variety of forms but by funding such an agency, the society gains positive impacts that have a value exceeding the investment. Ideally, a public agency that performs as intended can be a value multiplier. Constant fine tuning of the performance is recommended. Careful thought is advisable whenever considering adjusting the resource allocation levels of public entities. This is especially so when the adjustments are targeting net beneficial bodies – otherwise there is a risk of sacrificing valuable resources.

On one hand, the relatively scarce prior research around this particular phenomenon presented some challenges to carrying out the research to a successful completion. But on the other hand, clear future potential and opportunities emerge from the same exact reason. The research presented here is needed and justifiable and the research findings are being currently applied in practice to enhance the operations of the main unit of analysis and beyond. The methodology contribution has potential added value, generalizability, and re-usability, within this field and also outside it.

From an agency’s management point of view, quantitative and qualitative tools that demonstrate the benefits and effectiveness of the agency’s services are of crucial importance. Demonstrating efficiency and effectiveness as well as wider socio-economic impacts justifies (or falsifies) the whole existence of the administration’s prevailing structure and service architecture. The internal budget competition between public administration sectors leads to the increasing need to show and prove that there are impacts and that they are in the desired direction. Hence neither upper level administrations’ such as ministries nor any individual
subordinate agencies’ management can afford not to measure and show the case for
the impact of its services or other produce.

‘Measure twice, cut once.’

*Traditional Finnish proverb, commonly used by carpenters.*
References


150


Mulley C & Ho C (2013) Evaluating the impact of bus network planning changes in Sydney, Australia, Transport Policy, 30: 13-25


OECD (2010) OECD Public Governance Reviews, Finland, Working Together to Sustain Success, Assessment and Recommendations


OECD (2014b) International Regulatory Co-operation and International Organisations: the Cases of the OECD and the IMO. OECD Publishing


154


Sedlmeyer P (2014) From associations to intuitive judgment and decision making: Implicitly learning from experience. The Routines of Decision Making, 83-100


Trafi (2014) Trafi Annual Report 2013. In Finnish only


Original publications


Reprinted with permission of the copyright holders. The named Journals are the original sources of publication for the above mentioned four articles. Original publications are not included in the electronic version of the dissertation.
614. Roivainen, Antti (2017) Three-dimensional geometry-based radio channel model: parametrization and validation at 10 GHz
616. Yliniemi, Juho (2017) Alkali activation-granulation of fluidized bed combustion fly ashes
617. Iljana, Mikko (2017) Iron ore pellet properties under simulated blast furnace conditions: investigation on reducibility, swelling and softening
619. Schuss, Christian (2017) Measurement techniques and results aiding the design of photovoltaic energy harvesting systems
622. Li, Xiaobai (2017) Reading subtle information from human faces
623. Luoto, Markus (2017) Managing control information in autonomic wireless networking
626. Lappi, Tuomas (2017) Formation and governance of a healthy business ecosystem
630. Palosaari, Jaakko (2017) Energy harvesting from walking using piezoelectric cymbal and diaphragm type structures

Book orders:
Granum: Virtual book store
http://granum.uta.fi/granum/
Petri Mononen

SOCIO-ECONOMIC IMPACTS OF A PUBLIC AGENCY – ENHANCING DECISION SUPPORT FOR PERFORMANCE MANAGEMENT