Het verkeerskundig
laboratorium
voor studenten

ITS EDU LAB

Ex-post evaluation of network-wide Dynamic Traffic Management

Proposal for a theory-based evaluation methodology

Suerd Polderdijk

March 2009

Rijkswaterstaat

TU Delft

Ministerie van Verkeer en Waterstaat

Rijkswaterstaat

Ex-post evaluation of network-wide Dynamic Traffic Management

Proposal for a theory-based evaluation methodology

Suerd Polderdijk

March 27, 2009

MSc Thesis Transport, Infrastructure and Logistics
Delft University of Technology

Colophon

Published by: ITS Edulab, Delft
ITS Edulab is a cooperation between the Rijkswaterstaat centre for Transport and Navigation and the Delft University of Technology

Information: Henk Taale
Email: Henk.Taale@rws.nl

Author: Suerd Polderdijk
Delft University of Technology
Master Transport, Infrastructure and Logistics

Graduation committee:
Prof. Dr. Ir. Serge P. Hoogendoorn
Committee chairman,
Delft University of Technology
Faculty of Civil Engineering and Geosciences
Dr. Ir. John H. Baggen
Delft University of Technology
Faculty of Technology, Policy and Management
Dr. Ir. Sascha Hoogendoorn-Lanser
Rijkswaterstaat, Centre for Transport and Navigation
Dr. Ir. Henk Taale
Rijkswaterstaat, Centre for Transport and Navigation
Ir. Feiko van der Veen
DHV B.V., Environment and Transportation
Dr. Ir. Andreas Hegyi / Dr. Ir. Francesco Viti
Delft University of Technology
Faculty of Civil Engineering and Geosciences

Date: March 27, 2009

Status: Final

Summary

Dynamic Traffic Management (DTM) is a means to increase road infrastructure efficiency. This increase is achieved by matching infrastructure demand and supply, in time as well as over networks.

Decision-making about DTM in the Netherlands is hampered by a lack of knowledge of impacts, costs and benefits of network-wide DTM. A suitable evaluation methodology is required to gain this knowledge. The current evaluation methodology of Rijkswaterstaat is documented in the "Guideline for model and evaluation studies of utilisation measures[1]" and the "Guideline for ex-post evaluation of utilisation measures on the main road network[2]". This methodology is developed for evaluation of local DTM measures and may no longer be suitable to evaluate network-wide DTM. Rijkswaterstaat will therefore design a new evaluation guideline within the "National evaluation program[3]".

Research objective

The objective of this thesis is to identify strengths, weaknesses, opportunities and threats of the current ex-post evaluation methodology for DTM of Rijkswaterstaat and to propose a new evaluation methodology

A SWOT analysis is used to identify strengths, weaknesses, opportunities and threats of the current evaluation methodology. Trends are towards network-wide and coordinated use of DTM. Due to these trends identified weaknesses of the current evaluation methodology will become more apparent in the future.

The extent in which three existing alternative evaluation methodologies (CONVERGE methodology, Finnish ITS evaluation methodology and Canadian Finnish ITS evaluation methodology) provide improvements for the identified weaknesses is studied. It is concluded that improvements are offered to an extent, but weaknesses remain in the studied alternative methodologies.

In this report a new methodology for DTM evaluation is therefore proposed. This methodology uses a theory-based approach. Current DTM evaluation approaches focus on expected final impacts of DTM measures. The concept of theory-based evaluation is that every DTM design is based on theory and this project theory should be the basis of evaluation. Therefore evaluation should not just focus on expected final impacts of DTM, but also on assumptions that explain *how* these expected final impacts are achieved.

[1] In Dutch: "Leidraad model- en evaluatiestudies benuttingsmaatregelen"

[2] In Dutch: "Richtlijn voor de ex-post evaluatie van benuttingsmaatregelen op het hoofdwegennet"

[3] In Dutch: "Landelijk evaluatieprogramma"

“Logic models” can be used to make project theory explicit. Logic models are diagrams that describe how DTM measures are expected to achieve their expected impacts. The proposed evaluation methodology offers a step-by-step approach to develop logic models for DTM projects. Based on these logic models, relevant evaluation questions are formulated in a structured manner. The formulated evaluation questions and logic models are then used to select appropriate evaluation methods, indicators and experimentation set-ups. Disturbing external influences on evaluation are explicitly taken into account in these selections.

Logic models include assumptions about the (technical) functioning of DTM measures, road user behaviour and impacts. Therefore the proposed evaluation methodology links various evaluation types common in DTM evaluation to each other, such as technical, user-response and impact evaluation.

Using a case study and a SWOT analysis, strengths, weaknesses, opportunities and threats of the proposed evaluation methodology are identified. The case study of the “Field Test Traffic Management Amsterdam[4]” shows that the methodology can be applied to a real-life, network-wide DTM project.

It is concluded that the proposed evaluation methodology is better suited to cope with current trends in DTM than the current evaluation methodology. The proposed evaluation methodology:

- Is flexible to take any impact of DTM into account;
- Integrates various evaluation types and project phases into one evaluation;
- Explicitly evaluates assumptions about the functioning of DTM measures and road user response to them; and
- Reduces the likelihood that measured impacts not caused by DTM measures are attributed to them.

If properly applied, it is expected that the proposed evaluation methodology results in evaluations that are more useful, convincing and reliable than using the current evaluation methodology. However, further work to expand the methodology and test it in practice remains to be done.

It is recommended to use the proposed evaluation methodology as a basis for the new evaluation guideline for DTM projects of Rijkswaterstaat. The development of logic models should be integrated into the DTM design process, for example by integrating it into the “Sustainable traffic management[5]” process. Not only DTM evaluation, but also DTM design will benefit from this.

[4] In Dutch: “Praktijkproef Verkeersmanagement Amsterdam”

[5] In Dutch: “GebiedsGericht Benutten (GGB)”

Samenvatting

Dynamisch Verkeersmanagement (DVM) is een middel om de benutting van het verkeerswegennet te verhogen. Deze verhoging van de benutting wordt bereikt door verkeersvraag en -aanbod op elkaar af te stemmen, zowel in tijd als over het netwerk.

Besluitvorming over DVM in Nederland wordt belemmerd door een gebrek aan kennis van de effecten, kosten en baten van netwerkbreed DVM. Een geschikte evaluatiemethodiek is benodigd om deze kennis te verkrijgen. De huidige evaluatiemethodiek van Rijkswaterstaat is vastgelegd in de “Leidraad model- en evaluatiestudies benuttingsmaatregelen en “Richtlijn voor de ex-post evaluatie van benuttingsmaatregelen op het hoofdwegennet” Deze evaluatiemethodiek is ontwikkeld voor de evaluatie van lokale DVM maatregelen en is mogelijk niet langer geschikt om netwerkbreed DVM te evalueren. Binnen het “Landelijk Evaluatieprogramma” zal Rijkswaterstaat dan ook een nieuwe evaluatieleidraad ontwikkelen.

Doelstelling van het onderzoek

Doel van deze scriptie is het identificeren van sterke en zwakke punten, kansen en bedreigingen van de huidige methodiek voor ex-post evaluatie van DVM projecten van Rijkswaterstaat en het voorstellen van een nieuwe met methodiek.

Met behulp van een SWOT analyse zijn sterke en zwakke punten, kansen en bedreigingen van de huidige evaluatiemethodiek geïdentificeerd. De trends in DVM zijn in de richting van de netwerkbrede en gecoördineerde inzet van DVM maatregelen. Door deze trends zullen de geïdentificeerde zwakke punten van de huidige evaluatiemethodiek in de toekomst duidelijker zichtbaar worden.

De mate waarin drie alternatieve evaluatiemethodieken (CONVERGE methodiek, Finse ITS evaluatiemethodiek en Canadese ITS evaluatiemethodiek) verbeteringen bieden voor de geïdentificeerde zwakke punten is onderzocht. Hoewel verbeteringen worden aangeboden, blijven zwakke punten aanwezig in de bestudeerde alternatieve methodieken.

In dit rapport wordt dan ook een nieuwe evaluatiemethodiek voor DVM projecten voorgesteld. Deze methodiek gebruikt een theorie-gebaseerde benadering. Huidige evaluatiemethodieken richten zichten zich met name op verwachte (eind-)effecten van DVM maatregelen. Theorie-gebaseerde evaluatie stelt dat elk ontwerp van een DVM project gebaseerd is op een theorie en dat deze projecttheorie de basis van evaluatie zou moeten zijn. Evaluatie zou zich daarom niet alleen op verwachte (eind-)effecten van DVM moeten richten, maar ook op de aannames die uitleggen *hoe* deze verwachte effecten bereikt worden.

Een “logisch model” kan gebruikt worden om deze projecttheorie expliciet te maken. Een logisch model is een diagram dat beschrijft hoe de DVM maatregelen naar verwachting de gewenste (eind-)effecten zullen creëren. De voorgestelde evaluatiemethodiek biedt een stapsgewijze aanpak om een logisch model voor een DVM project te ontwikkelen. Aan de hand van dit logisch model worden vervolgens gestructureerd evaluatievragen geformuleerd. Op basis van deze vragen en het logisch model worden geschikte evaluatiemethoden, indicatoren en experimentele opzet gekozen. Verstorende externe invloeden op de evaluatie worden expliciet in deze keuzes meegenomen.

Een logisch model bevat aannames over het (technische) functioneren van DVM maatregelen, gedrag van weggebruikers en resulterende (eind-)effecten. De voorgestelde methodiek verbindt daarom op een logische wijze de verschillende evaluatietypen die gangbaar zijn binnen DVM met elkaar, zoals technische evaluatie, gebruikersevaluatie en effectevaluatie.

Met behulp van een casusstudie en een SWOT analyse zijn de sterke en zwakke punten, kansen en bedreigingen van de voorgestelde evaluatiemethodiek geïdentificeerd. De casusstudie van de “PraktijkProef Amsterdam” toont aan dat de methodiek toegepast kan worden op een bestaand, netwerkbreed DVM project.

Geconcludeerd wordt dat de voorgestelde evaluatiemethodiek beter aansluit bij de huidige trends in DVM dan de huidige evaluatiemethodiek van Rijkswaterstaat. De voorgestelde evaluatiemethodiek:

- Is algemeen toepasbaar om elk effect van DVM te evalueren;
- Integreert de verschillende evaluatietypen en projectfasen in één evaluatieaanpak;
- Evalueert expliciet de aannames over technisch functioneren en het gedrag van weggebruikers; en
- Vermindert de kans dat gemeten effecten die niet zijn veroorzaakt door de DVM maatregelen toch aan deze maatregelen worden toegeschreven.

Verder onderzoek is benodigd om de methodiek verder te ontwikkelen en in de praktijk te testen. De voorgestelde evaluatiemethodiek zal naar verwachting resulteren in een evaluatie die nuttiger, overtuigender en betrouwbaarder is dan met behulp van de huidige evaluatiemethodiek.

Het wordt Rijkswaterstaat dan ook aanbevolen de voorgestelde evaluatiemethodiek te gebruiken als basis voor de nieuwe evaluatieleidraad voor DVM projecten. Verder is het aan te bevelen de ontwikkeling van logische modellen te integreren in het ontwerpproces, bijvoorbeeld door het te integreren in het “Gebieds Gericht Benutten” (GGB) proces. Niet alleen de evaluatie van DVM projecten, maar ook het ontwerp van DVM maatregelen heeft hierbij baat.

Voorwoord

“Our engineers have done some testing and evaluation, and overall we concluded this was an interesting option to discover information.”
David L. Tennenhouse, Senior Research Scientist, M.I.T.

Dit rapport is geschreven in het kader van mijn afstudeeronderzoek voor de master Transport, Infrastructure and Logistics aan de TU Delft. Dit afstudeerwerk heb ik mogen doen in het “ITS Edulab”, een samenwerkingsverband tussen de TU Delft en de Dienst Verkeer en Scheepvaart van Rijkswaterstaat. Bij dezen mijn dank voor de collega’s bij het Edulab, de projectteams van “FileProof A10” en de “PraktijkProef Amsterdam” en de afdeling Verkeersmanagement Wegennet voor de gezelligheid, adviezen en uiteraard potjes tafelvoetbal.

Uiteraard een hartelijk dankwoord voor mijn afstudeercommissie. Andreas, Feiko, Francesco, Henk, John, Sascha en Serge, ik heb met elk van jullie met veel plezier samengewerkt. Bedankt voor jullie scherpe tips, prettige begeleiding, enthousiasme en gezelligheid!

Manus en Wout, het waren jullie tips en ideeën die me de gekozen richting in hebben gestuurd. Onze twee afspraken hebben daarmee een belangrijke bijdrage aan het resultaat geleverd, dankjewel!

Last but zeker not least, Merel, dankjewel voor het doorlezen en dat je bereid was al mijn saaie afstudeerverhalen aan te horen...

Delft, 27 april 2009

List of English – Dutch translations

English	Dutch
Before and after study	Voor- en nameting of nul- en éénmeting
Centre for Transport and Navigation	Dienst Verkeer en Scheepvaart
Control strategy	Regelstrategie
Dynamic Traffic Management	Dynamisch Verkeersmanagement
Field Test Traffic Management Amsterdam	Praktijkproef Verkeersmanagement Amsterdam
Guideline for model and evaluation studies of utilisation measures	Leidraad model- en evaluatiestudies benuttingsmaatregelen
Guideline for ex-post evaluation of utilisation measures on the main road network	Richtlijn voor de ex-post evaluatie van benuttingsmaatregelen op het hoofdwegennet
Improving traffic flow on the A10 ring road	Verbeteren doorstroming Ring A10
Main road network	Hoofdwegennet
National evaluation program	Landelijk evaluatieprogramma
Network Vision North-Holland	Netwerkvisie Noord-Holland
Policy Note on Mobility	Nota Mobiliteit
Rijkswaterstaat Traffic Centre the Netherlands	Rijkswaterstaat Verkeerscentrum Nederland
Secondary road network	Onderliggend wegennet
ShoulderDRIP	BermDRIP
Sustainable Traffic Management	Gebieds Gericht Benutten
Utilisation	Benutting
Vehicle hours of delay	Voertuigverliesuren

Table of contents

1. Introduction

1.1 Introduction to Dynamic Traffic Management

Utilisation[6] concerns synchronising infrastructure demand and supply, both in time and over the network. This synchronisation is done in such a way that network performance is optimised using existing physical infrastructure.

Measures to increase utilisation of existing road infrastructure are receiving increasing attention in the Netherlands. Construction of new roads and the introduction of road pricing are postponed. This has led to an increase of attention for utilisation. The Ministry of Transport, Public Works and Water Management has underlined this trend, by incorporating “utilisation” as one of the pillars of Dutch national transportation policy. This policy is documented in the “Policy note on Mobility[7]” (Ministry of Transport, Public Works and Water Management, 2006).

An instrument to increase road infrastructure utilisation is Dynamic Traffic Management (DTM). The “dynamic” aspect concerns coupling current (or predicted) traffic conditions to measures being taken. These measures for example consist of ramp metering and Dynamic Route Information Panels (DRIPs).

1.2 Problem definition

From local to network-wide DTM
Dynamic Traffic Management is currently mainly implemented locally in an isolated and reactive manner (Hoogendoorn and Van Zuylen, 2008). DTM measures are applied to relieve a local bottleneck, individual DTM measures are not coupled and thus do not cooperate and DTM measures react to current traffic conditions.

DTM may not be exploited to its full potential with this approach. Local DTM may solve problems locally, but creates new problems at other locations. Ramp metering may for example improve traffic flow on the main road network but also shift queues to urban roads. Reactive DTM may aim to reroute traffic via alternative routes, only to cause new congestion on these alternative routes. Uncoordinated DTM could create situations where one DTM measure is worsening a problem another DTM measure is trying to solve.

[6] In Dutch: “Benutten”

[7] In Dutch: “Nota Mobiliteit”

DTM may be exploited to its full potential using a network-wide, anticipative and coordinated approach. By using a network-wide approach to DTM design, remaining capacity available in networks can be utilised. Traffic can be controlled in such a way that for example important links in networks remain free of congestion or a bottleneck of air quality is relieved. This however requires cooperation between DTM measures; a form of coupling between DTM measures is often a necessity. Operating DTM in an anticipative manner could prevent bottlenecks from occurring, instead of relieving them. Experience with network-wide DTM is currently being acquired in the Netherlands and a large scale field test of network-wide DTM is planned for 2010 in and around Amsterdam, see Schwarz, Van Schie, and Verroen (2007).

Need for insight into costs and benefits of network-wide DTM
Despite the expected benefits, decision-making around network-wide DTM is hampered by a lack of knowledge about impacts of network-wide DTM. Knowledge about impacts on traffic flow of *local* DTM measures in regular conditions has been acquired, at least to an extent (see for example Rijkswaterstaat, 2008b). However, impacts of DTM measures in non-regular conditions (incidents, severe weather conditions, events) are mostly unknown, as are impacts of DTM on emissions, noise hindrance and traffic safety (Ministry of Transport, Public Works and Water Management, 2008). Knowledge about impacts of *network-wide* DTM is mainly limited to model predictions and theoretical estimates (Ministry of Transport, Public Works and Water Management, 2008).

Besides gaining insight into impacts of network-wide DTM, the decision-making process may benefit from a framework to interpret its costs and benefits (Rijkswaterstaat, 2008a). Costs and benefits of network-wide DTM are largely unknown. Insight into (societal) costs and benefits of network-wide DTM might speed up the decision making process, leading to an implementation of DTM that is more geared towards demands of society and allows expenditures for DTM to be justified.

Current evaluation guidelines are not intended for network-wide DTM
By starting the “national evaluation program[8]”, the Dutch Centre for Transport and Navigation intends to increase its knowledge of costs and benefits of DTM. The goal of the “national evaluation program” is to “allow better and well-considered choices regarding the implementation of traffic management to be made” (Rijkswaterstaat, 2008a). The program aims to achieve this goal by supervising evaluation of DTM projects, standardizing evaluations when possible and gathering knowledge gained in these evaluations.

[8] In Dutch: “Landelijk evaluatieprogramma”

To be able to achieve the goals of this program, a suitable guideline for evaluating network-wide DTM projects is required. Based on experiences gained with evaluating local DTM projects, the "Guideline for model and evaluation studies of utilisation measures[9]" has been composed as the standard for evaluating DTM measures (both ex-ante and ex-post). Next to this the "Guideline for ex-post evaluation of utilisation measures on the main road network[10]" (Rijkswaterstaat Traffic Centre the Netherlands, 2008) has been written. This guideline is meant to perform a "quick ex-post evaluation" of DTM measures focussed on improving traffic flow. From now on "the current evaluation methodology" will be used to refer to these two documents.

In this methodology focus is on impacts of local DTM measures on traffic flow, on the main road network, in regular conditions. The following remarks can therefore be made about this methodology:

- The guidelines focus on impacts on *traffic flow*. How to determine impacts on for example road user behaviour, traffic safety, noise hindrance and emissions is not explicitly documented. The evaluator is referred to other documents.
- The methodology is meant for *impact* evaluation alone and do not include other evaluation types. The guidelines for example do not provide a framework for interpreting costs and benefits of DTM.
- The methodology is focused on impacts of *local* DTM on the *main road network* and does not offer guidance to determine impacts of network-wide DTM, on both the main and secondary road network.

The Centre for Transport and Navigation of Rijkswaterstaat has noted these possible shortcomings of the current evaluation methodology. Revision of the current evaluation guidelines is therefore incorporated as objective of the "National evaluation program".

Summarizing, the problem definition is formulated as follows:

Problem definition

A suitable evaluation guideline is required to gain insight into impacts of network-wide Dynamic Traffic Management (DTM) and to support decision-making around DTM. DTM can have impacts on for example traffic flow, travel time reliability, road user behaviour, noise hindrance, emissions and traffic safety. The current evaluation methodology of Rijkswaterstaat is however focused on evaluation of impacts of *local* DTM measures on traffic flow and may no longer be suitable to evaluate *network-wide* DTM projects.

[9] In Dutch: "Leidraad model- en evaluatiestudies benuttingsmaatregelen"

[10] In Dutch: "Richtlijn voor de ex-post evaluatie van benuttingsmaatregelen op het hoofdwegennet"

1.3 Research objective and research questions

Based on the problem definition the following research objective is formulated:

Research objective

> The objective of this research is to identify strengths, weaknesses, opportunities and threats of the current ex-post evaluation methodology for DTM of Rijkswaterstaat and to propose a new evaluation methodology

To be able to meet the research objective, the following research questions are answered in this report:
- What are strengths, weaknesses, opportunities and threats of the current evaluation methodology?
- To what extent do alternative evaluation methodologies offer an improvement for the identified weaknesses?
- What potential improvement offered in evaluation-related literature for the remaining weaknesses is most promising?
- How can this most promising improvement be incorporated into a proposal for a new evaluation methodology?
- Is it possible to apply the proposed evaluation methodology to a real-life, network-wide DTM project?
- What are strengths, weaknesses, opportunities and threats of the proposed evaluation methodology?

Answering these research questions leads to the answer of the central research question of this report:

Central research question

> How do strengths, weaknesses, opportunities and threats of the proposed evaluation methodology relate to those of the current evaluation methodology of Rijkswaterstaat?

The remainder of this introduction explains how these research questions are answered and what this research yields. A delineation of the research scope is presented in Section 1.4. Section 1.5 explains what scientific and practical relevance the resulting research has. The research approach is explained in Section 1.6. Finally an overview of the structure of the remainder of the report is presented in the reading guide of Section 1.7.

1.4 Research scope

Within the research questions the research scope is already defined to an extent. In this section the research scope will be delineated further.
- Strengths, weaknesses, opportunities and threats (SWOTs) of evaluation methodologies are not identified in readability (textual quality) or practicality (quality of the report structure) of evaluation guidelines.
- Research focus is on ex-post evaluation of DTM measures. The proposed evaluation methodology can however also be applied to ex-ante evaluation.

- Research focus is on roadside DTM measures and not on in-car or cooperative roadside/in-car systems. Research focus is thus on Intelligent Transportation Systems (ITS) for infrastructure and not on the following ITS clusters (Panou and Bekiaris, 2004):
    - ITS for private vehicles;
    - ITS for public transport;
    - ITS for commercial vehicles;
    - Cooperative infrastructure/in-vehicle systems.

  The research may however also provide insights for these ITS clusters.
- Focus of the methodology proposed in this report is on evaluation of DTM measures in regular situations. The methodology is however suitable for the evaluation non-regular situations as well.
- The proposed evaluation methodology will not yet include socio-economic evaluation. Further research is required to incorporate socio-economic evaluation into the proposed evaluation methodology. To this end Appendix A includes a discussion of methods for socio-economic evaluation of DTM projects.
- Data collection, data selection, data analysis and feedback of evaluation results are outside the scope of the methodology proposed in this report. Further research is required to incorporate these steps into the proposed evaluation methodology.

1.5 Research relevance

Achieving the objective of this research within the defined scope has both scientific and practical relevance, which is explained below.

Scientific relevance
This research:
- Provides insight into shortcomings of current DTM evaluation methodologies;
- Identifies and analyses potential solutions for these shortcomings;
- Introduces the theory-based evaluation approach to the field of DTM evaluation;
- Proposes a new evaluation methodology based on the theory-based evaluation approach.

Practical relevance
This research:
- Produces an overview of SWOTs of the current evaluation methodology. Based on these SWOTs recommendations are made for Rijkswaterstaat on when (not) to apply this methodology;
- Proposes an evaluation methodology that can be used as the basis of a new evaluation guideline for Rijkswaterstaat which is able to cope with current trends in DTM;
- Proposes an evaluation methodology and case study which provide a starting point for the "Field Test Amsterdam", an important project for the future of DTM in the Netherlands;
- Proposes an evaluation methodology that stimulates cooperation between decision-makers, project designers and evaluators;

- Proposes an evaluation methodology that may:
    - Increase insight into costs and benefits of DTM;
    - Assist the decision-making process around DTM;
    - Increase public support for DTM;
    - Increase the effectiveness of DTM measures.

1.6 Research approach

This section describes how the research questions defined in Section 1.3 are answered and how the scientific and practical relevance defined in Section 1.5 is achieved.

The research starts with a SWOT analysis of the current evaluation methodology of Rijkswaterstaat. Advantages of SWOT analysis over a listing of advantages and disadvantages are that it:
- Makes a distinction between attributes of the methodology under analysis (strengths and weaknesses) and attributes of its environment (opportunities and threats). This helps to identify possible sources of improvement.
- Requires a clear definition of the objective of ex-post evaluation. This helps to focus the identification of SWOTs.

An explorative case study, review of evaluation guidelines and literature on (DTM) evaluation are used to identify SWOTs.

Alternative evaluation methodologies are then studied to determine the extent in which they provide improvements over the current evaluation methodology of Rijkswaterstaat. Focus is on identifying improvements for weaknesses of the current evaluation methodology. A review of evaluation guidelines is used to identify these improvements.

The research then focuses at weaknesses not (sufficiently) improved in alternative evaluation methodologies. Literature on evaluation is used to identify and describe potential improvements for these remaining weaknesses. The "most promising" improvement is then selected by checking three criteria. This improvement is used in the proposal for a new evaluation methodology.

The proposed evaluation methodology uses a theory-based evaluation approach. A "logic model", a graphical representation of project theory, is chosen as the basis of the methodology. Literature on logic models and theory-based evaluation is used to design a step-by-step approach to design logic models and develop evaluation plans based on these logic models.

A SWOT analysis of the proposed evaluation methodology is then performed. A case study is used to determine whether it is possible to apply the proposed evaluation methodology to a real-life project. The case study and literature on theory-based evaluation and logic models are used to complete the SWOT analysis.

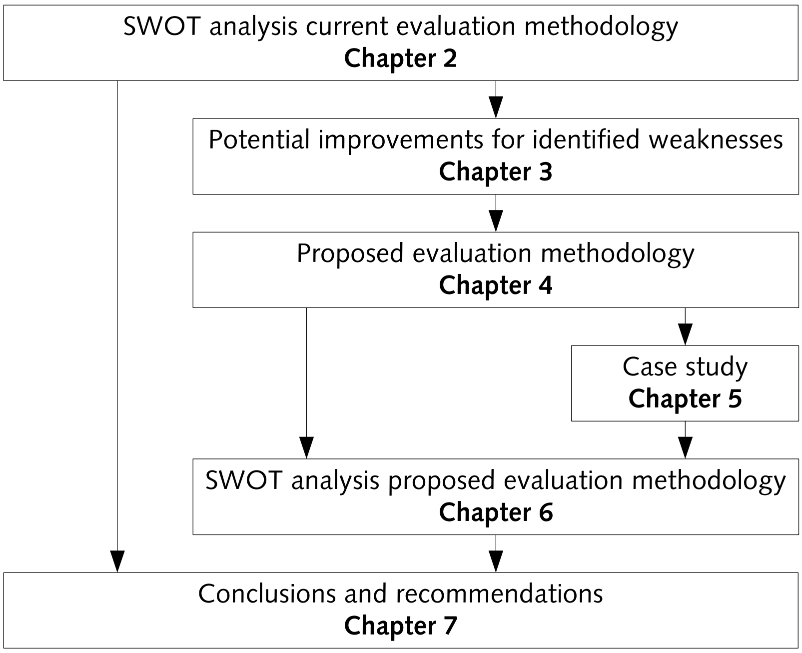
Finally the central research question is answered by comparing results of the SWOT analyses of the current and proposed evaluation methodologies.

1.7 Reading guide

The contents of this report follow the structure of the research approach described in Section 1.6. This section describes the report structure and contents of each chapter. First an overview of the report structure is given, followed by a description of contents of each chapter.

Research overview
Figure 1.1 graphically summarises the report structure.

Figure 1.1 Report structure



Chapter 2
In Chapter 2 a SWOT Analysis of the current evaluation methodology is performed. The chapter starts with defining the objective of ex-post evaluation in Section 2.2. Next, a description of the characteristics of the current evaluation methodology is given in Section 2.3. To give an indication of SWOTs, an explorative case study of the (attempted) application of this methodology to a network-wide DTM project is presented in Section 2.4. The case used is the ex-post evaluation of the project "Improving traffic flow on the A10 ring road[11]", part of the "FileProof" program. A comparison of characteristics of the methodology with literature on (DTM) evaluation is then made to further clarify strengths and weaknesses of the methodology in Sections 2.5 and 2.6. Finally, trends in DTM are described to identify opportunities and threats in Sections 2.7 and 2.8.

[11] In Dutch: "Verbeteren doorstroming ring A10"

Chapter 3
In Chapter 3 potential improvements for the identified weaknesses of the current evaluation methodology of Rijkswaterstaat are identified. First the extent to which three alternative evaluation methodologies offer improvements for these weaknesses is researched in Section 3.2. The remainder of the chapter will focus on the weaknesses not improved in these three studied evaluation methodologies.

For these remaining weaknesses three potential improvements are explored. First of all, three alternative experimental designs are discussed in Section 3.3. Secondly, statistical and modelling tools are described in Section 3.4. Finally, the theory-based evaluation approach is introduced in Section 3.5. The chapter ends with a selection of potential improvements that are used in the proposed methodology in Chapter 4.

Chapter 4
A new evaluation methodology, using a theory-based approach, is proposed in Chapter 4. The chapter starts with defining the scope of the proposal. Next, the logic model, the "core" of the methodology, is discussed in Section 4.2. Finally, a step-by-step approach is proposed to create a logic models for DTM projects and create evaluation plans based on these logic models in Sections 4.4 and 0. The approach is clarified with an example.

Chapter 5
The proposed evaluation methodology is applied to a real-life, network-wide DTM project in Chapter 5. The case used is the "Field Test Traffic Management Amsterdam[12]", a network-wide and coordinated application of DTM due to start in 2010. An introduction to this case is given in Section 5.2. The logic model for part of this project is developed in Section 5.3. This logic model is used for an evaluation approach in Section 5.4.

Chapter 6
In Chapter 6 a SWOT analysis of the proposed evaluation methodology is performed. The objective of ex-post evaluation defined in Chapter 2 will be used for this SWOT analysis again, allowing a fair comparison to be made between the current and proposed methodology in Chapter 7, the final Conclusions and Recommendations of this report. Strengths and weaknesses are described in Sections 6.2 and 6.3, opportunities and threats in Sections 6.4 and 6.5.

[12] In Dutch: "Praktijkproef Verkeersmanagement Amsterdam"

2. SWOT-analysis current evaluation methodology

2.1 Introduction

In this chapter, a SWOT-analysis is performed for the current evaluation methodology of Rijkswaterstaat. This chapter's central question is:

Research question of Chapter 2

What are strengths, weaknesses, opportunities and threats of the current evaluation methodology?

This Chapter results in:

- Strengths of the current evaluation methodology that are helpful to achieve the objective of ex-post evaluation;
- Weaknesses of the current evaluation methodology that are harmful to achieve the objective of ex-post evaluation;
- Opportunities for the current evaluation methodology that are helpful to achieve the objective of ex-post evaluation;
- Threats to the current evaluation methodology that are harmful to achieve the objective of ex-post evaluation;

See Table 2.1.

Table 2.1: SWOT analysis
Overview of the four aspects of a SWOT analysis. Based on (Wikimedia Foundation Inc., 2008)

	Helpful to achieve the objective of ex-post evaluation	**Harmful** to achieve the objective of ex-post evaluation
Internal origin (attributes of the methodology)	S	W
External origin (attributes of the environment)	O	T

The objective of ex-post evaluation is defined in Section 2.2. Next, characteristics of the object under SWOT-analysis, the current evaluation methodology, are described in Section 2.3. An explorative case study of the application of the current evaluation methodology to a network-wide DTM project is performed in Section 2.4. The explorative case study illustrates difficulties of ex-post evaluation of a network-wide DTM project to the reader and helps to identify SWOTs of the current evaluation methodology. The explorative case is the ex-post evaluation of the network-wide DTM project "Improving traffic flow on the A10 ring road[13]", part of the "FileProof" program.

Finally, SWOTs of the current evaluation methodology are described in more detail using literature on (DTM) evaluation. Three sources were used to identify SWOTs:

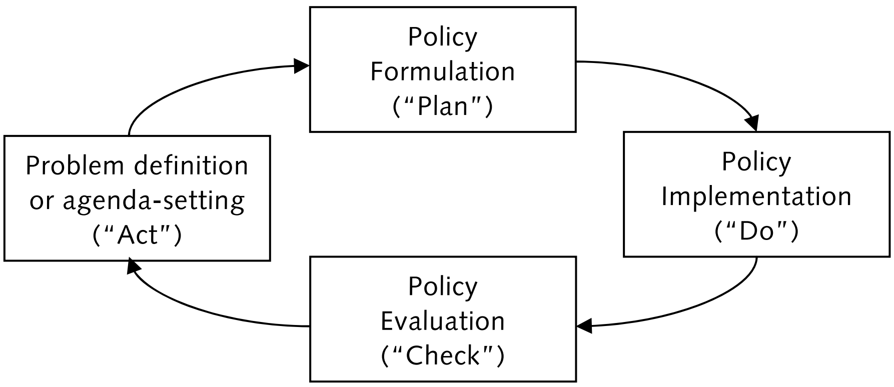
- Literature on DTM evaluation and evaluation in general;
- Difficulties encountered by the evaluation team of the explorative case study;
- A review of evaluation guidelines.

The chapter concludes with an overview of SWOTs of the current evaluation methodology. The next chapter describes potential improvements offered in literature for the identified weaknesses.

2.2 Objective of ex-post evaluation

To be able to state the objective of ex-post evaluation, which is needed to perform SWOT-analysis, the role that ex-post evaluation plays in policy-making needs to be known. In science of public administration, the process of agenda setting, policy design and implementation is often shown as a cyclic process: the policy cycle. An example of a basic representation of the policy cycle is shown in Figure 2.1.

Figure 2.1: Policy cycle
Basic representation of the policy cycle. The equivalent four policy phases used by Rijkswaterstaat are included between brackets.
Sources: (Palumbo, 1987), adapted, (Rijkswaterstaat Dienst Verkeer en Scheepvaart, 2008)



[13] In Dutch: "Verbeteren doorstroming ring A10"

Evaluation plays a role in each of these four phases of the policy cycle (Palumbo, 1987):

- In the agenda setting and problem definition phase, the objectives of evaluation are to determine or forecast size and geographical distribution of problems, to determine or forecast needs of stakeholders, and to identify target groups or areas.
- In the policy formulation phase, evaluation serves to identify alternative means to achieve policy objectives and to determine effectiveness of alternatives. Furthermore evaluation must assess the acceptance of alternatives by public and stakeholders. An evaluation in this phase and the previous phase is referred to as ex-ante evaluation.
- In the policy implementation phase, evaluation checks whether policy is implemented properly, i.e. in line with policy design. An evaluation in the implementation phase is called ex-durante or intermediate evaluation, within Rijkswaterstaat also referred to as monitoring.
- In the policy evaluation phase evaluation is referred to as ex-post evaluation, the evaluation type studied in this research.

An ex-post evaluation is thus necessary to "close" the policy cycle by connecting the policy implementation phase to the problem definition and agenda-setting phase. According to the "Guideline for model and evaluation studies of utilisation measures[14]" the objective of evaluation is "to support decision making in the future" (Rijkswaterstaat Adviesdienst Verkeer en Vervoer, 2002). Since it is hard to perform a SWOT-analysis using this broad definition, a more specific objective of ex-post evaluation is formulated for this research. The objective of ex-post evaluation, used to perform SWOT-analyses in this research, is:

Objective of ex-post evaluation of DTM projects
Adapted from (Giorgi and Tandon, 2002)

> Assessing the functioning and impacts of DTM measures and providing information on the use and allocation of public resources or the efficiency of DTM measures.

Ex-post evaluation thus includes a *learning* aspect and an *accountability* aspect. *Learning* deals with how and when DTM measures work and how future DTM measures can be improved. Traffic engineers will generally be most interested in this aspect. A*ccountability* deals with the extent in which benefits of DTM measures outweigh costs and whether such DTM measures should be considered again in the future. Decision-makers will generally be most interested in this aspect. Due to restrictions on evaluation resources, a trade-off usually has to be made between these two aspects.

[14] In Dutch: "Leidraad model- en evaluatiestudies benuttingsmaatregelen"

2.3 Current evaluation methodology

The object under SWOT-analysis is described in this section. This object is the current ex-post evaluation methodology for DTM projects of Rijkswaterstaat.

The current evaluation methodology aims at determining impacts of measures to increase road infrastructure utilisation, such as DTM measures. Focus is on determining direct, traffic related impacts of individual measures, in regular conditions, on the main road network.

The current evaluation methodology for DTM projects is documented in two guidelines. These guidelines are the "Guideline for model and evaluation studies of utilisation measures" and the "Guideline for ex-post evaluation of utilisation measures on the main road network[15]". From now on "Guideline for model and evaluation studies" and "Guideline for ex-post evaluation" will be used to refer to these documents.

This section starts with a description of these guidelines, followed by a description of the approach used to determine impacts of DTM measures.

Evaluation guidelines
The "Guideline for model and evaluation studies" (Rijkswaterstaat Adviesdienst Verkeer en Vervoer, 2002) is an aid for model studies (ex-ante evaluation) and evaluation studies (ex-post evaluation). It is not meant as a prescriptive guideline, but provides a frame of reference and suggestions for evaluators.

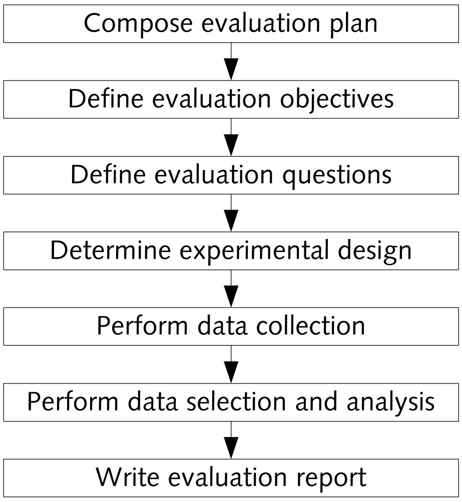
In this guideline both ex-ante and ex-post evaluation is considered supportive to the "Sustainable traffic management[16]" process. "Sustainable traffic management" is a procedure in which various road authorities (and other relevant stakeholders) work together on the basis of equality to design integral and sustainable traffic management solutions" (Rijkswaterstaat, 2002).

The "Guideline for model and evaluation studies" describes a step-by-step approach to ex-post evaluation. This step-by-step approach is illustrated in Figure 2.2.

[15] In Dutch: "Richtlijn voor de ex-post evaluatie van benuttingsmaatregelen op het hoofdwegennet"

[16] In Dutch: "GebiedsGericht Benutten (GGB)"

Figure 2.2: Steps in ex-post evaluation in "Guideline for model and evaluation studies"



How to perform these steps is not explicitly prescribed in the "Guideline for model and evaluation studies". Instead, considerations and suggestions are given to evaluators at each step.

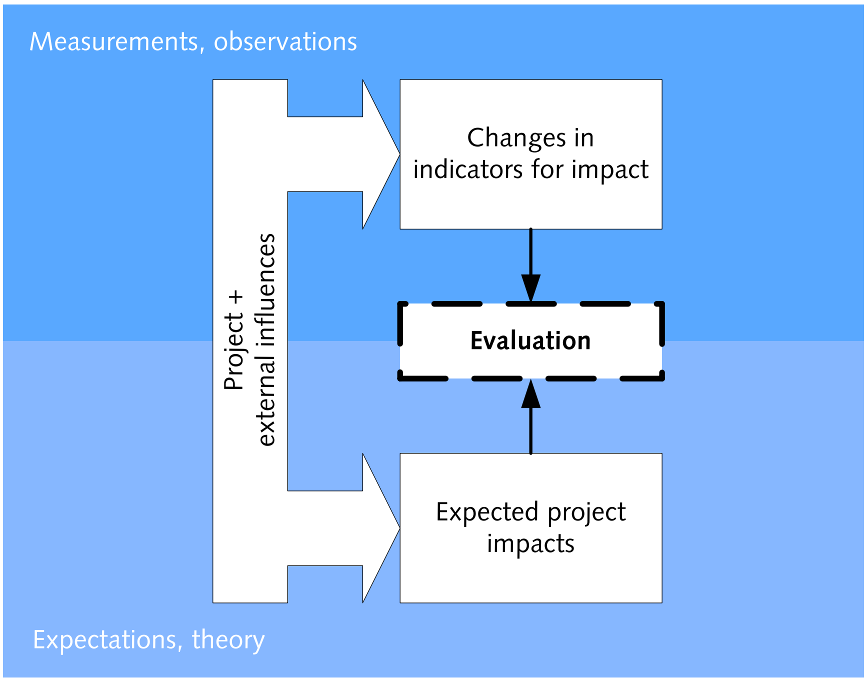
More prescriptive is the "Guideline for ex-post evaluation" (Rijkswaterstaat Traffic Centre the Netherlands, 2008). This guideline is focused on ex-post evaluation of traffic management measures aimed at improving traffic flow on the main road network. It is meant to perform a "quick evaluation" to estimate impacts of (groups of) DTM measures. To this end, it gives a detailed prescription for:

- Experimental design;
- Data collection;
- Data selection and analysis.

The evaluation approach to determine impacts of DTM described as "most common" in the "Guideline for model and evaluation studies" and explicitly prescribed in the "Guideline for ex-post evaluation" is described below.

Evaluation approach
To determine impacts of DTM measures, the current evaluation methodology determines changes of indicators for expected impacts of DTM measures. This is required since impacts of DTM measures on for example "accessibility" cannot be determined without a clear definition of this impact. Based on this definition, indicators for these impacts should be chosen. Measured changes in these indicators after realisation of DTM measures are an indication for impacts of DTM measures on "accessibility". See Figure 2.3 for an illustration.

Figure 2.3: Current evaluation approach



To determine these changes in indicators, the current evaluation methodology uses an experimental design called "before and after study[17]". This approach starts with determining values of indicators before DTM measures are implemented, in other words before the intervention in the traffic system is made. This is called before study. Next, after implementing the intervention, values of the same indicators are determined again. This is called after study. Based on differences in values of indicators between before study and after study, impacts of DTM measures can be estimated. When possible, statistical significance of impacts should be determined.

However, besides DTM measures, various external influences affect both actual impacts of DTM measures and value of indicators of these impacts. This is also included in Figure 2.3. Examples of these external influences are road works, incidents, weather conditions and spatial and economic developments. In the current evaluation methodology these external influences are referred to as "disturbing variables[18]".

[17] In Dutch: "voor- en nameting" of "nul- en éénmeting". The "Guideline for model and evaluation studies" does not explicitly prescribe a before and after study, but describes it as "most common".

[18] In Dutch: "verstorende variabelen"

In the current evaluation methodology a process called "matching"[19] is used to deal with these external influences. Impacts of external influences are removed from measurements as much as possible. To this end, measurements in which external influences have large impacts are removed from the dataset. This should result in datasets of before and after measurements in which impacts of (remaining) external influences is *equal*.

After removing measurements on which external influences had large impacts, a check is made whether the remaining impacts of external influences is comparable in before and after datasets. This is done by determining the number of vehicle kilometres travelled or traffic volume in networks, and checking whether they are comparable in before and after study.

Besides this "matching" approach the "Guideline for model and evaluation studies" mentions two possibilities for dealing with external influences:

- Measuring impacts of known external influences and reducing determined changes of indicators by these impacts. These measurements can also be done in a "control area" where DTM measures do not have impacts. Impacts of external influences can also be estimated based on previously gained knowledge.
- Alternating evaluation, in which DTM measures are switched on and off at random days. Impacts of external influences on before and after study can then be assumed equal.

Table 2.2 summarises the characteristics of the two evaluation guidelines of Rijkswaterstaat.

Table 2.2 Overview of characteristics of studied evaluation methodologies

	Guideline for model and evaluation studies	Guideline for ex-post evaluation
Evaluation focus	Ex-ante and ex-post impact evaluation	Ex-post impact evaluation
Considered impact groups of DTM	Focus on traffic flow. Safety, noise, emissions, user-acceptance mentioned but not explicitly described	Traffic flow
Experimental design to determine impacts	Various designs described. Before and after study "most common".	Before and after study
Approach to dealing with external influences	Various approaches described. Matching "most common".	Matching

[19] The "Guideline for model and evaluation studies" describes "matching" as "most common". The "Guideline for ex-post evaluation" prescribes matching.

2.4 Explorative case study current evaluation methodology

In this section an attempt to apply the evaluation methodology, which was described above, to a network-wide DTM project is studied. This explorative case study will illustrate SWOTs of the current evaluation methodology, which will be researched in more detail in the next section. The case studied is the ex-post evaluation of project "Improving traffic flow on the A10 ring road[20]", part of the "FileProof" program. From now on "FileProof A10" will be used to refer to this project. Innovative about this project is a network-wide approach to DTM design and coordination between DTM measures.

After an introduction of the project, its ex-post evaluation will be described. The section will conclude with an overview of problems encountered by the evaluation team while creating the evaluation plan based on the current evaluation methodology. This overview will help to find SWOTs of the current evaluation methodology in evaluation-related literature.

Introduction to the explorative case
The objective of "FileProof A10" is to design and implement a broadly supported set of DTM measures, as part of the "Network vision North-Holland[21]", which improves traffic flow of the Ring Road of Amsterdam (A10) by control of traffic flows (Rijskwaterstaat, Directie Noord Holland, 2007). In the "Network vision North-Holland" (Gemeente Amsterdam, Provincie Noord-Holland, Rijkswaterstaat, Stadsregio Amsterdam and Regionaal Orgaan Amsterdam, Not dated) Rijkswaterstaat, the Province of North-Holland, the Urban Region of Amsterdam and the Municipality of Amsterdam have comitted themselves to administrative and operational cooperation to increase accessibility of the region. An important part of this vision is to "keep the traffic on the Ring A10 flowing", as this will prevent spillback of congestion towards connecting motorways and urban roads.

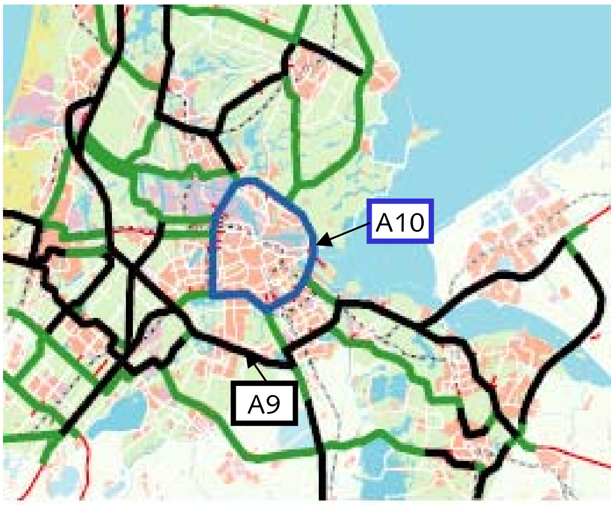
This network vision requires a prioritisation of the regional road network to be made. Traffic on roads with a high priority need to keep flowing, possibly using roads with a lower priority as a buffer for congestion. This should lead to "graceful degradation" of the road network in peak conditions, where low priority roads become congested first, and the highest priority road, Ring A10, remains flowing for as long as possible. The prioritisation made for the road network around Amsterdam can be seen in Figure 2.4.

[20] In Dutch: "Verbeteren doorstroming ring A10"

[21] In Dutch: "Netwerkvisie Noord-Holland"

Figure 2.4: Prioritisation of the road network around Amsterdam

According to the network vision North-Holland. Blue roads (A10 ring road) are priority 1, black roads priority 2, green roads priority 3. Source: (Gemeente Amsterdam et al, Not dated)



To be able to keep traffic on the Ring A10 flowing as much as possible, a control philosophy was designed in "FileProof A10". This philosophy comprises the following control strategies[22], approaches to alleviate the most important sources[23] of congestion on the A10:

- Rerouting of traffic "on a distance of the ring road". This can be done by rerouting traffic via other motorways such as the A9;
- Rerouting of traffic via non-congested sections of the A10 ring road;
- Optimising traffic flow on the A10 by small infrastructure and road design changes;
- Improving outflow of traffic from the A10 to urban roads;
- Improving outflow of traffic from the A10 to other motorways;
- When necessary controlling inflow of traffic onto the A10 from the four main axes in Amsterdam;
- When necessary controlling inflow of traffic onto the A10 from other urban roads;
- When necessary controlling inflow of traffic onto the A10 from other motorways.

To implement these four control strategies, a set of Dynamic Traffic Management measures was designed for both the main road network and the secondary road network. This set of DTM measures consists of:

- Installation of Dynamic Route Information Panels (DRIPs) to provide (route) information for road users;
- Infrastructure adaptations such as lengthening of merging and exit lanes;
- Expansion of monitoring installations, such as installing cameras on urban roads to measure travel times at important urban access routes and queue detection systems at on-ramps.
- Installation of ramp metering installations.

These DTM measures are implemented in phases, over a time-span of approximately a year. At the time of writing it is expected that the final component, the control software, will be operational in the final months of 2009.

[22] In Dutch: "regelstrategieën"

[23] In Dutch: "kiemen"

Besides a network-wide approach to DTM design, another innovative aspect is coordinated DTM control. The control software will allow ramp metering installations to be coordinated with urban traffic signals and other ramp metering installations.

Difficulties during the design of an evaluation plan

The ex-post evaluation of "FileProof" A10 is divided into three parts: collection of traffic data, an impact evaluation and a (road) user-acceptance evaluation. The case study focuses on the impact evaluation, since the current evaluation methodology is focused at impact evaluation. The objectives of the impact evaluation are determining impacts of DTM measures on:

- Vehicle delay hours[24];
- Travel times for (different classes of) road users;
- Travel time reliability.

These impacts should be determined for both the main road network and the secondary road network. In addition, a comparison is made between costs and benefits of the project by attaching a monetary value to the reduction of vehicle delay hours.

The impact evaluation is tendered to an external consultant. According to tender documents the ex-post impact evaluation should be carried out based on the current evaluation methodology (using the "Guideline for model and evaluation studies").

However, the consultant's evaluation team encountered difficulties when designing an evaluation plan based on the current evaluation methodology. The network-wide scale of "FileProof A10" and the phased implementation of DTM measures caused these difficulties. These difficulties will be described below.

The evaluation team expected it to be very difficult to determine (statistically significant) impacts at a network-wide level based on a single before and after study. The evaluation team identified several reasons for this:

- A result of the phased implementation is that after implementing all DTM measures, experience road users have with DTM measures varies per measure. Road users have had several months to understand some DTM measures and have adapted their behaviour accordingly, whilst road users are not yet familiar with other measures that were installed later.

[24] In Dutch: "voertuigverliesuren"

- Due to the long implementation period of DTM measures and the large scale of the study area the number and impacts of external influences increase. External influences include road works, events, the credit crunch and fuel price fluctuations. These increases of the number and impacts of disturbing influences have two effects. First of all, "removing" disturbing influences from the dataset will lead to a small dataset. Secondly, the likelihood that unknown external influences are influencing measured impacts increases. The likelihood that impacts caused by external influences are inadvertently attributed to DTM measures also increases.
- Due to the long implementation period, some external influences have lasting impacts on measured changes of indicators. Their impacts cannot be removed by removing measurements from the dataset.
- Expected impacts of DTM measures are small compared to impacts of external influences.
- Mobility management, economic growth, changes in fuel prices and possibly DTM measures themselves influence traffic volume during the time it takes to implement all measures. Origin-Destination patterns may have shifted. Checking whether all external influences have sufficiently been removed by comparing traffic volume in before and after study may not be feasible.

Another encountered difficulty is that measuring no or only minor impacts at a network-wide scale does not necessarily mean that no impacts are achieved (both positive and negative) at local scales. Local authorities are particularly interested in local impacts. Although the current evaluation methodology is focussed at determining these local impacts, it does not provide guidelines to distinguish between the various scales of impacts.

A potential solution for the two issues described above is to divide the network into sub-networks and determine impacts of DTM measures per sub-network. Per sub-network a different dataset can be selected for before and after study. Next, impacts per sub network can be summed-up to gain network-wide impacts. However, because datasets of before and after study are different for each sub-network, this may come down to "comparing apples and oranges".

2.5 Strengths of current evaluation methodology

The explorative case study illustrated several weaknesses of the current evaluation methodology. In Section 2.6 these weaknesses are defined in detail. Besides weaknesses the current evaluation methodology off course also has strengths. This section starts with a summary of these strengths, followed by explanation.

Strengths of the current evaluation methodology, that are helpful to achieve the defined objective of ex-post evaluation, are that it:

- Is based on a straightforward and intuitive approach;
- Is applicable to determine impacts on traffic flow of various projects without much adaptation;
- Has been applied frequently; experience is gained with applying the methodology;
- Connects evaluation to the policy cycle.

These strengths are explained below.

Straightforward and intuitive
The methodology used for impact evaluation is based on a before and after study, which according to Haight (2002) and Wholey, Hatry and Newcomer (1994) is a straightforward and intuitive approach. The approach measures indicators before and after implementation of DTM measures. Impacts of DTM measures are derived from resulting differences. This approach appeals to intuition. Before and after study is explicitly prescribed in the "Guideline for ex-post evaluation" and mentioned as "most common" in the "Guideline for model and evaluation studies".

Applicable to various projects without much adaptation
The methodology is applicable to determine impacts on traffic flow of various projects without much adaptation. The methodology can be applied irrespective of DTM design. The methodology does not place strong demands on project planning and monitoring. The only demand on realisation of DTM projects is that before study has to be done before actual realisation of DTM measures.

Experience gained
Experience has been gained with the methodology within Rijkswaterstaat and consultants. The approach has been used to evaluate many (local) DTM projects over the past two decades (see Rijkswaterstaat Adviesdiens Verkeer en Vervoer, 2002).

Evaluation connected to policy cycle
The "Guideline for model and evaluation studies" offers an aid for both ex-ante and ex-post evaluation of DTM projects. The role each evaluation type has in the policy cycle is explicitly defined, by defining the role of evaluation in the "Sustainable traffic management" process. This may increase usefulness of evaluation results

2.6 Weaknesses of current evaluation methodology

The explorative case study in Section 2.4 has shown that the current evaluation methodology has several weaknesses, which became apparent due to the network-wide scale of "FileProof A10" and the phased implementation of DTM measures. In this section these weaknesses will be described in more detail. The next section describes trends in DTM, identifying opportunities and threats to achieve the objective of ex-post evaluation.

Weaknesses of the current evaluation methodology that are harmful to achieve the defined objective of ex-post evaluation are that it:

- Does not analyse the functioning of DTM measures and road user response to them (technical and user-response evaluation);
- Does not provide information on the use and allocation of public resources (socio-economic evaluation);
- Does not focus on all relevant DTM impacts;
- Focuses on before and after study to determine DTM impacts. This experimental design is weak at proving causality between DTM measures and measured changes of indicators.

These weaknesses are explained below.

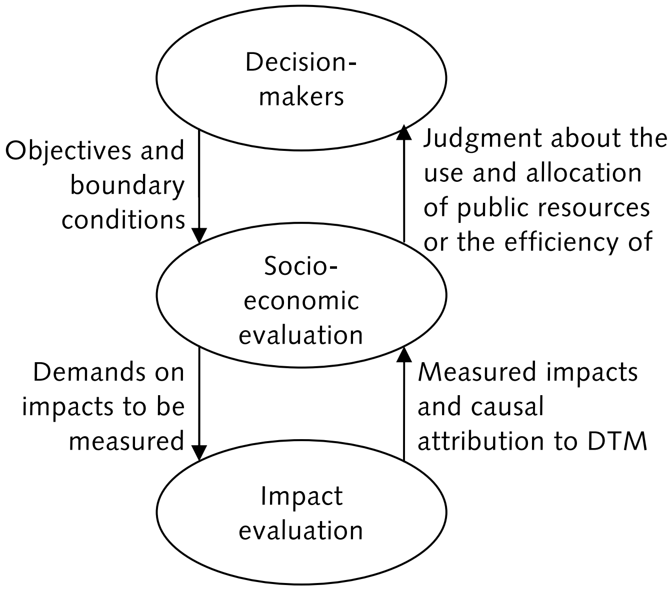
Does not include technical and user-response evaluation
The current evaluation methodology does not explicitly evaluate the the functioning of (individual) DTM measures. In Section 2.2 the objective of ex-post evaluation was defined as: "Assessing the functioning and impacts of DTM measures and providing information on the use and allocation of public resources or the efficiency of DTM measures". The current evaluation methodology meets only part of this objective. The methodology assesses "impacts of DTM measures", but does not explicitly assess their "the functioning". According to Hall, Miller and Khatak (1996) this problem is inherent to the before and after study experimental design: it does not provide insights into the functioning of (individual) DTM measures. The methodology does not show which DTM measures work and which do not, when they work and why.

For DTM measures it is not sufficient to assess whether their technical functioning is according to design specifications. Success of DTM measures also depends on (unpredictable) road user responses to DTM measures (Newman-Askins, Ferreira and Bunker, 2003). These user-responses should also be assessed. Two evaluation types are thus missing in the current evaluation methodology: technical evaluation and user-response evaluation.

Does not include socio-economic evaluation
The current evaluation methodology does not "provide information on the use and allocation of public resources or the efficiency of DTM measures". A socio-economic evaluation is required to meet this part of the objective of ex-post evaluation.

Socio-economic evaluation is strongly related to impact evaluation. It uses impacts and evidence of causal attribution of impact evaluation as input. On the other hand, socio-economic evaluation demands which impacts should be measured in impact evaluation. Figure 2.5 shows this relationship.

Figure 2.5: Relationship between decision-makers, socio-economic evaluation and impact evaluation



An ex-post socio-economic evaluation is required to show to what extent impacts generated by DTM projects compensate for invested public resources. On the one hand it allows "better" choices to be made about future allocations of public resources. On the other hand a socio-economic evaluation is essential to show that chosen DTM measures were superior over available alternatives and expenditures were thus justified (Moore and Pozdena, 2004). Information socio-economic evaluation provides may improve future decision-making about DTM.

Does not focus on all relevant DTM impacts
A third weakness of the current evaluation methodology is that it does not focus on all relevant DTM impacts. In order to perform socio-economic evaluation, all impacts of DTM measures need to be determined. Three reasons can be given why the current evaluation methodology does not focus on all relevant DTM impacts.

First of all the, current evaluation methodology is focused at impacts of DTM in regular conditions. However, next to impacts in regular conditions, it is also relevant to research impacts of DTM in non-regular conditions (Ministry of Transport, Public Works and Water Management, 2008).

Secondly, the methodology focuses at impacts of DTM on traffic flow, and not for example on road user behaviour, traffic safety or emissions. The "Guideline for model and evaluation studies" includes these impact types as relevant impact types of DTM, but does not explicitly describe how to determine these impacts.

Thirdly the methodology is focused at direct impacts of DTM. A complicating factor in DTM evaluation is that DTM impacts evolve over various timescales (Ministry of Transport, Public Works and Water Management, 2008). Direct impacts of DTM measures are for example an improvement of traffic flow and a reduction of congestion. These direct impacts also lead to indirect impacts, since road users adapt to new situations. These indirect impacts consist of changes of destination, route, departure time and mode choices of road users. A new equilibrium is formed. On the long run this new situation may even affect spatial-economic development. In the new equilibrium congestion in networks can be at the same level as before implementation of DTM measures. However, more trips were made possible, which have an economic value.

This third reason can be overcome by selecting appropriate indicators for impacts of DTM measures. The "Guideline for model and evaluation studies" does not explicitly mention this issue in the selection of indicators. The "Guideline for ex-post evaluation" does recommend several indicators, all of which are focused on direct impacts of DTM measures.

Weak at proving causality between DTM measures and measured changes of indicators

The current evaluation methodology is strictly an impact evaluation, which should provide two insights (BandA Group, not dated). Firstly, an ex-post impact evaluation shows the extent in which objectives have been achieved. Secondly, an ex-post impact evaluation tries to explain why impacts have been achieved, including the extent in which results can be causally attributed to DTM measures. Causal attribution refers to the problem of how and how safely one may infer that DTM measures are responsible for measured changes in indicators.

It is hard to prove causal attribution of measured changes in indicators to DTM measures using a before and after study experimental design (Haight, 2002; Hall et al, 1996; Rossi, Lipsey, and Freeman, 2004; Wholey et al, 1994). Therefore sole reliance on a before and after study is often not justified. Haight (2002) describes it as follows: "A principal difficulty is that the traffic system after the intervention is usually different from the system before the intervention, besides impacts of the measures being evaluated. It is usually difficult, if often impossible, to separate the various influences on evaluation outcomes." The current evaluation methodology mostly relies on before and after study, so the possibility that a measured change in indicators is wrongly attributed to DTM measures is present.

This weakness is worsened by two factors. First of all, as Newman-Askins et al (2003) and Hall et al (1996) point out, impacts of DTM measures are likely to be small compared to impacts of external influences. And secondly, the functioning of DTM measures and road user responses to these measures are not evaluated. The possibility exists that impacts are attributed to measures that in fact do not function as expected. Without proving causality and evaluating the functioning of DTM measures and responses of road users to them, it will often be incorrect to extrapolate impacts measured in one location to another location (Newman-Askins et al, 2003).

This weakness will become more and more apparent when the number and impacts of external influences increase. Nevertheless, according to the "Guideline for model and evaluation studies" the current evaluation methodology can also be applied for evaluation of groups of DTM measures. It does state that when evaluating network-wide DTM, "determining statistically significant effects will become harder […], since differences between before and after study is influenced by more factors as study areas grow" (Rijkswaterstaat Adviesdienst Verkeer en Vervoer, 2002).

2.7 Opportunities for current evaluation methodology

Opportunities for the current evaluation methodology can be exploited to emphasise strengths or reduce threats and weaknesses. In this section one such opportunity is described. This opportunity for the current evaluation methodology that is helpful to achieve the defined objective of ex-post evaluation is that it:

- Can be adapted to include other impacts.

These opportunities are explained below.

Can be adapted to include other impacts
The approach used in the methodology is in principle not restricted to impacts of DTM on traffic flow. If suitable indicators for traffic safety, road user behaviour and emissions are formulated, the before and after study approach is applicable to determine these impacts as well.

It is, however, difficult to adapt the methodology to determine impacts of DTM in non-regular conditions, such as incidents or events. Since each non-regular situation is unique, finding comparable non-regular situations for before and after study is a difficult task.

2.8 Threats to current evaluation methodology

In the remainder of this chapter, threats to the current evaluation methodology are described. This section shows that three trends in DTM will make weaknesses of the current evaluation methodology more apparent. These trends can thus be considered threats to achieve the objective of ex-post evaluation using the current evaluation methodology.

These threats are:
- Diversification of objectives of DTM;
- Increasingly network-wide approach to DTM design;
- Increasing coordination between DTM measures.

Below an explanation is given.

Diversification of objectives of DTM

A trend is to use DTM to achieve different objectives than improving traffic flow in regular conditions. Air quality and safety are no longer only seen as boundary conditions for DTM. Reducing emissions or increasing traffic safety are now seen as objectives of DTM, see for example (Van Leusden, Bakker, Hepp, and Vreeswijk, 2008). DTM is also applied to optimise traffic safety and traffic flow in non-regular conditions, such as road works, events and accidents. Some authors argue that in these fields the largest impacts of DTM can be achieved (Immers, Meurs, Schuurman, Van Berkum, Van Kooten and Van Wee, 2008).

For two reasons this diversification of objectives is a threat to the current evaluation methodology. First of all, the current evaluation methodology is focused on impacts on traffic flow in regular conditions. Secondly, the need for socio-economic evaluation increases. Some form of trade-off between for example impacts on traffic flow, safety and air quality impacts will need to be made. Furthermore, some of these impacts will affect road users, while others will impact non-users (Newman-Askins et al, 2003; Zhang, Kompfner, White and Sexton, 1998). Socio-economic evaluation is needed to make trade-offs between impacts and impact groups.

Increasingly network-wide approach to DTM design

A second trend is that DTM measures are designed using a network-wide approach instead of a local approach. Both the main and secondary road networks are included in this network-wide approach. This requires coordination between road authorities (and other stakeholders) and their objectives. This coordination is required since Dutch road infrastructure management is fragmented; Rijkswaterstaat manages the main road network, the provinces are responsible for the provincial roads and the municipalities deal with the remaining secondary road network.

To streamline the process network-wide DTM design, the process of "Sustainable traffic management" has been introduced in the Netherlands to serve as a guideline. "Sustainable traffic management" is a procedure in which various road authorities (and other relevant stakeholders) work together on the basis of equality to design integral and sustainable traffic management solutions" (Rijkswaterstaat, 2002).

The scope of DTM projects increases due to the network-wide DTM design approach. Realisation periods increase due to the number and complexity of DTM measures installed. Geographical scope naturally increases as well.

This increased scope, both in time and over the network, increases the number and impact of external influences on DTM projects. For example, the resulting larger study area of evaluation increases the number of incidents and events that affect evaluation. The larger separation in time of before and after measurements increases impacts of long-term influences such as economic developments, mobility policies and changes in origin-destination patterns.

The weakness at proving causality between measured changes in indicators and DTM measures will become more apparent. As concluded in section 2.6, this weakness is inherent to before and after study. Due to the network-wide DTM design approach the number and impact of external influences increase. Therefore, the likelihood that changes in indicators not caused by DTM measures are attributed to these DTM measures also increases.

The need for socio-economic evaluation also increases due to the network-wide DTM design approach. Two reasons can be given:

- The number of stakeholders involved in DTM projects and therefore organisational complexity increases (Baggen, Nuijten and Van Luling, 2004). The increased number of stakeholders involved leads to an increase in objectives and interests. These objectives may differ from those of Rijkswaterstaat and the methodology should thus be able to take these into account. Furthermore the need for evaluation to show whether investments in DTM are worthwhile increases.
- Impacts of network wide DTM will appear over different geographical and timescales. Impacts will range from local impacts to corridor wide or network-wide improvements and from incident-specific benefits to benefits which take months or even years to materialize (Thill, Rogova and Yan, 2004). Socio-economic evaluation is required to take these geographic and timescales of DTM impacts into account.

Increasing coordination between DTM measures

Another trend is towards coordinated control of DTM measures, instead of isolated control of individual DTM measures. For example, Kotsialos, Middelham and Papageorgiou (2005) describe an algorithm for coordination of ramp metering. Besides coordination between single types of DTM measures (in this case ramp-metering) advances have been made in coordination between different DTM measures. For example, Kamel, Benasser, and Jolly (2008) describe a control approach for the combined coordination of ramp-metering and variable speed limits.

Coordination between DTM measures increases the need for technical and user response evaluation. As systems become more complex, it will become more important to evaluate what works, when it works and why. Impacts of coordinated DTM may not be as expected because a single element does not function as expected, or road users do not respond to measures as expected. When performing only impact evaluation these causes may not be found.

2.9 Conclusions

In this chapter the following research question was answered:

Research question of Chapter 2

What are strengths, weaknesses, opportunities and threats of the current evaluation methodology?

Using the policy cycle, the following objective of ex-post evaluation was selected to identify SWOTs of the current evaluation methodology:

Objective of ex-post evaluation of DTM projects

Assessing the functioning and impacts of DTM measures and providing information on the use and allocation of public resources or the efficiency of DTM measures.

The identified SWOTs of the current evaluation methodology are summarized in Table 2.3.

Table 2.3: SWOT analysis of current evaluation methodology
Based on (Wikimedia Foundation Inc., 2008)

	Helpful to achieve the objective of ex-post evaluation	**Harmful** to achieve the objective of ex-post evaluation
Internal origin (attributes of the methodology)	S • Straightforward, intuitive • Applicable to various projects with little adaptation • Experience gained • Evaluation connected to policy cycle	W • No technical and user-response evaluation • No socio-economic evaluation • Does not take all irrelevant impacts of DTM into account • Weak at proving causality DTM measures – measured changes indicators
External origin (attributes of the environment)	O • May be adapted to evaluate other impacts	T • Diversification of objectives • Increasingly network-wide approach to DTM design • Increasing coordination of DTM measures

Current trends in DTM will make the identified weaknesses of the current evaluation methodology more apparent in the future. The only way to mitigate these threats is to improve the current evaluation methodology. In the next chapter, potential improvements for the identified weaknesses will be studied. A new evaluation methodology will be proposed in Chapter 4 based on findings of Chapter 3.

3. Potential improvements evaluation methodology

3.1 Introduction

In the previous chapter weaknesses of the current evaluation methodology were identified. Furthermore, it was shown that current trends in DTM are threats for achieving the objective of ex-post evaluation with the current methodology. The need to improve weaknesses of the current evaluation methodology therefore increases.

In this chapter, potential improvements for weaknesses of the current evaluation methodology are identified. The most promising improvement is used in the proposed evaluation methodology in Chapter 4. This chapter provides an overview of the state-of-the art in DTM-evaluation. Furthermore, it identifies weaknesses for which insufficient improvements are currently offered in DTM-related literature.

The identification of potential improvements starts with DTM evaluation methodologies developed in other countries. In the first section three of these alternative methodologies are described. After this description sub-conclusions about the extent in which these three alternative methodologies offer improvements for the identified weaknesses are drawn. The following question is thus answered:

Research question 1 of Chapter 3

To what extent do alternative evaluation methodologies offer an improvement for the identified weaknesses?

The research then focuses on the remaining weaknesses. In Section 3.3, alternative experimental designs are discussed. Next, statistical and modelling methods are briefly discussed. Finally in Section 3.5 an alternative approach to evaluation, called theory-based evaluation, is introduced.

This chapter concludes with determining which potential improvement for the remaining weaknesses is most promising. The criteria to be used to determine which improvement is "most promising" are given after the next section. Scoring the potential improvements on these criteria answers the following question:

Research question 2 of Chapter 3

What potential improvement offered in evaluation-related literature for the remaining weaknesses is most promising?

In Chapter 4 the chosen potential improvement is applied in a proposal for a new evaluation methodology. Within the scope of this report socio-economic evaluation will not be included in this proposal. However, in Appendix A a discussion about the applicability of methods for socio-economic evaluation to DTM projects is presented.

3.2 Improvements in existing evaluation methodologies

The extent in which alternative evaluation methodologies offer improvements for the current evaluation methodology is studied in this section. Three alternative evaluation frameworks are described and analysed:

- The European CONVERGE methodology.
- The Canadian ITS evaluation methodology.
- The Finnish ITS evaluation methodology.

These three methodologies were selected from the available methodologies worldwide since they have a spread in origin (EU country, EU, non-EU country) and age (1998, 2002 and 2007) and may therefore provide a good overview of knowledge gained over various countries through the years. This section starts with an introduction to these three methodologies, after which conclusions are drawn about the extent in which they provide improvements for the weaknesses of the current evaluation methodology of Rijkswaterstaat.

CONVERGE evaluation methodology
As part of the CONVERGE evaluation methodology a "Guidebook for assessment of transport telematics applications" (Zhang et al, 1998) has been developed. It gives general guidance and recommendations for both ex-ante evaluation (called "assessment" in the guidebook) and ex-post evaluation (called "validation" in the guidebook) of ITS projects.

The CONVERGE evaluation methodology consists of seven steps, shown in Figure 3.1. Differences with the current evaluation methodology of Rijkswaterstaat are the explicit inclusion as separate evaluation steps of:

- A definition of users and stakeholders needs[25];
- A pre-evaluation of expected impacts to focus the actual evaluation.

[25] These steps are indirectly included in the "Guideline for model and evaluation studies" by regarding evaluation as supportive to the "Sustainable traffic management" process.

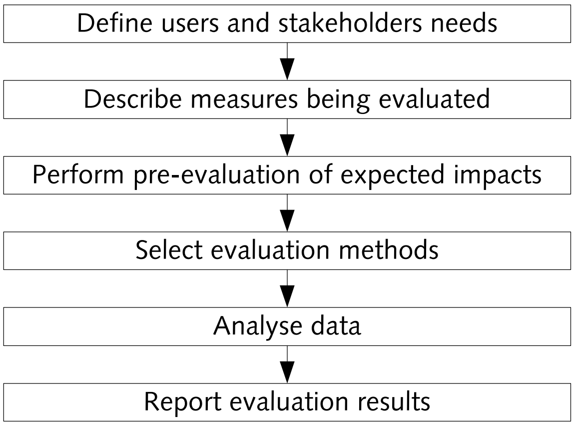


Figure 3.1: Steps in ex-post evaluation in CONVERGE evaluation guidebook

Where the current evaluation methodology of Rijkswaterstaat focuses on impact evaluation, the CONVERGE guidebook lists six types of evaluation[26]:

- Technical evaluation, which evaluates whether implemented systems function according to technical specifications;
- Impact evaluation, which focuses on impacts on for example safety, environment, transport efficiency, user behaviour and modal split.
- User-acceptance evaluation, which focuses on preferences and opinions of (road) users;
- Socio-economic evaluation, which focuses on (societal) costs and benefits. The CONVERGE guidebook lists two possible methods for this type of evaluation: Cost-Benefit Analysis and Multi-Criteria Analysis. It does not state a preference for either method;
- Financial evaluation, in which several financial parameters of projects are determined such as initial and running costs, rate of return and the payback period.

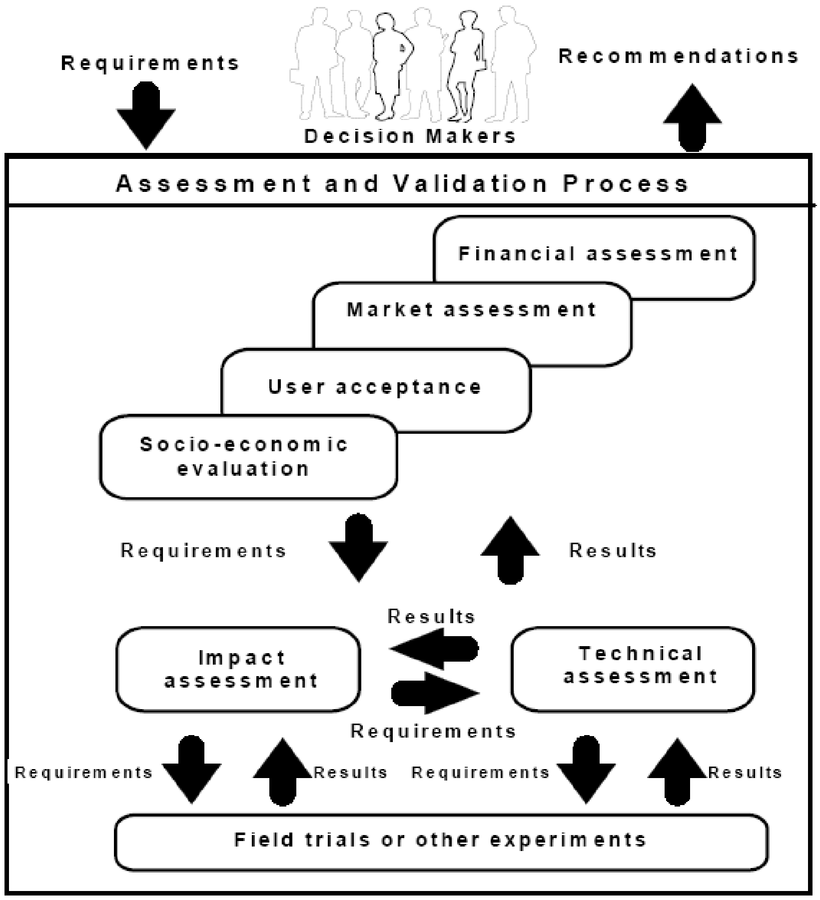
The CONVERGE methodology offers several improvements over the current evaluation methodology of Rijkswaterstaat. Firstly, the use and allocation of public resources is analysed in socio-economic evaluation. Secondly, a wider range of DTM impacts is taken into account, although only in regular conditions. Thirdly, it analyses the functioning of individual measures and road user responses.

A weakness of the methodology is that each evaluation type is mostly considered separately and is not explicitly linked to others types. An exception is the explicit link between impact and socio-economic evaluation. Furthermore the CONVERGE guidebook does state that the six evaluation types can be interrelated. See Figure 3.2 for an overview of the six evaluation types and some possible interrelations. However, what these interrelations are, how to deal with them and when to perform which evaluation type is not made explicit.

[26] Referred to as "assessment objectives" in the CONVERGE guidebook

This has two implications. First of all, information of technical and user-response evaluation is not explicitly used to support conclusions of impact evaluation. If this information is used, it can be made more plausible that measured changes in indicators are attributed to DTM measures. Secondly, not explicitly linking the various evaluation types may reduce the potential to learn from evaluation. It will be harder to determine what works, when it works, what it causes and why.

Figure 3.2 Evaluation types included in the CONVERGE guidebook
The figure displays the six evaluation types of the CONVERGE methodology and shows some possible interrelations between evaluation types.
Source: (Zhang et al, 1998)



The approach of the CONVERGE methodology to impact evaluation can be classified as before and after study. This approach to impact evaluation can be summarised as follows:

- Select appropriate indicators for expected impacts;
- Define the reference case for comparison;
- Determine data collection methods;
- Define an approach to “matching” conditions of the reference and study case as much as possible;
- Statistical considerations to determine the required number of measurements (or simulation runs);
- Define the measurement or simulation plan;
- Check the integrity of measurements or simulation runs.

This approach is similar to the before and after study used in the current evaluation methodology of Rijkswaterstaat. The weakness at proving causality between DTM measures and measured changes in indicators thus remains.

Canadian ITS evaluation methodology

The second methodology studied is the Canadian ITS evaluation guideline (Bruzon and Mudge, 2007). This methodology consists of an impact evaluation and socio-economic evaluation and is mainly focused on ex-post evaluation. Socio-economic evaluation consists of cost-benefit analysis.

The Canadian evaluation guideline describes an approach summarised in Figure 3.3. The steps are similar to the current evaluation methodology of Rijkswaterstaat. Main differences are the inclusion of expected impacts definition and Cost Benefit Analysis as separate steps.

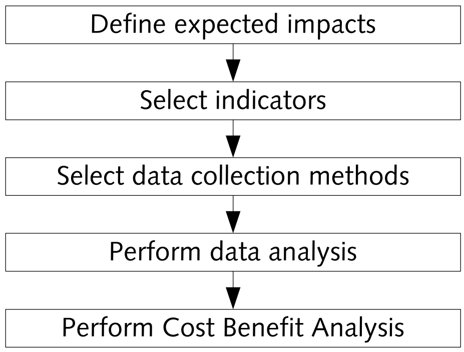


Figure 3.3: Steps in ex-post evaluation in Canadian ITS evaluation methodology

For impact evaluation six types of benefits of ITS are taken into account:
- Safety;
- Mobility;
- Efficiency and Productivity;
- Energy and the Environment;
- Security;
- Customer Satisfaction.

The experimental design used to determine these impacts is a before and after study. How to determine impacts of the ITS-project based on the before and after datasets and how to deal with external influences is not explicitly described in the Canadian ITS evaluation guideline.

Two improvements for weaknesses of the current evaluation methodology can be identified. The Canadian ITS evaluation guideline takes a wider range of impacts into account compared to the current evaluation methodology of Rijkswaterstaat. It also provides information on the use and allocation of public resources via the Cost-Benefit Analysis.

However, the Canadian ITS evaluation methodology also:
- Is weak at proving causality between DTM measures and measured changes of indicators. This weakness is even more apparent than in the current evaluation methodology of Rijkswaterstaat since it does not explicitly take external influences into account.
- Does not provide information on the functioning of individual measures and road user response to them.

Finnish ITS evaluation guideline

Similarly to the CONVERGE guideline, the Finnish guideline for ITS evaluation (Finnish Ministry of Transport and Communications, 2002) takes various evaluation types into account. Besides impact evaluation the guideline also includes:

- Impact evaluation, which includes a broad description of potential impacts of ITS;
- Socio-economic evaluation, consisting of a Cost Benefit Analysis within a qualitative framework;
- Legal and institutional evaluation;
- Technical evaluation;
- Market evaluation;
- Human-machine interface evaluation.

The methodology is applicable for both ex-ante and ex-post evaluation of ITS projects.

The approach of the guideline is presented in Figure 3.4. Main differences compared to the current evaluation methodology of Rijkswaterstaat are the inclusion of defininig expected impacts and socio-economic evaluation as separate steps.

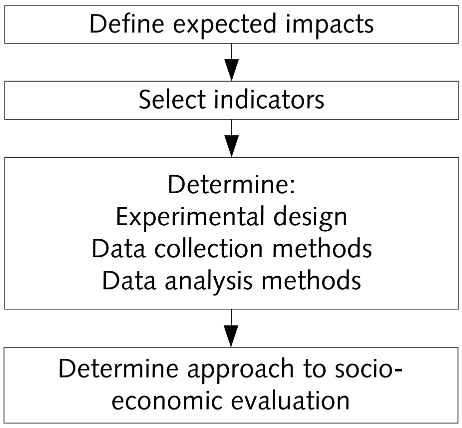


Figure 3.4: Steps in Finnish ITS evaluation methodology

The experimental design used for impact evaluation is a before and after study. After study is performed twice, shortly after implementation of DTM measures and a year after implementation. The approach to dealing with external influences is a comparable "matching" approach as used in the current evaluation methodology of Rijkswaterstaat.

It can be concluded that the Finnish ITS evaluation methodology offers similar improvements as the CONVERGE methodology. It also takes a wider range of impacts into account and includes technical, user-response and socio-economic evaluation.

Its weaknesses are also similar to the CONVERGE methodology. The first weakness is that it does not state when what type of evaluation should be performed and how the evaluation types are interrelated. Secondly, impact evaluation is based on the same before and after study approach as the current evaluation methodology of Rijkswaterstaat. The weakness at proving causality between DTM measures and measured changes of indicators remains.

Conclusions about existing evaluation methodologies
Table 3.1 summarises the characteristics of the three studied alternative evaluation methodologies. For comparison with the current evaluation methodology of Rijkswaterstaat, the two evaluation guidelines of Rijkswaterstaat are included. The following question can now be answered by using Table 3.1:

Research question 1 of Chapter 3

To what extent do alternative evaluation methodologies offer an improvement for the identified weaknesses?

The following improvements for the weaknesses of the current evaluation methodology of Rijkswaterstaat can be identified in the three studied alternative methodologies:
- The studied alternative methodologies take a wider range of impacts into account, although only under regular conditions. Especially the Finnish evaluation methodology includes an extensive listing of potential impacts of ITS and relevant indicators.
- All three studied evaluation methodologies provide information on the use and allocation of public resources via a socio-economic evaluation. Nevertheless, there is no consensus about preferred methods for socio-economic evaluation. In Appendix A a discussion of the suitability of three methods for socio-economic evaluation is presented.
- The Finnish and CONVERGE methodologies provide information on the functioning of individual DTM measures and road user responses. These methodologies include technical and user-response evaluation.

Two weaknesses remain in the studied alternative methodologies:
- The studied evaluation methodologies do not make interdependencies between the various evaluation types and various impacts explicit. The potential to learn from evaluation is not fully exploited therefore.
- The weakness at proving causality between DTM measures and measured changes of indicators is also present in the three studied evaluation methodologies. Two causes for this can be given. Firstly, impact evaluation in the three studied evaluation methodologies is based on a similar before and after study approach as used in the current evaluation methodology. The CONVERGE guidebook and Finnish guideline use a similar before and after study to impact evaluation. The Canadian guideline does not take external influences explicitly into account at all. Secondly, information of technical and user-response information is not explicitly used to support conclusions of impact evaluation.

Table 3.1 Overview of characteristics of studied evaluation methodologies

	Guideline for model and evaluation studies	Guideline for ex-post evaluation	Canadian ITS evaluation methodology	European CONVERGE methodology	Finnish ITS evaluation methodology
Evaluation focus	Ex-ante and ex-post	Ex-post	Ex-post	Ex-ante and ex-post	Ex-ante and ex-post
Evaluation types included	• Impact	• Impact	• Impact • Socio economic • Financial	• Impact • Socio-economic • Financial • Technical • User-acceptance	• Impact • Socio-economic • Financial, legal and institutional • Technical • Human-machine interface and interaction • Market
Socio-economic evaluation	None	None	CBA	CBA or MCA	CBA within qualitative framework
Technical, user-response evaluation	No	No	Yes	No	Yes
Evaluation types explicitly linked	N.A.	N.A.	No	No	No
Considered impact groups of ITS	• Focus on traffic flow • Safety noise and emissions mentioned but not explicitly described	• Traffic flow	• Safety • Mobility • Efficiency and productivity • Energy and the environment • Security • Customer satisfaction	• Safety • Environmental conditions • Transport efficiency	• Transport network and costs • Fleet and its costs • Accessibility • Travel time and its predictability • Traffic safety • Noise, emissions and energy • Valuations and comfort
Experimental design to determine impacts	Various designs described. Before and after study "most common.	Before and after study	Before and after study	Before and after study	Before and after study
Approach to dealing with disturbing variables	Various approaches described Matching[27] "most common".	Matching	Matching	Matching	Matching

[27] See Section 2.3 for a description of the `matching` process.

In the following sections potential improvements that address one or both of these weaknesses are discussed. In the conclusions the most promising improvement is selected to be incorporated into the proposal for a new evaluation methodology in Chapter 4. The potential improvement that is "most promising" meets three criteria:

- Does it increase confidence about causality between DTM measures and measured changes of indicators?
- Does it increase the potential to learn from the evaluation: does it help to determine what works, when it works, what it causes and why?
- Is it in principle applicable to any DTM project?

3.3 Alternative experimental designs

Three alternative experimental designs are discussed which may provide insights into the causality between DTM measures and measured changes of indicators. Three alternative evaluation designs could be improvements over before and after study: before and after study with a control area, alternating evaluation and trend- or timeline analysis (Grontmij Traffic and Infrastructure, 2002; Haight, 2002; Hall et al, 1996).

Control area
The first possible improvement is to include a control area in before and after study. The researched area with DTM measures is compared with a network without DTM measures. In the same time-period, a before study and after study of the same indicators are done for both the study area and the control area. By also determining changes of indicators in the control area, impacts of external influences (such as economic trends and fuel price changes) can be recognised. Measured changes of indicators in the study area can then be corrected for these influences.

This experimental design is an approximation of the design that would provide most confidence about causality between DTM measures and measured changes of indicators: "random assignment". In random assignment DTM measures are only applied to a randomly selected group of road users. A comparable test and control group are formed due to this random assignment on the same network. In DTM projects with road-side measures this experimental design is not feasible.

Although the control area seems a logical and useful addition to before and after study, in reality of DTM projects it is hardly feasible (Grontmij Traffic and Infrastructure, 2002; Hall et al, 1996). It is hard to find a road section with a comparable infrastructure supply (road section capacity, but also availability of alternative routes) and comparable traffic volume (including similar composition of vehicle types and through/local traffic), not to mention how hard it is to find a road *network* that will allow a fair comparison between the control area and study area.

Alternating evaluation
A second possible alternative to before and after study is alternating evaluation, which (Haight, 2002) describes as the "week on/week off approach". In this approach DTM measures will only be "turned on" at a randomly selected number of days, thereby approximating random assignment.

This approach has an important advantage over before and after study. This advantage is that long-term and lasting impacts of external influences will be present in both before and after study datasets. For example, when fuel price increases or road works occur after before study is performed, the comparison between before and after study is distorted. In alternating evaluation this would not distort the comparison since impacts of external influences will be present in both datasets.

Practical considerations will nonetheless also limit the applicability of this experimental design. Firstly, public authorities will need to be convinced of the added value of the approach. They will have to accept that after a long period of decision-making "their" DTM measures are finally operational, only to see them deliberately switched off at random days. Secondly, road users may notice the approach, which may lead to reduced effectiveness of DTM measures (road users start to ignore DTM measures, since they tend to be "turned off" quite often anyway) or may lead to worsening of public opinion of DTM measures (road users may feel the measures are a waste of public funding since they seem to be "broken" 50% of the time). Thirdly, road users need a certain "habituation period" to get used to DTM measures, which may not be available anymore if alternating evaluation is used. Alternating evaluation is thus most applicable to measures that are not clearly visible to road users (Rijkswaterstaat Adviesdienst Verkeer en Vervoer, 2002).

Trend or timeline analysis
The third alternative experimental design is trend or timeline analysis. This would require an extended before study and plotting of values of indicators on a timescale. This will allow long-term trends in indicators to be recognized. DTM measures should then create some form of breaches in these trends; the sizes of these breaches indicate impacts of DTM measures. The sizes of these breaches can be determined by mathematically describing trends and using this to predict values of indicators at the time of after study. Differences between measured values of indicators and predicted values would then be more accurate estimates for impacts of DTM measures.

This approach has two drawbacks. Trend breaches with other causes than DTM measures may not be recognised as such and may wrongfully be attributed to DTM measures. Furthermore, the approach requires an extended before study and after study, possibly over multiple years. This may not be feasible.

Conclusions about alternative experimental designs
Each of the three experimental designs studied in this research may reduce the likeliness that effects are wrongfully attributed to DTM measures. Besides this advantage, each of the experimental designs studied has drawbacks or practical limitations. Therefore, they are not applicable to any DTM project. Furthermore, the alternative experimental designs studied do not improve the potential to learn from evaluation. By changing the experimental design, one does not learn more about the functioning of DTM measures and road user responses to them.

3.4 Statistical and modelling tools

In the previous section three alternative designs were studied that may increase confidence of causality between DTM measures and measured changes of indicators. It was concluded that each of these methods has drawbacks or practical limitations. In this section two methods will be discussed which may also provide insights into the causality between DTM measures and measured changes of indicators, without changing the before and after study experimental design.

Statistical tools
Statistical (data-mining) methods such as principal component analysis or common factor analysis could be used to identify principal components or factors that best explain variations of indicators at each measurement point. DTM measures will form part of these principal components. A difficult step would be the interpretation of results: which principal components correspond to which DTM measures?

Tsekeris and Stathopolous (2006) show that traffic flow variations in a large-scale arterial network can be described by only a small set of principal components. These principal components indicate the common sources of the variability of traffic flow. "In turn, this facilitates the deeper understanding and a more plausible interpretation of the factors contributing to the long-term evolution of the main characteristics of urban network traffic" (Tsekeris and Stathopolous, 2006). Principal Component Analysis could thus be a useful tool to determine impacts of DTM measures on traffic flow.

Hybrid simulation-measurements approach
Secondly, traffic models could be used to recreate the reference case, what would have happened without DTM measures, based on data from after study. This hybrid simulation/measurements approach for ex-post evaluation is thus a form of "reverse engineering".

This hybrid approach has a major advantage, but also puts strains on data collection and requirements for the traffic simulation model used. A major advantage of the approach is that one does not need to know exact impacts of external influences at the time of after study, since they will be present in both after study and the reconstructed (simulated) reference case. However, to reconstruct the reference case correctly it is a necessity to have sufficient information on the functioning and (local) impacts of each DTM measure and road user responses to them. Moreover, sufficient information needs to be gathered to be able to reconstruct traffic situations at the time of after study. Finally, the traffic simulation model used should be able to deal with these inputs.

The evaluation of A10-west road works (Taale, Bootsma and Schuurman, 2002) is an example of an evaluation which links ex-ante simulation study to ex-post measurements. In this evaluation results of an ex-ante simulation study were compared with actual measurements. No examples of the application of a true hybrid ex-post evaluation, where a traffic simulation study is based on ex-post measurements, were found however.

The hybrid approach is a promising approach when a traditional approach for ex-post evaluation, based on measurements, fails to provide sufficient proof of the causal relation between DTM measures and measured changes in indicators. A pitfall of the approach would be to base inputs of the simulation model used on unproven assumptions on the functioning of DTM measures, thus creating "a self-fulfilling prophecy". Instead, the simulation model should be fed with actual traffic data and (proven) microscopic impacts of DTM measures.

Conclusions about statistical and modelling tools
Both described methods provide insights into the causality between DTM measures and measured changes of indicators. The described approaches are in principle applicable to any DTM project. Statistical tools do not improve the potential to learn from evaluation. Simulation models may help to learn about impacts of DTM measures. However, simulation models are based on assumptions of the functioning of DTM measures and road user behaviour. If this is incorrectly modelled, simulation results will not provide (correct) insights into what works, when it works, what it causes and why.

3.5 Theory-based evaluation

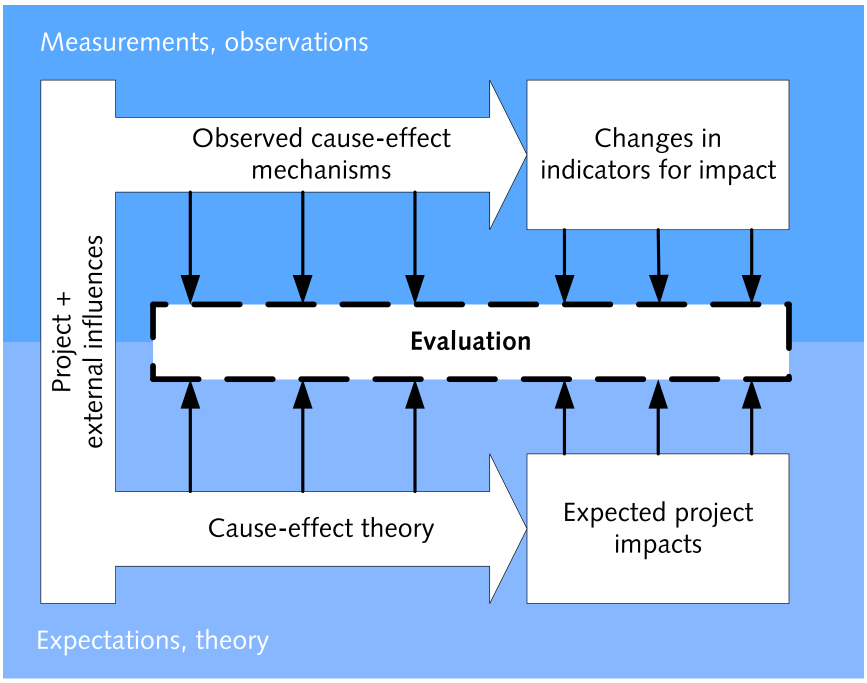
In this section an approach to evaluation is introduced as an alternative to the approach used in the methodologies studied in this research. "Current approach" is used to refer to the before and after study approach used in these evaluation methodologies. The alternative approach is called theory-based evaluation. "Theory" in this context refers to the assumptions underlying DTM design that explains how DTM measures are expected to lead to achieve their final (often large scale, long-term) impacts.

The concept of theory-based evaluation and differences compared to the current approach are explained below. Next, a motivation for using the theory-based evaluation approach in DTM projects is given. Finally, logic models, useful tools to make project theory explicit, are discussed in the final part of this section.

Concept of theory based-evaluation
The concept of theory-based evaluation is that the design of every project is based on theory and that this project theory should be the basis of evaluation (Chen and Rossi, 1987; Rossi, Lipsey and Freeman, 2004; Weiss, 1995; W.K. Kellogg Foundation, 1998). In this context "project theory" refers to an explanation of how project designers expect projects to achieve their final (often large scale, long-term) impacts. The word "expected" is used on purpose, since every theory is based on explicit or implicit assumptions. Theory-based evaluation argues that these underlying assumptions are what evaluation should focus on, instead of only measuring indicators for final impacts of projects.

Figure 3.5 clarifies the concept of theory-based evaluation. In evaluation planning one formulates expected impacts of projects and cause-effect theory of how projects are expected to cause these impacts. One also determines which external influences can affect project impacts and cause-effect mechanisms. Ex-post evaluation then determines whether cause-effect theory matches observed cause-effect mechanisms as well as whether expected project impacts match measured changes in indicators for these impacts.

Figure 3.5: Theory-based evaluation approach



Compare Figure 2.3 to Figure 3.5 to identify differences with the current evaluation approach. In the current approach one formulates an expectation about impacts of projects. For example in the CONVERGE and Finnish methodologies formulating expected impacts is explicitly included as an evaluation step. Ex-post evaluation then determines whether expected project impacts match measured changes in indicators for these impacts. External influences on these indicators are taken into account. The current evaluation approach does not focus on cause-effect theory of projects, which explains *how* expected final impacts are achieved.

Origins of theory-based evaluation

Theory-based evaluation has its origins in social sciences, where it is mainly applied to evaluate complex social programs, such as programs that aim to reduce child abuse or aim to increase social engagement of minority groups. Proponents of theory based-evaluation argue that it is better able to test causal relationships between activities taken as part of social programs and expected final impacts, compared to an approach of only determining changes in indicators for these impacts (Chen and Rossi, 1987; Pawson and Tilley, not dated; Rogers, 2000; Rossi, Lipsey and Freeman, 2004; Weiss, 1998; W.K. Kellogg Foundation, 1998).

Social programs share several similarities with DTM projects. These similarities may be an indication that the theory-based approach may be applicable to evaluate DTM projects as well. Furthermore, this may indicate that strengths of the theory-based approach also apply to the evaluation of DTM projects.

Both DTM projects and social programs produce impacts that require a sharp definition in order to be measurable. Social programs may for example aim to “improve social integration of minority groups” while DTM projects may aim to “improve traffic flow”. Without a definition of “improved social integration” or “improved traffic flow” and the selection of relevant indicators these impacts are hard to measure.

Both DTM projects and social programs try to achieve these impacts in a complex world, affected by numerous and often unknown external influences. Indicators that one can use to measure impacts of DTM projects or social programs are affected by these external influences. Furthermore, expected impacts of DTM projects or social programs on indicators are often low compared to expected impacts of external influences. The “signal-to-noise-ratio” is thus low in both. Causally attributing measured changes in indicators to DTM projects or social programs should therefore be done with care.

Finally, both network-wide DTM projects and social programs need to make assumptions about the behaviour of individuals. Social programs need to make assumptions on the response of individuals to social programs. Designers of DTM measures need to make assumptions on road user behaviour.

Motivation for using theory-based evaluation in DTM projects
The DTM evaluation methodologies studied in this research focus on expected final impacts of DTM. A before and after study approach is used for impact evaluation. It was concluded that this approach is weak at proving causality between DTM measures and measured changes of indicators.

Evaluating project theory (assumptions that explain *how* these expected final impacts are achieved) can reduce this weakness. It can be argued that two DTM evaluation methodologies studied in this research (the CONVERGE methodology and Finnish ITS evaluation methodology) evaluate project theory to an extent, since they include technical and user-response evaluation. However, information of technical and user response evaluation is not explicitly used to support conclusions of impact evaluation. Arguments for using theory-based evaluation in DTM projects are explained in more detail below.

A first argument for using theory-based evaluation in DTM projects is that it avoids an important pitfall that threatens evaluation using an approach of only determining changes in indicators for expected final impacts of projects (Chen and Rossi, 1987; Pawson and Tilley, not dated; Rogers, 2000; Rossi, Lipsey and Freeman, 2004; Weiss, 1998; W.K. Kellogg Foundation, 1998). The likelihood that measured impacts are wrongfully attributed to DTM projects is reduced by basing evaluation on project theory. Tracking micro-stages of impacts as they evolve makes it more plausible that measured changes in indicators are due to the evaluated project and not to external influences. One can be more confident about the causal relationship between DTM measures and measured changes in indicators.

Theory-based evaluation can also increase our knowledge of the functioning of DTM measures and road user response to them. By evaluating project theory, the approach helps to determine which measures work and which do not, when they work, what they cause and why. By increasing our knowledge of the functioning of DTM measures, DTM design can be improved in the future.

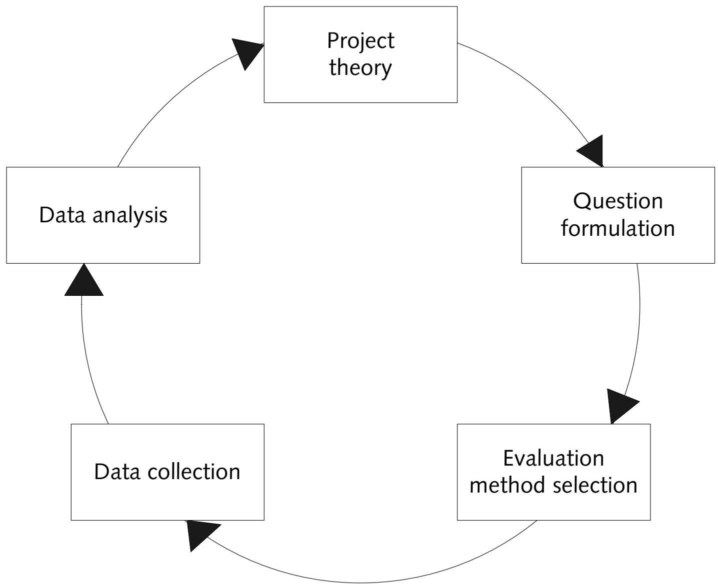
Finally, the theory-based approach links the various evaluation types already common in the evaluation of DTM projects, such as technical evaluation, user-response evaluation and impact evaluation. Project theory includes assumptions about the technical functioning of DTM measures, road user responses to these DTM measures and resulting impacts. The theory-based approach thus allows evaluators to create a clear overview of relevant evaluation types and their interdependencies.

Logic models as tools for theory-based evaluation

For theory-based evaluation to be performed, project theory needs to be made explicit. It is thus required to explicitly state what resources (such as human resources, financial resources, partnerships, materials and equipment) projects use, what their final impacts are expected to be and how each resource is expected to contribute to these final impacts. In many cases, resources are not expected to contribute to final, long-term impacts *directly* but *indirectly*. These resources generate outputs, which are expected to lead to micro-scale or short-term outcomes. These outcomes are then expected to lead to mesoscopic or intermediate outcomes and finally to large scale, long-term impacts. By making this chain of assumed interim outcomes explicit, project theory is uncovered.

The resulting project theory is then used to formulate evaluation questions. The selection of evaluation methods, data collection and data analysis is specifically aimed at these evaluation questions, as shown in Figure 3.6. After data is analysed, a comparison with project theory is made. Were assumptions made in project theory correct? And were expected impacts achieved? Why (not)? This cycle can be run through several times during the various phases of project design, implementation and evaluation.

Figure 3.6: Steps in theory-based evaluation approach
Source: (Pawson and Tilley, not dated), adapted.



Useful tools to make project theory explicit are logic models. A logic model is a graphical depiction of "a chain of causal assumptions linking project resources, activities, intermediate outcomes and ultimate goals" (Wholey, 1987). In practice many different terms besides "logic model" are used for this concept, such as "impact pathways", "theory of change", "program model", "outcome line", "causal model" "cause map" and "action theory" (Rogers, 2000; Rossi, Lipsey and Freeman, 2004).

Logic models come in various shapes, sizes and arrangements. The basic elements of logic models are shown in Figure 3.7.

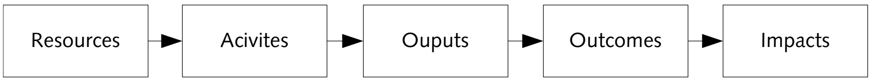


Figure 3.7: Elements of logic models
Source: (W.K. Kellogg Foundation, 2004)

Logic models consist of five elements, each of which may have sub-elements. In (W.K. Kellogg Foundation, 2004) these elements are defined:

- *Resources* are what projects use as inputs to be able to achieve their desired final impacts. These resources may include both tangible and intangible inputs such as human resources, financial resources, partnerships, materials and equipment.
- *Activities* are what is done with resources within the scope of projects.
- *Outputs* are direct effects of these undertaken activities.
- *Outcomes* are expected effects of these direct outputs, often involving responses of project target groups.
- These outcomes are a necessity to achieve desired final project *Impacts*.

Each box-arrow combination in logic models is an "if-then" relationship, which is based on an assumption. For example:

- **If** this resource is available, **then** this activity can be undertaken;
- **If** this activity is undertaken, **then** these outputs will be generated;
- **If** these outputs are generated, **then** these outcomes will be caused;
- **If** these outcomes are caused, **then** this impact will be achieved.

Thus each component of logic models "is the outcome of the successful attainment of the previous component and, in turn, is a boundary condition to attainment of the next component" (Verma and Burnett, 1999).

A simple example is given below to clarify the concept of logic models as shown in Figure 3.7. In Section 4.2 a proposal for the elements of logic models for DTM projects is made. An example of such a logic model for a DTM project is given in Sections 4.4 and 0.

Besides project activities, *external influences* may affect assumptions throughout logic models. Figure 3.7 does not explicitly include these as separate elements of logic models. Due to their importance, these are included as separate elements in the logic model of the example below. They are also included in the proposal for logic models for DTM projects in Section 4.2. The further one moves away from outputs in logic models, the weaker influences of outputs will generally become and the higher the likelihood of external factors having an influence will be (Rogers, 2000).

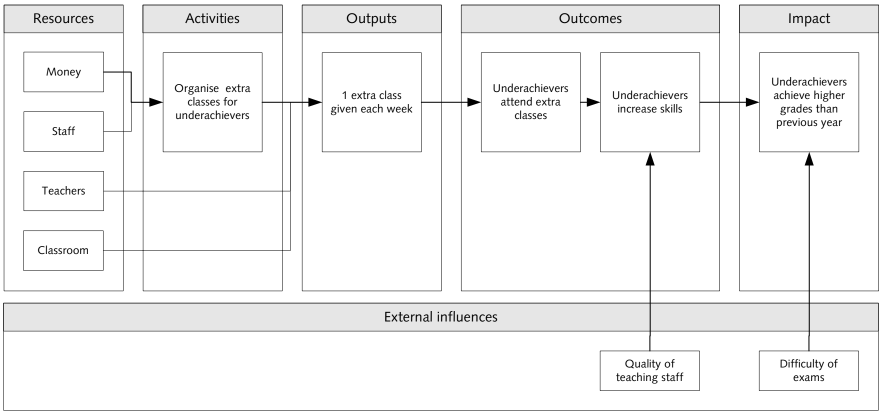
Example: Simple logic model for social program
This simple example is meant illustrate the concept of logic models as shown in Figure 3.7

Example: extra teaching for underachieving children
A school wants to improve grades of children they consider "underachievers". The school assumes extra teaching will improve their grades. Therefore the school will organise an extra lecture once a week for these children.

Example: Simple logic model for social program, continued

The school invests resources, money and support staff, to organise extra classes, the activity of this social program. The output is one extra class given each week. Teachers and a classroom are extra resources needed for these classes to be given. Expected outcomes of the organised extra classes are that underachievers attend these classes and therefore increase their skills. This increase in skills leads to the expected final impact of the program: underachievers achieve higher grades than in the previous year.

This results in the following logic model for this simple social program:



Two examples of external influences are included. Whether skills of underachievers increase also depends on changes in the quality of teaching staff. Whether higher grades are achieved also depends on the difficulties of exams.

Examples of assumptions visualised with the logic model are:
- Sufficient money, support staff, teachers and classrooms are available;
- Underachievers attend the organised extra classes;
- Attending extra classes increases skills of underachievers;

Theory-based evaluation should evaluate these assumptions.

The description above may give the impression that logic models are based on a single causal path, read from left to right. A logic model may however have multiple causal paths (Rogers, 2000). Some of these alternative causal paths may be complementary, some may be mutually exclusive and sometimes several causal paths may be a necessity to achieve a single outcome.

Conclusions about theory-based evaluation

Theory-based evaluation increases confidence about causality between DTM measures and measured changes of indicators. In principle the approach can be applied to any DTM project. Furthermore, theory-based evaluation increases the potential to learn from evaluation: it determines what works, when it works, what it causes and why.

3.6 Conclusions

In this chapter potential improvements for weaknesses of the current evaluation methodology were analysed. In this chapter the following questions was answered first:

Research question 1 of Chapter 3

For which of the identified weaknesses do alternative evaluation methodologies not offer an improvement?

It was concluded that the three studied alternative methodologies offer improvements to some extent, however:

- The potential to learn from evaluation is not fully exploited, since interdependencies between the various evaluation types and various impacts are not made explicit.
- The weakness at proving causality between DTM measures and measured changes of indicators remains.

Based on these conclusions, three criteria were formulated to determine which potential improvement(s) for these remaining weaknesses are "most promising":

- Does it increase confidence about causality between DTM measures and measured changes of indicators?
- Does it increase the potential to learn from evaluation: does it help to determine what works, when it works, what it causes and why?
- Is it in principle applicable to any DTM project?

These criteria help to answer the following research question:

Research question 2 of Chapter 3

What potential improvement offered in evaluation-related literature for these remaining weaknesses is most promising?

In Table 3.2 the scoring of the studied potential improvements on these three criteria is summarised. Both a hybrid simulation/measurements approach and theory based evaluation score a "yes" on each criterion. Nevertheless, the simulation approach must be based on correct assumptions about the functioning of DTM measures and road user responses to be useful. Theory-based evaluation explicitly evaluates these assumptions. In the next chapter, a new evaluation methodology based on the theory-based approach is therefore proposed.

Table 3.2 Scoring of the studied potential improvements on three criteria

Potential improvement	Improves causality	Improves learning	Practically Applicable
Before and after study with control area	No	No	Yes
Alternating evaluation	Yes	No	No
Trend or timeline analysis	Yes	No	No
Data mining techniques	Yes	No	Yes
Hybrid simulation/measurements approach	Yes	Yes[28]	Yes
Theory-based evaluation	Yes	Yes	Yes

[28] See remark in description above table

4. Proposed evaluation methodology

4.1 Introduction

In the previous chapter, theory-based evaluation was selected as a promising alternative to the approach used in the current ex-post evaluation methodology for DTM of Rijkswaterstaat. In this chapter the theory-based approach is used in a proposal for a new evaluation methodology. The central question of this chapter is:

Research question of Chapter 4

How can this most promising improvement be incorporated into a proposal for a new evaluation methodology?

Focus of the methodology proposed in this report is on project theory formulation, evaluation question formulation and evaluation method selection. Using the theory-based approach, assumptions in DTM design, external influences and their interdependencies are mapped. Based on this project theory, evaluation questions are formulated and evaluation methods are chosen for each evaluation question. Focus is on these steps, since they are not (extensively) discussed in existing evaluation methodologies studied in Chapter 3.

Focus of this report is not on data collection, data analysis and feedback of evaluation results. Since focus is not on feedback of evaluation results, socio-economic evaluation is also outside the scope of this report. Figure 4.1 shows the focus of this report, in Section 4.6 a reflection on this scope is presented.

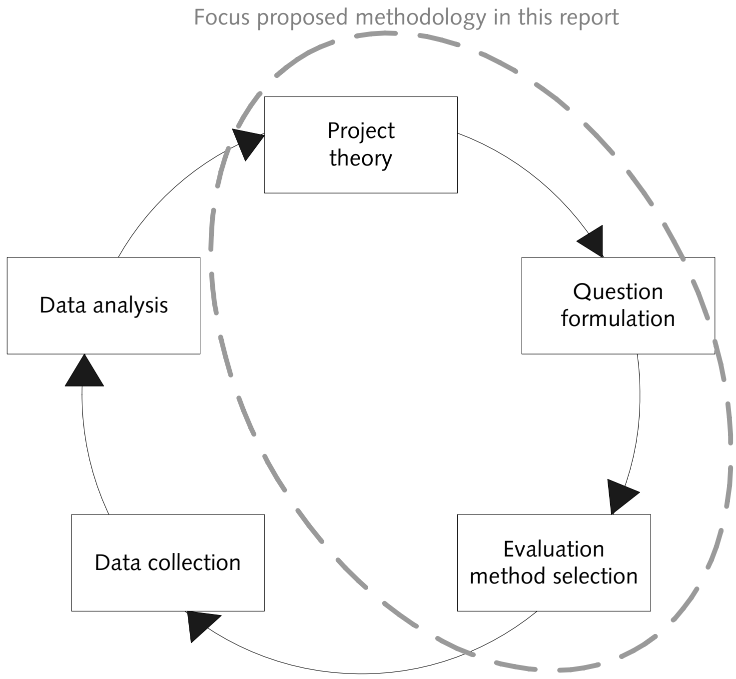


Figure 4.1: Focus of methodology proposed in this report
Source: Pawson and Tilley (not dated), adapted. Further work is required to include steps not part of the focus of this report.

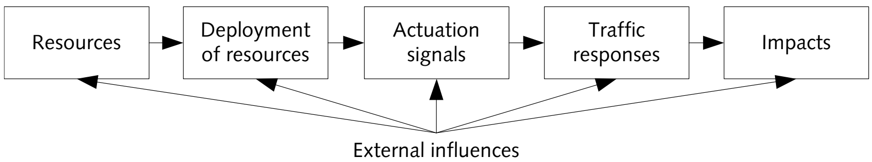
Logic models, introduced in the previous chapter, are useful tools to make project theory explicit. In Section 4.2 a proposal for logic models for DTM projects is presented. After this in Section 4.4 a step-by-step approach to create logic models for DTM projects is proposed. This approach is continued in Section 4.5 by formulating evaluation questions and selecting evaluation methods to answer these questions. In order to clarify the proposed approach it is applied to an example in both sections. At the end of this chapter, a reflection is given on the proposed evaluation methodology and its scope.

4.2 Logic models for Dynamic Traffic Management projects

In the previous chapter logic models were introduced as tools to make project theory explicit. In this section a proposal is made for logic models specific for DTM projects.

Proposal for logic models for DTM projects
A proposal for the elements of logic models specific for DTM projects is shown in Figure 4.2. See Figure 3.7 for the elements of logic models as they are used in social programs. The proposed logic models consist of six elements, which may each consist of a set of sub-elements. A description of each of the six elements is given below, with examples for clarification.

Figure 4.2: Proposed elements of logic models for DTM projects



Resources are what projects use as inputs to be able to achieve desired final impacts. These resources may include human resources (staff in traffic management centres, motorway inspection teams), hardware (ramp metering installations, variable message signs, monitoring equipment) and software (control algorithms).

Deployment of resources is what is being done with resources to influence traffic conditions. This could be the calculation of on-ramp flow rates at ramp meters or the selection of rerouting scenarios. The more precise one can describe when, under what conditions and how one expects resources to be deployed, the more useful logic models will be for evaluation purposes.

Actuation signals are *direct* outputs of the deployment of resources, which are intended to influence traffic conditions. Ramp metering installations for example display "green" or "red" signals or route advice may be displayed on Dynamic Route Information Panels.

Traffic responses are responses of traffic to actuation signals. This involves assumptions about road user response. The "green" or "red" signals of a ramp meter may be expected to lead to an increased traffic flow on the motorway and smoother merging of traffic. The displaying of route advices on route information panels may be expected to cause road users to switch to alternative routes, which may in turn reduce traffic volumes on congested roads. As with the description of the deployment of resources, the more precise one can state when one expects which effect, the more effective evaluation can be.

Impacts are what DTM projects aim to achieve and for which the expected *traffic response* is a necessity. Expected impacts are based on project objectives, such as increased traffic safety or accessibility. If expectations about sizes of impacts exist, these can be included into logic models.

Final impacts may not only consist of what is desired to change, but may also include what is *not* desired to change. Examples of this may be found in boundary conditions of DTM projects such as "no deterioration of traffic safety" or "no increase in emissions". Logic models can include separate paths that explain how resources will be used to prevent that these boundary conditions are violated.

External influences may affect the functioning of resources, actuation signals, traffic responses to actuation signals and impacts DTM projects aim to create. External influences are explicitly added as separate elements to the proposed logic model for DTM projects due to their importance, as discussed in Chapter 2. The further one moves away from actuation signals in logic models, the weaker actuation signal influence will generally become and the higher the likelihood of external factors having an influence will be (Rogers, 2000). Thus the more one moves to impacts at the right hand side of logic models, the more external influences need to be taken into account.

4.3 Introduction to proposed evaluation methodology

In Section 4.4 an approach to develop logic models for DTM projects is presented. This logic model then forms the basis for the design of theory-based evaluation approaches in Section 4.5.

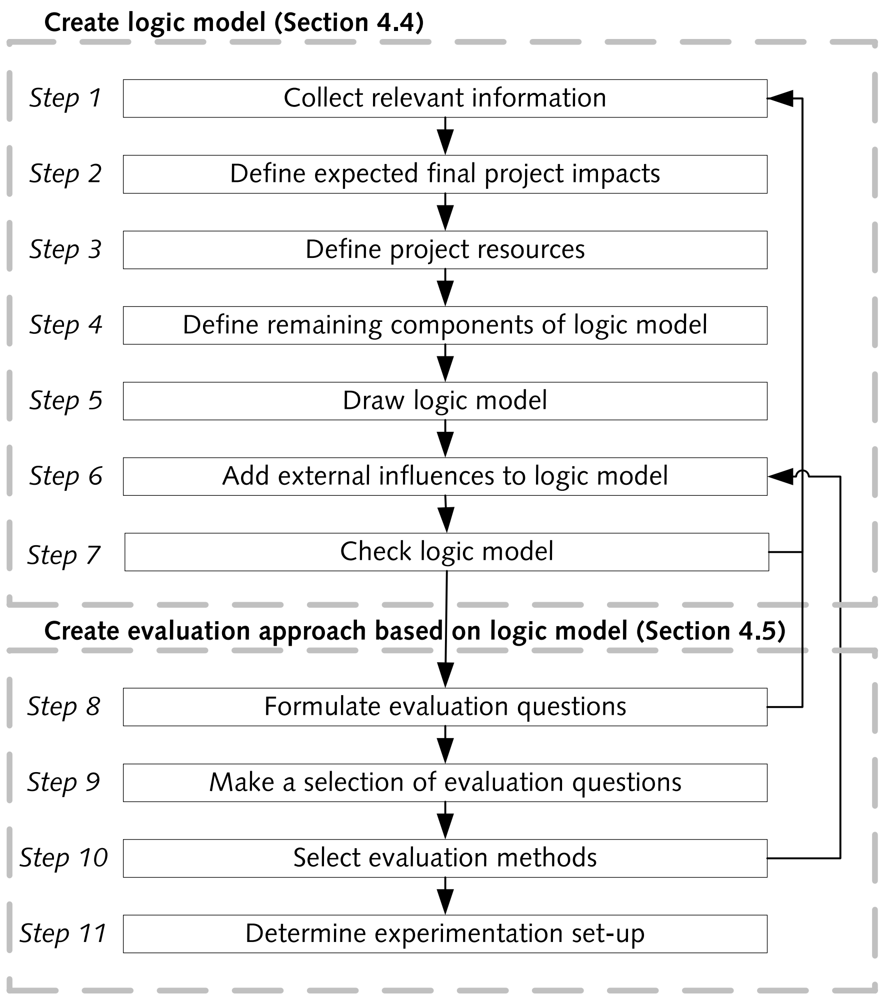
The proposed methodology is focused on creating an evaluation approach when DTM design is finished. See the reflection in Section 4.6 for remarks about relations between logic model development, DTM design and decision-making.

An approach to develop logic models is proposed which minimises the likelihood that (unintended) impacts or traffic responses caused by DTM projects are missed in evaluation. The approach starts with defining left and right boundaries of logic models. Defining expected project *impacts* (right boundary) and *resources* (left boundary) does this. Resources can then be connected to impacts by reasoning in two directions. This identifies the *deployment of resources*, *actuation signals* and *traffic responses*. After this *External influences* are added to logic models. Finally, a check is made whether the developed logic model is plausible, acceptable and complete. Project theory is now made explicit with the developed logic model.

The resulting logic model is used to generate an evaluation approach. Hereby evaluation questions are derived directly from it. After this a selection of evaluation questions is made, thus determining evaluation focus. Trade-offs between evaluation objectives are made explicit. The developed logic model then provides an instrument to select appropriate evaluation methods and experimentation set-up to answer each evaluation question. External influences are explicitly taken into account.

An overview of the resulting steps of the proposed evaluation methodology is presented in Figure 4.3.

Figure 4.3: Overview of steps of proposed evaluation methodology



Three feedback loops are explicitly included in the methodology, although more may exist in practice:

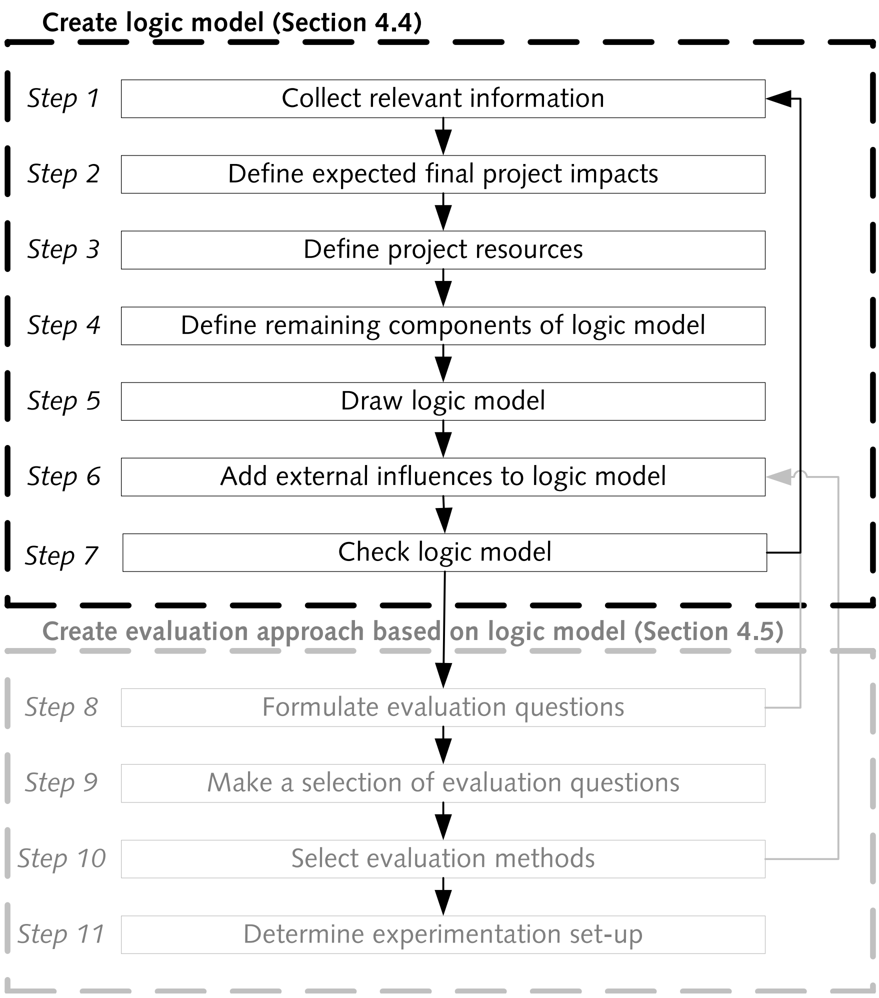
- After checking logic models one may find that information is missing and return to step 1.
- When formulating evaluation questions one may discover that evaluation questions cannot be answered unambiguously and return to step 1. A sharper definition of developed logic models is then required.
- Selected evaluation approaches may help to uncover new external influences in step 6.

In the next sections steps of the methodology are explained in detail. By way of illustration, the proposed methodology is applied to a simple DTM example. This example returns at each step in textboxes.

4.4 Creating logic models

This section describes an approach to create logic models for DTM projects. In the next section the approach is continued by creating an evaluation approach based these logic models. See Figure 4.4 for an overview of the steps of the methodology that are discussed in this section.

Figure 4.4: Steps of the proposed methodology that are discussed in Section 4.4
Steps discussed in this section in black steps discussed in the next section in grey



Result of the approach of this section should be a logic model that is plausible, accepted and complete. Elements of the developed logic model should be defined in such a way that they are not open to multiple interpretations. The logic model can then be translated into evaluation questions that can be answered unambiguously in the next section.

Step 1: collect relevant information
Before a logic model can be created, all documents already produced about the researched DTM project and relevant for its logic model should be collected. The following documents may be relevant to create logic models:

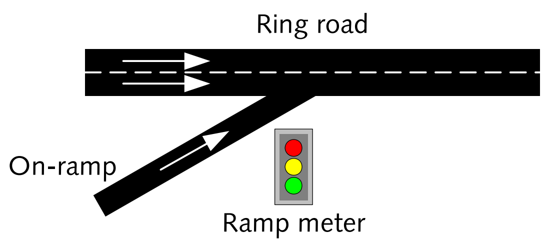
- Project plans;
- Network visions[29] and;
- Other deliverables of the "Sustainable traffic management[30]" process;
- (Functional) designs of DTM measures;

Project designers, decision-makers and key stakeholders can be interviewed in case information is missing.

Example: Ramp metering, step 1

Example: ramp metering
The DTM example that is used to illustrate the approach is a fictitious case concerning a single ramp metering installation on an on-ramp to a ring road. A schematic illustration is shown below. Design of the installation is completed and realisation of the project is planned. The evaluator is requested to evaluate the functioning and impacts of the ramp metering installation (a socio-economic evaluation is outside evaluation scope). The evaluator has gathered design documents of the installation and policy documents in which the decision to install the ramp metering is explained.

Note that the example used is intended to illustrate the proposed evaluation methodology. The example is deliberately simplified and is not meant to be (completely) realistic. Furthermore the example is not an application of network-wide DTM. In the next Chapter a case-study of a network-wide DTM project is presented.



[29] "Network visions" are documents developed by various road authorities in which a combined policy for the main and secondary road network is formulated
[30] In Dutch: "GebiedsGericht Benutten (GGB)"

Step 2: define expected final project impacts
Expected final project impacts can generally be found in information gathered in step 1, for example in project objectives. To determine expected final project impacts it may be necessary to first define problems the researched DTM project is designed to solve (or reduce). Expected final project impacts used in logic models should be discussed with decision-makers and key stakeholders.

The sharper impacts can be defined, the more useful the logic model will be for evaluation. An example of an impact that requires a sharper definition is "increased accessibility of city x". This impact can be interpreted in various ways. Is this impact, for example, the perception of accessibility of city x by road users? Or is this a reduction of average travel times to and from city x via road y? Or is this an increase of the total number of vehicle-kilometres on all access roads to city x? And when is this impact expected?

Note that impacts defined in project objectives may not be "final" project impacts, since impacts of DTM evolve over various timescales. See the short discussion in Chapter 2 for more details. Section 4.6 will return briefly on this issue.

Example: Ramp metering, step 2

The objective of the ramp metering project is to "keep the ring road flowing". The evaluator feels that this impact should be defined more precisely in order to be able to evaluate it. Designers expect the ramp metering to have an impact when total flow of vehicles is near to ring road capacity, but does not yet exceed it. This is generally at the beginning and ends of rush hours. By preventing ring road capacity being exceeded, traffic jams can be prevented or at least shortened. After discussion with decision-makers and project designers expected final project impact is therefore defined as *"increased average flow*[31] *at the ring road during rush hours"*.

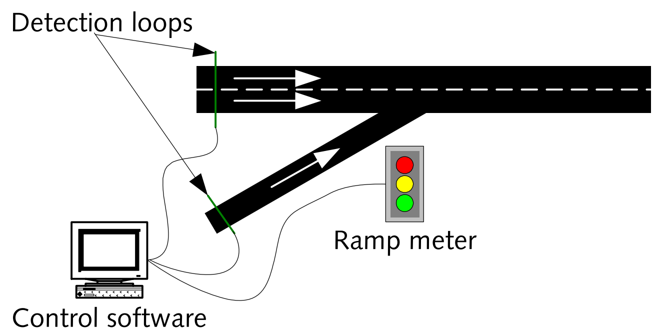
A key boundary condition for the ramp metering installation was defined with stakeholders: queues of vehicles waiting on the on-ramp should not spill back to secondary roads. A detection loop is installed at the on-ramp. Ramp metering stops once queues pass this detection loop.

Step 3: define project resources
Resources are required inputs to be able to achieve final project impacts. This may include human resources and hard- and software. These resources can be identified in project documents and in interviews. The resulting list of resources should then be discussed with decision-makers and project designers.

[31] It can be argued that "increased average flow" is still not an indicator fully explaining impacts of ramp metering, but for simplicity it is used in this example.

Example: Ramp metering, step 3

Since socio-economic evaluation is not in evaluation scope, the evaluator chooses not to incorporate financial and human resources into his "resources" list. He includes technical resources needed to achieve the expected final project impact. The ramp metering installation consists of two detection loops, a ramp meter, and control software, see the figure below.



Step 4: define remaining components of logic model
Once resources and expected final project impacts have been defined, resources need to be connected to impacts. This is generally a difficult step to take at once, since these connections are based on assumptions that may not be explicitly mentioned in project documents.

For this reason, an extra step is introduced to reduce these difficulties. In this step, the deployment of resources within project scope, actuation signals this generates and expected traffic responses mentioned in gathered information sources are collected. A useful way to structure this information is in a table (McLaughlin and Jordan, 1999), with resources, deployment of resources, actuation signals, traffic responses, and impacts as columns, see Table 4.1. The more explicit one can state under which conditions one will deploy which resources and where and which traffic responses are expected, the better DTM projects can be evaluated. To ensure no information is missed and all views and expectations are incorporated in the table, contents and completeness of the table should be verified with decision-makers and project designers.

Table 4.1 Layout of table to be used to create logic models

Resources	**Deployment of resources**	**Actuation signals**	**Traffic responses**	**Impacts**
What is needed for the project, what do we "put in"?	What will we do with our resources?	What are *direct* outputs of the deployment of resources?	What effects do we expect from our actuation signals, in many cases through road users?	What do we think we will achieve with the project, what do we "get out"?

Example: Ramp metering, step 4

The evaluator has collected information on the expected functioning and effects of the ramp metering installation. If total flow on the on-ramp plus the ring road exceeds a threshold value of 4000 vehicles/hour, the ramp metering installation will start operating. It will calculate appropriate on-ramp flow rates and based on that calculate signal intervals. “Green” is displayed every interval, allowing one car to enter the ring road. Ramp metering will stop operating once queues on the on-ramp pass the detection loop or if the ring road is congested despite ramp metering.

There is some debate between project designers about how actuation signals of the installation (the displaying of a “green” or “red” signal) will generate the expected final impact. Three versions are mentioned:

- The restriction of traffic flow on the on-ramp reduces total traffic flow from the ring road plus the onramp. This improves traffic flow downstream of the on-ramp, leading to an increased average traffic flow downstream of the on-ramp;
- Road users divert to other on-ramps, reducing the number of vehicles using the on-ramp. The lower number of vehicles actually using the on-ramp will increase average traffic flow downstream of the on-ramp;
- The ramp metering “smoothens” inflow of vehicles from the on-ramp by preventing “platoons” of vehicles to enter the ring road. Due to reductions of “disturbances” from the on-ramp, average traffic flow will increase downstream of the on-ramp.

The evaluator decides to take each of the three alternative theories into account and based on his information, the evaluator constructs the following table:

Resources	**Deployment of resources**	**Actuation signals**	**Traffic responses**	**Impacts**
• Control software • Detection loop ring road • Detection loop on-ramp • Ramp meter	• Calculate signal intervals • Measure flow ring road • Detect queue on-ramp	• If control conditions are met, ramp metering installation displays “green” each interval and allows one car to pass and “red” otherwise	• Reduced inflow onto ring road • Smoothened merging of vehicles • Road users divert to other on-ramps	• Increased average flow downstream on-ramp during rush hours • Queues on on-ramp do not spill back to secondary roads

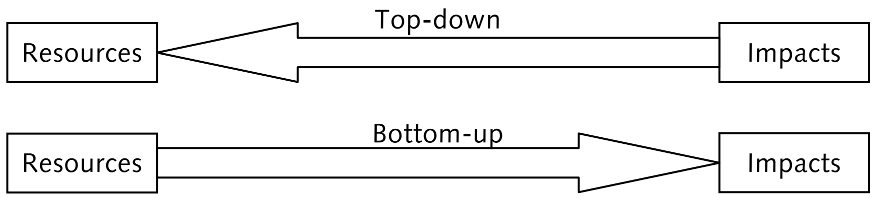
The evaluator, decision-makers and project designers agree that this table should be used to create the logic model.

Step 5: draw logic model

In this step of creating logic models, information of Table 4.1 is used to actually draw them. The resulting logic model will connect resources to deployment of resources, deployment of resources to actuation signals, actuation signals to traffic responses and traffic responses to impacts. All elements in the table are to be given a place in the logic model, leading to a plausible explanation of how resources are used to create expected final impacts.

The drawing of logic models can be done in two ways: top-down and bottom-up, see Figure 4.5. In the top-down approach one starts with final project impacts and connects this to traffic responses. After traffic responses are connected, actuation signals are connected to traffic responses. The deployment of resources that creates these actuation signals are then added, followed by resources required. The bottom-up approach works vice-versa, starting with resources and ending with final impacts. Both approaches can be used to complement each other.

Figure 4.5: Two approaches to creating logic models
The top-down approach starts with final project impacts and works its way down to explain how project resources will contribute to these impacts. The bottom-up approach works vice-versa, starting with resources



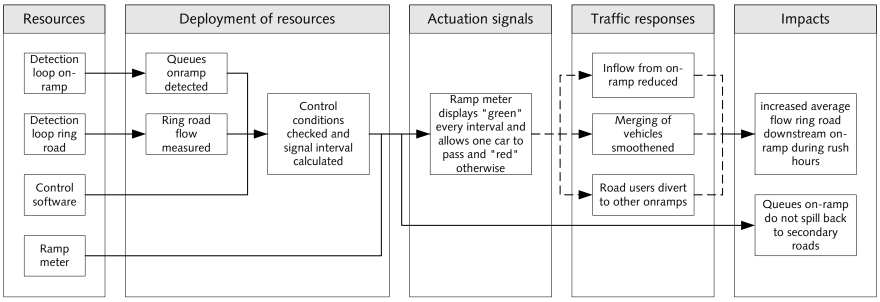
The actual drawing of logic models is usually the most difficult step. While drawing logic models one will often discover that information is missing since assumptions made in DTM design were not explicitly documented. This missing information should be discussed with decision-makers, project designers and key stakeholders. For the time being "black boxes" can be used in logic models. Arrows leading into black boxes and out of black boxes are known, but assumptions about "internal" functioning of black boxes remains unknown. Suppose for example accidents need to be detected, but how this will be done is not yet known. In this case, a black box can be included in the logic model that states "Detect accident". In later versions of the logic model one can make explicit how these accidents are expected to be detected.

A possible method that may be helpful to uncover assumptions included in DTM design is to create (hypothetical) scenarios for the deployment of DTM measures. A certain traffic condition is assumed, which DTM measures are expected to improve. What resources are required to improve this traffic condition, how these resources will be deployed, what actuation signals will be produced and how traffic is expected to respond is analysed step by step. Assumptions made in each step are then uncovered.

Expectations on how project impacts will be achieved may vary. This can be incorporated in logic models. Logic models then have alternative paths leading to final impacts. Evaluation can take several paths into account and evaluate which of the alternative theories is "most likely". An example of a logic model with alternative causal paths is included in the example in the textbox below.

Example: Ramp metering, step 5

By connecting elements of the table created in step 4, the evaluator has drawn a logic model for the ramp metering project. Under "traffic responses" the evaluator has included three alternatives, only one of which has to be correct to achieve expected final impacts. The three alternatives can also be complementary, each contributing to achieve expected final impacts. To indicate this relationship the evaluator uses a dashed line. This leads to the following logic model (the logic model is shown on a larger scale in Appendix B):



By "reading" the logic model, a plausible explanation should now be given of how resources are used to achieve expected final impacts. In this example *assumptions* in the logic model (to be tested in evaluation) can be "read" as follows (bold letters indicate arrows in the model):

- **If** the detection loop on the on-ramp is functioning **then** queues on the on-ramp will be detected correctly.
- **If** the detection loop on the ring road is functioning **then** correct ring road flows are measured.
- **If 1)** correct ring road flow are measured **and if 2)** queues on the on-ramp are detected correctly **then** the control software will check control conditions (ring road flow exceeds 4000 vehicles per hour and queues on the on-ramp are not past the detection loop) and calculate signal intervals.
- **If** control conditions are checked and signal intervals are calculated **then 1)** the ramp meter will display "green" every interval and will allow one car to pass and "red" otherwise **and 2)** queues on the on-ramp will not spill back to secondary roads.
- **If** the ramp meter displays "green" every interval and allows one car to pass and "red" otherwise **then 1)** inflow from the on-ramp will be reduced **and/or 2)** road users will divert to other on-ramps **and/or 3)** the merging of vehicles onto the ring road will be "smoothened".
- **If 1)** inflow from the on-ramp is reduced **and/or if 2)** road users divert to other on-ramps **and/or if 3)** the merging of vehicles onto the ring road is "smoothened" **then** average traffic flow downstream of the on-ramp during rush hours will be increased.

Step 6: add external influences to logic model
Due to the importance of external influences in ex-post evaluation of DTM projects, as shown in Chapter 2, external influences are dealt with in a separate step in the proposed approach. In this step external influences are added to logic models. This is done by asking at each element of logic models, "what other events, projects or developments, not in control of the studied DTM project, may have a (positive or negative) influence on this part of the logic model?" The resulting external influences are then given their expected places in logic models. Some external influences only influence final impacts of DTM projects, while others may influence microscopic traffic responses.

Two types of external influences can be distinguished. The first type of external influences affect mechanisms DTM measures intend to set in motion directly. An example is the effect in-car navigation systems can have on whether route-advice shown on a roadside information panels is followed. Secondly, external influences can affect indicators one intends to use for evaluation of DTM projects, for example impacts economic growth can have on traffic flow on a certain road. The difference is explained further in Figure 4.6.

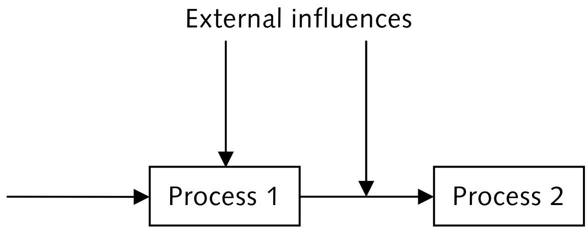


Figure 4.6: Two types of external influences
The first type of external influences affects process 1 itself, thus also affecting the output of process 1 and thus affecting process 2. The second type of external influence only affects the input for process 2, but does not affect process 1.

This distinction will be used in the selection of evaluation methods in Section 4.5. After evaluation methods and relevant indicators have been selected in Section 4.5, it is important to check lists of external influences again, since this may uncover more external influences on chosen indicators that were not considered before.

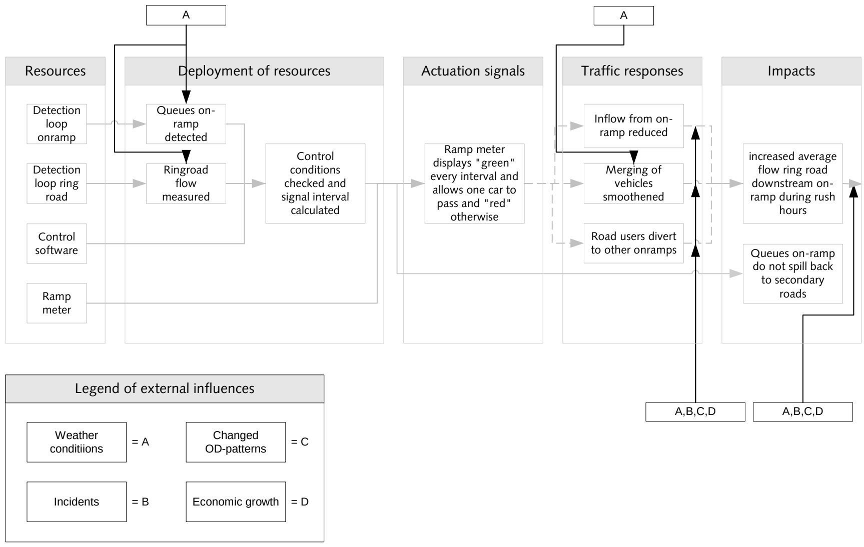
Example: Ramp metering, step 6

The evaluator, decision-makers and project designers take a closer look at each assumption that was stated using the logic model at step 5. They conclude that the main external influences to be taken into account in evaluation are external influences that may affect numbers of road users that use the on-ramp or that pass the ring road downstream of the on-ramp. These influences are:
- Incidents;
- Changes in origin-destination patterns;
- Weather conditions;
- Economic growth and resulting changes in infrastructure demand;

Example: Ramp metering, step 6, continued

Weather conditions are the only external influence that may directly affect whether the ramp metering installation functions correctly. The other external influences will likely only affect measurements used to evaluate the ramp metering project. Therefore the evaluator makes a distinction between these two types in the logic model. He places influences on processes directly “on the boxes” and above the logic model and indirect influences on subsequent processes “on the arrows” and below the logic model. To keep the logic model readable, a legend is used for external influences and the logic model itself is made grey.

This results in the following logic model (This logic model is shown on a larger scale in Appendix C):



The evaluator, decision-makers and project designers agree that this logic model will be used as a basis for development of the evaluation plan.

Step 7: Check logic model

Before creating an evaluation approach based on the developed logic model, it is recommendable to check it. This check prevents that impacts of DTM are missed in evaluation and to reach agreement on the developed logic model.

The developed logic model should meet three criteria:

- It should be *plausible*, what means that it presents a plausible explanation for how project resources are expected to lead to final impacts;
- It should be *complete*, what means that it includes all possible impacts of the studied DTM project and includes all possible causal paths which connect project resources to final impacts. The methodology focuses on expectations about the functioning and impacts of DTM measures (project theory), it is thus possible that impacts that are not considered in project theory remain unnoticed (Verma and Burnett, 1999);

- It should be *accepted*, what means that decision-makers, key stakeholders, project designers and evaluators agree on its contents. By achieving consensus about the developed logic model a common understanding is reached about project scope, its expected final impacts and expectations on how these final impacts are achieved.

Completeness of logic models will be the toughest to check. Plausibility and acceptability can be checked by discussing logic models with decision-makers, key stakeholders, project designers and evaluators. Even when project theory in logic models is accepted and plausible, it may not yet be complete. If project theory in logic models is incomplete, this may cause DTM impacts to be missed in evaluation.

Two methods can be used to check completeness. First of all, ex-ante evaluation can be used to check project theory before actual realisation, for example by studying evaluation reports of comparable projects and simulation study. Secondly a checklist of relevant DTM impacts, external influences and possible traffic responses can be used. This checklist can evolve as knowledge of DTM increases. See for example the checklist of relevant DTM impacts included in the Finnish ITS evaluation methodology (Finnish Ministry of Transport and Communications, 2002).

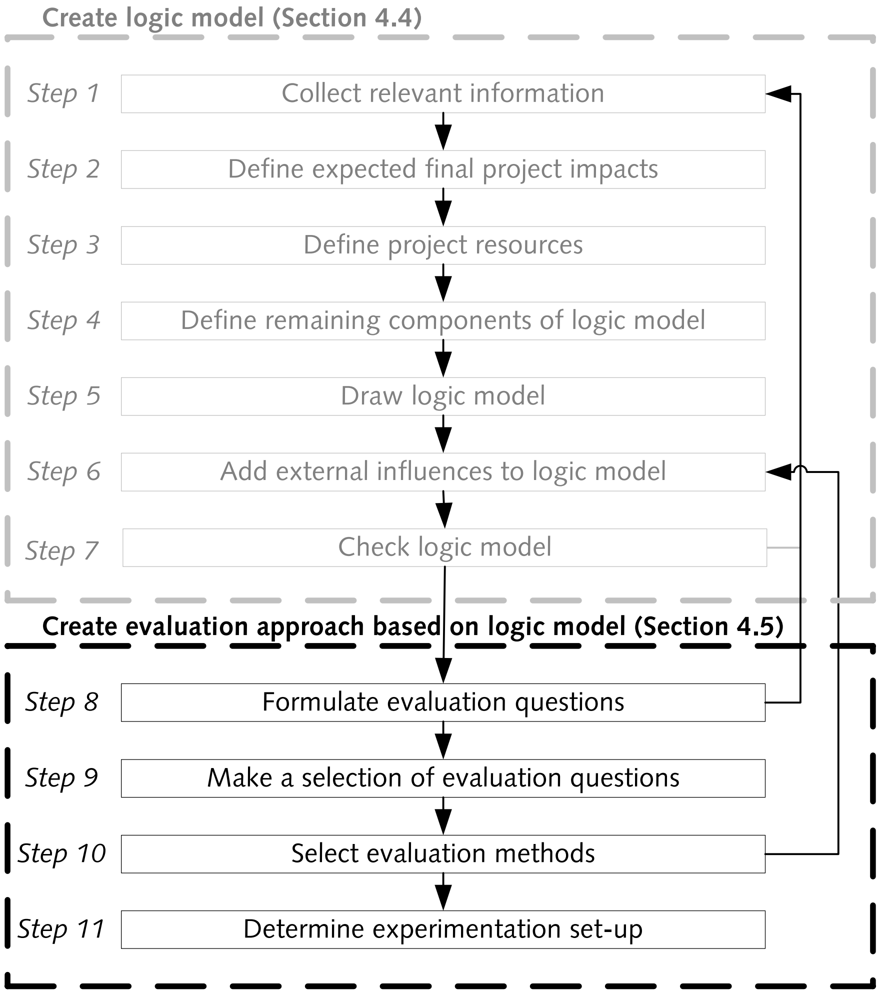
Several iterations may be necessary to discuss and check logic models, gather new information and update them. A feedback loop to step 1 of the methodology may thus be required here.

4.5 Creating evaluation approaches with logic models

The previous section resulted in the creation of logic models for DTM projects, graphical depictions of chains of causal assumptions linking project resources to expected final impacts. The logic model is thus a depiction of project theory, which forms the basis of theory-based evaluation.

In this section the step-by-step approach of the previous section is continued. Considerations for the development of evaluation approaches, based on developed logic models, are given in this section. Evaluation approaches developed in this section consist of a selection of evaluation questions, selection of evaluation methods and experimentation set-up. See Figure 4.7 for an overview of steps of the methodology that are discussed in this section.

Figure 4.7: **Steps of the proposed methodology that are discussed in** Section 4.5 Steps discussed in this section in black, steps discussed in the previous section in grey



Step 8: Formulate evaluation questions
In step eight evaluation questions are formulated for each "box-arrow combination" of logic models developed in the previous section. Each of these box-arrow combinations is an assumption in project theory. Depending on evaluation methods chosen, evaluation questions may also be phrased as hypotheses. Hypotheses are statements (thus not ending with a question mark) that are true or false and can be statistically tested to be rejected or confirmed.

This step may result in evaluation questions that are trivial or which are already common to ask in DTM evaluation. Advantages of theory-based evaluation approach are that it shows interdependencies between various evaluation questions and may lead to evaluation questions that might not have been considered using existing evaluation approaches.

If resulting evaluation questions cannot be answered unambiguously, elements of logic models need to be defined more sharply. Questions that can be asked to refine logic models are: Under what conditions will each assumption hold? What are expected sizes of effects (in a measurable indicator)? A feedback loop to step 1 of the methodology may thus be required here.

Example: Ramp metering, step 8

Using the logic model with external influences developed in step 6, the evaluator creates the following list of possible evaluation questions:

1. Does the ring road detection loop correctly measure traffic flow on the ring road upstream of the on-ramp? And during severe weather conditions?
2. Does the on-ramp detection loop correctly detect queues? And during severe weather conditions?
3. If traffic flow on the ring road exceeds 4000 vehicles/hour and queues on the on-ramp are not past the detection loop, does the control software calculate correct signal intervals?
4. Does the ramp meter display "green" every interval, does it only allow one car to pass "green and does it display "red" otherwise once it has started operating?
5. Is inflow from the on-ramp reduced due to the operation of the ramp meter?
6. Does the operation of the ramp meter lead to a smoother merging of vehicles onto the ring road? And during severe weather conditions?
7. Do road users divert to other on-ramps due to the operation of the ramp meter?
8. Does the reduced inflow from the on-ramp and/or the smoother merging of vehicles onto the ring road and/or the diversion of road users to other on-ramps lead to an increased average flow downstream of the on-ramp during rush-hours?

Step 9: Make a selection of evaluation questions
The previous step has likely resulted in a long list of evaluation questions; it may not be possible to answer all of these questions using the limited financial, technical and human resources available for evaluation. It may therefore be necessary to focus evaluation on questions that are deemed "most important".

A new question than logically arises: how does one determine which evaluation questions are "most important"? No strict guideline can be given to answer this question, since answers depend on preferences of project designers, decision-makers and key stakeholders. The following aspects should be taken into account however:

- The importance attached to evaluation question by project designers, decision-makers and key stakeholders. This preference may be highly subjective, but neglecting this preference may reduce the persuasiveness of evaluation or the likelihood that evaluation results are properly used.
- The uncertainty about corectness of assumptions. Some evaluation questions are based on assumptions which have already been confirmed in various projects under various conditions. In such circumstances investing resources into answering evaluation questions may not be sensible. However, if new technologies are applied, or existing technologies are applied in new conditions, or assumptions have simply never been tested, relating evaluation questions should be given more priority.

- (Expected) consequences for evaluated projects if assumptions prove incorrect or remain untested. It is required to identify "key" assumptions in project theory. Not testing "key" assumption may reduce the validity and/or persuasiveness of evaluation.
- The ease and likelihood at which evaluation questions can be answered (with sufficient confidence) during evaluation. Answering evaluation questions related to the left-hand side of logic models will generally be easier than answering evaluation questions related to the right-hand side. On the right-hand side, project impacts may only be small in comparison with impacts of external influences.
- Resources available for the evaluation. These resources include available monitoring and measuring equipment, simulation models and financial and human resources. Another important resource is time available for evaluation, since some evaluation questions cannot be answered simultaneously.

The process of selecting evaluation questions and determining evaluation focus incorporates an important trade-off that has to be made in evaluation. In Chapter 2 it was stated that the objective of ex-post evaluation has two aspects: the *learning* aspect and the *accountability* aspect. The learning aspect is mainly present in evaluation questions on the left-hand side of logic models, while the accountability aspect is mainly present in evaluation questions on the right hand side of logic models. A trade-off has to be made between these two aspects, due to the scarcity of resources available for evaluation. By selecting evaluation questions using this methodology this trade-off is made explicit.

A potentially efficient approach is to cluster evaluation questions into subgroups. The last evaluation questions in subgroups are the only questions which are included in evaluation scope. Only if these last questions are answered negatively, does one answer the other questions in subgroups. This will then determine which assumptions in logic models were incorrect.

Example: Ramp metering, step 9

After discussing the proposed set of evaluation questions with project designers and decision-makers, it is decided to include questions 2, 3, 5, 7 and 8 in the evaluation plan and not to include questions 1, 4 and 6 in the evaluation plan:

<u>Question 1:</u>
The ring road detection loop is already installed and has been verified before using traffic counts; the evaluation team decides it is not necessary to verify its functioning again.

<u>Question 4:</u>
The ramp meter has been tried and tested at multiple locations before and has proven to function according to specifications. It is assumed the ramp meter will work according to specifications in this project as well.

Example: Ramp metering, step 9, continued

Question 6:
The evaluation team decides that it is too difficult to define an objectively measurable indicator to answer this question.

Step 10: Select evaluation methods
Various methods can be used to answer evaluation questions selected in step 9. For example, the before and after study experimental design of the current evaluation methodology can be used effectively if its weaknesses are taken into account (see Chapter 2). Other possible methods are a hybrid simulation/measurements approach and data mining approaches described in Chapter 3, literature reviews, surveys amongst road users and observations by experts.

No strict guideline can be made for the selection of evaluation methods, but again a few considerations are given. The following aspects should be taken into account:
- The type of evaluation questions and (geographic and time) scales involved.
- The type, number and expected impacts of external influences. A hybrid simulation/measurements approach may be very suitable for situations where measurements would be disturbed by many unknown external influences. However, measurements or observations may be more suitable if external influences directly affect mechanisms DTM projects intend to set in motion. If external influences are numerous and their impacts are unknown and evaluation resources are available, a combination of methods is advisable.
- The persuasiveness of evaluation methods. Decision-makers may trust outcomes of an evaluation based on measurements more than one based on simulation outputs.
- Available evaluation resources.

After an evaluation approach has been chosen, new external influences may be identified. It is therefore prudent to review external influences identified during the creation of logic models again. Therefore a feedback loop to step 6 exists here.

Example: Ramp metering, step 10

Using considerations stated above the evaluator decides to use the following evaluation methods per evaluation question:

Question 2:
The functioning of the loop will be verified directly after it has been installed by comparing its output with traffic counts and visual observations.

Question 3:
The functioning of the control software will be verified using logs of detection loops, the ramp meter and the control software.

Example: Ramp metering, step 10, continued

Question 5:
The evaluator decides to determine inflow onto the ring road from the on-ramp for several rush hour periods using after study measurements. Using this data the team will reconstruct what inflow onto the ring road would have been without the ramp meter. This may indicate whether the ramp meter has sufficient space to store its queues during rush hours to actually have an effect on traffic flow downstream of the on-ramp.

Question 7:
By using a before and after study it will be determined whether the total number of vehicles that uses the on-ramp in rush hour periods reduces. If this is the case the evaluator will check whether several likely alternative on-ramps have received a substantial increase of traffic or perform a survey amongst likely users of the on-ramp.

Question 8:
The evaluator decides that before and after study is required to detect if any changes in average traffic flow downstream of the on-ramp during rush hours occur. Since many external influences affect this before and after study, the evaluator decides to combine it with a hybrid measurements/simulation approach. The outcomes of traffic flow calculations of evaluation question 5 together with traffic counts will be used as inputs for a simulation model. The model will then be calibrated based on actual measured flows. Next, the situation without the ramp meter will be modelled. Impacts of the ramp metering installation will be estimated by comparing results of before and after study measurements and simulation study.

Step 11: Determine experimentation set-up
Once evaluation methods are chosen, experimentation set-up can be determined. This final step of the methodology proposed in this report is not described in detail. Further work is required on this and following steps to refine the methodology. See Section 4.6 and recommendations made in Section 7.4 for details.

Experimentation set-up depends on many factors, including:
- Available monitoring equipment. Available monitoring equipment will determine which measurements are available to answer evaluation questions.
- Statistical considerations and required confidence levels. Some form of comparison between reference situations and situations after implementation of DTM measures will have to be made for most evaluation questions. In general the following holds:
  - The smaller impacts of DTM measures are, or;
  - The higher required confidence levels are, or;
  - The larger fluctuations in measurements are;

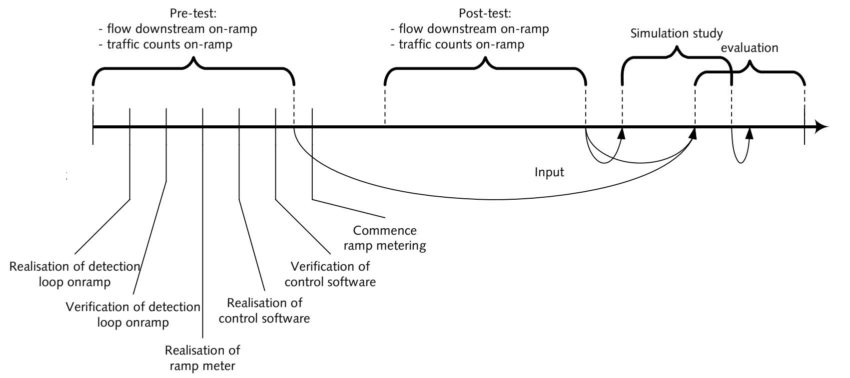
  The larger the required number of measurements or simulation runs will be (Zhang et al, 1998). It should be taken into account that the more measurements are required, the higher the likelihood becomes that (unknown) external influences are affecting measurements.

- Expected timing and sizes of impacts of external influences. If before and after study is used to answer evaluation questions, it should be performed in an environment where changes in impacts of external influences are negligible.
- Realisation of DTM measures. Realisation planning will affect the order in which evaluation questions can be answered. It also affects the ease at which evaluation questions can be answered. A long and phased realisation period allows the functioning of individual DTM measures and road user responses to these measured to be evaluated in detail. However, determining final impacts of all DTM measures based on measurements becomes harder. Due to the large time span between before and after study numbers and impacts of external influences will increase.

For each evaluation question, a separate experimentation set-up can be chosen. It should be kept in mind though that experimentation, evaluation and realisation planning should be considered as a whole. A combined planning for project realisation, experimentation and evaluation can be created once experimentation set-up is known.

Example: Ramp metering, step 11

The evaluator and project designers construct a combined realisation, experimentation and evaluation planning, which schematically might look like the following timeline:



4.6 Reflection on proposed evaluation methodology

In this final section a reflection on the proposed evaluation methodology is presented. This section starts with a reflection on the scope of the proposed evaluation methodology. Based on this further work and research required are identified. Finally, relations between the proposed evaluation methodology, DTM design and decision-making are discussed.

Reflection on scope of proposed evaluation methodology
Focus of the methodology in this report is on developing logic models and formulating evaluation questions based on these logic models. See Figure 4.1 for an illustration. Logic model development forces evaluators, decision-makers and project designers to cooperate and identify project assumptions and external influences. Evaluation questions are then derived from the logic model. These steps are not dealt with in detail in existing evaluation methodologies. Added value of the theory-based evaluation methodology proposed in this report is thus not found in using "new" evaluation methods or statistical tools, but in asking "the right questions" and using (existing) evaluation methods wisely to answer them.

Although the methodology proposed in this report is focused on ex-post evaluation of DTM projects for regular situations, it can also be used for ex-ante evaluation and evaluation for non-regular conditions. In Section 5.5 of the next chapter this will be explained using the case study.

As shown in this chapter, logic models are also useful tools to help determine evaluation methods, experimentation set-up, data collection methods and data analysis methods. The evaluation methodology proposed in this report thus forms a *basis* for an evaluation methodology that offers improvements for weaknesses of the current evaluation methodology of Rijkswaterstaat.

Further work and research required
Steps not discussed within this report are detailed evaluation design, data collection, data analysis and feedback of evaluation findings to project theory and to decision-makers. Further work and research is required to include these steps into the methodology. The evaluation methodologies studied in this report provide good sources of information.

Further research is also required to integrate socio-economic evaluation into the methodology. To this end Appendix A includes a discussion of the applicability of three methods for socio-economic evaluation for DTM projects.

As stated before, the methodology can also be applied for ex-ante evaluation and evaluation of non-regular situations. However, further work on the methodology is required to incorporate these aspects.

Relations between logic models, DTM design and decision-making
To design logic models, information is needed from DTM design and its objectives. Logic models explain which assumption project designers make and what final impacts decision-makers desire.

Therefore, the creation of logic models should not be the sole responsibility of project evaluators, but of decision-makers and project designers as well. The development of logic models is therefore an interdisciplinary activity, promoting communication and understanding between disciplines.

The earlier in the DTM design process the development of logic models is started, the more useful they will be for evaluation purposes. Developing logic models during the DTM design process "will provide legitimation for the evaluation approach chosen and will provide some guarantee that evaluation results will be used" (Verma and Burnett, 1999). Furthermore, the more time is available to develop and discuss logic models, the more assumptions can be uncovered and the sharper assumptions can be formulated.

The development of logic models from an early moment in the DTM design process has other benefits as well:

- By creating logic models, the design of DTM measures can be critically reviewed and improved, by asking questions such as:
    - Are assumptions made in logic models plausible?
    - Are there other, more effective ways, to reach these (intermediate) outcomes?
    - Are expected impacts actually helpful to solve problems?
- Logic models are potentially useful communication tools. They can help to create common understanding between decision-makers, key stakeholders, project designers and evaluators. Common understanding is created about expected impacts, required resources to achieve these impacts and how these resources will be used to achieve desired impacts.

Ideally, logic models of DTM projects should be products of the "Sustainable traffic management[32]" process. Currently the process of "Sustainable traffic management" is a linear process, which ends with the delivery of a governmental agreement on the realisation of DTM projects. This process is not yet a closed policy cycle, such as the example shown in Figure 2.1. Integrating implementation and evaluation of DTM projects into the "Sustainable traffic management" process will make it cyclic. Lessons learned in project implementation and evaluation are used for subsequent problem definition and project design.

As stated before, feedback of evaluation findings to decision-makers is outside the scope of this research. Therefore, the integration of logic model development and the subsequent theory-based evaluation approach into the "Sustainable traffic management" process, is not parts of this report.

[32] In Dutch: "GebiedsGericht Benutten (GGB)"

4.7 Conclusions

In this chapter a methodology for ex-post evaluation of DTM was proposed, based on the theory-based approach. Hereby the following question was answered:

Research question of Chapter 4

How can this most promising improvement be incorporated into a proposal for a new evaluation methodology?

To be able to apply theory-based evaluation to DTM projects, project theory first needs to be made explicit. Logic models are used for this purpose and adapted to DTM projects. These are graphical depictions of chains of causal assumptions linking project resources (human resources, hardware and software) to expected impacts. Based on these logic models, relevant evaluation questions are formulated in a structured manner. The formulated evaluation questions and logic models are then used to select appropriate evaluation methods, indicators and experimentation set-ups. External influences on evaluation are explicitly taken into account in these selections.

In the scope of this report the methodology ends with the selection of an evaluation approach and experimentation set-up. Further work is needed to complete the proposed evaluation methodology.

In this chapter the proposed methodology was only applied to a fictitious example of a local DTM project. Whether the approach can be applied to a real-life case of network-wide DTM, has not been discussed yet. Therefore, in the next chapter the proposed evaluation methodology is applied to a real-life, network-wide DTM project.

5. Case study proposed evaluation methodology

5.1 Introduction

In the previous Chapter a new methodology for ex-post evaluation of DTM projects was proposed. The methodology uses the theory-based approach, which was described in Section 3.5. In this section an attempt to apply the proposed evaluation methodology to a real-life, network-wide DTM project is described. The research question that is answered in this Chapter is:

Research question of Chapter 5

Is it possible to apply the proposed evaluation methodology to a real-life, network-wide DTM project?

Furthermore, the case study gives an indication of strengths, weaknesses, opportunities and threats of the proposed evaluation methodology, which will be used in the SWOT analysis of Chapter 6.

The chapter starts with a description of the case, the "Field Test Traffic Management Amsterdam[33]". Next, results of the step-by-step approach of the proposed evaluation methodology will be discussed, starting with the development of the logic model for (part of) the case and ending with the development of evaluation questions and considerations for the selection of evaluation methods and experimentation set-up. A reflection on the case study is given in Section 5.5.

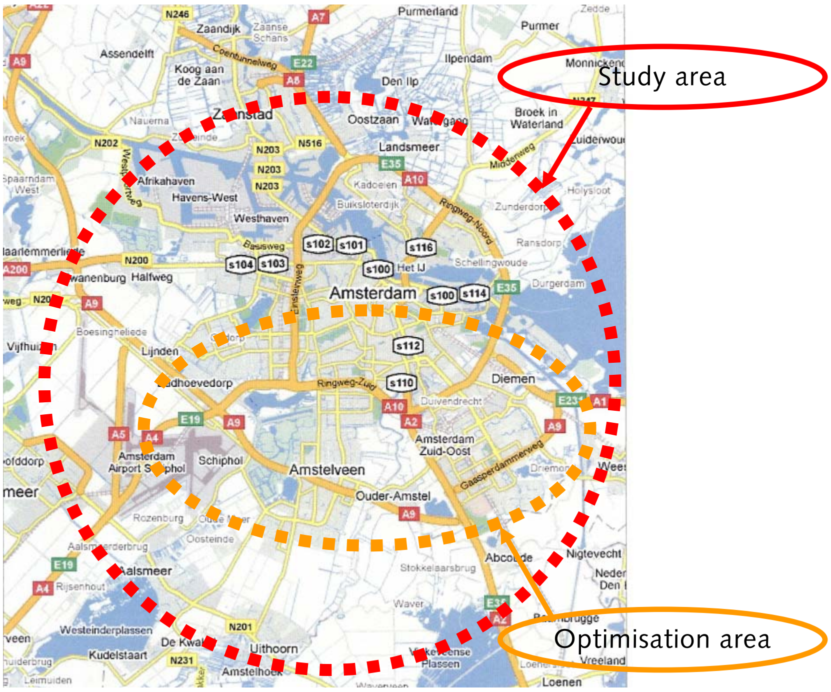
Findings of the case study are used in the next chapter to perform SWOT analysis of the proposed evaluation methodology. In this chapter literature on theory-based evaluation and logic models is combined with findings of the case study.

5.2 Introduction to the case

The "Field Test Traffic Management Amsterdam[33]" will be the first large-scale application of pro-active, coordinated and network-wide DTM in the Netherlands. From now on "Field Test" will be used to refer to this project. Where "FileProof A10", described in Section 2.4, focuses on improving traffic flow on the A10 ring road, the "Field Test" focuses on the development, experimentation, monitoring and evaluation of a prototype for anticipative, coordinated traffic management system that will be implemented on both the main and secondary road networks on the south side of Amsterdam. Such a traffic management system will be unique in the world. Figure 5.1 shows the part of the Amsterdam network that will be optimised within the scope of the "Field Test".

[33] In Dutch: "Praktijkproef Verkeersmanagement Amsterdam"

Figure 5.1: Geographic scope of the "Field Test Traffic Management Amsterdam"
Source: (Arane Adviseurs in Verkeer en Vervoer, 2008), adapted



The objective of this traffic management system is to improve traffic flow on the south side of Amsterdam, see the "optimisation area" in Figure 5.1. A secondary objective is to improve travel time reliability on the south side of Amsterdam. Boundary conditions are that emissions, noise hindrance, traffic safety and road user appreciation does not deteriorate (Rijkswaterstaat, 2009).

To be able to realise these objectives, similar control strategies (approaches that can be applied and combined to reduce traffic problems within networks) will be used as in "FileProof A10":
- Rerouting of traffic "on a distance of the ring road". This can be done by rerouting traffic via other motorways such as the A9;
- Rerouting of traffic via non-congested sections of the A10 ring road;
- Optimising traffic flow on the A10, A9 and N201;
- Improving outflow of traffic from the A10 to other motorways;
- Improving outflow of traffic to local roads, N201 and off ramps A9, A4;
- Controlling inflow of traffic from local roads (except the four main axes of Amsterdam) to the A10;
- Controlling inflow of traffic from the four main axes of Amsterdam to the A10;
- Controlling inflow of traffic from the A1, A2, A4 and A9 to the A10;

Like in "FileProof A10" these control strategies will be utilised to implement the "Network vision North-Holland[34]", see Section 2.4 for details.

[34] In Dutch: "Netwerkvisie Noord-Holland"

The main question the "Field Test" aims to answer is "is active, coordinated and network-wide traffic management a (cost-) effective means to improve the performance of networks (Rijkswaterstaat, 2009)?" Realisation and evaluation of the "Field Test" will be done in phases, starting in 2010. The integral final evaluation will be concluded in 2011, answering the main question.

To be able to answer this main question, evaluation of the "Field Test" includes evaluation of:
- Socio-economic impacts;
- Impacts on traffic flow at various scales;
- Impacts of travel time reliability;
- Road user appreciation;
- Impacts on emissions, noise hindrance and traffic safety.

The study area for these evaluations is shown in Figure 5.1.

Evaluation scope is reduced for the case study. The proposed methodology is used for evaluation:
- During and after realisation of the "Field Test".
- Of impacts on traffic flow and travel time reliability.
- Of one control strategy of the "Field Test". The control strategy "rerouting traffic on a distance of the ring road" was selected for this. This control strategy was selected because it is complex, incorporates both technical and user-response aspects and requires a network-wide approach.

It is expected that if it is possible to apply the proposed methodology with the described scope, it will be possible to apply it for all control strategies combined and all its impacts. In Section 5.5 a reflection on the selected scope for the case study is given.

5.3 Developing the logic model

In the previous section the control strategy of "Rerouting traffic on a distance of the ring road" was selected for the case study. This section starts with explaining this control strategy and developing a logic model for it. In the next section this logic model is used to develop a possible evaluation approach.

Rerouting at a distance of the ring road
"Rerouting at a distance of the ring road" is used to reroute traffic around congested parts of the A10 via other motorways. It can be used to relieve both regular congestion (congestion which appears at regular intervals during rush hours) and non-regular congestion (caused by for example incidents or events).

Communication to road users is done via “dWiSta panels”, a specific type of Variable Message Signs that are integrated into signposts. These panels are placed on points where road users decide on alternative routes. They inform road users about congestion and display route advices. Communication to road users may also be done using Dynamic Route Information Panels (DRIPs) or ShoulderDRIPs[35]

Figure 5.2 presents an example of the application of this control strategy. An incident causes congestion on the south side of the A10. Road users coming from the A1 are informed about this congestion. Road users that want to reach the A10 West are advised to use an alternative route via the A9. An example of a dWiSta panel is included in Figure 5.2.

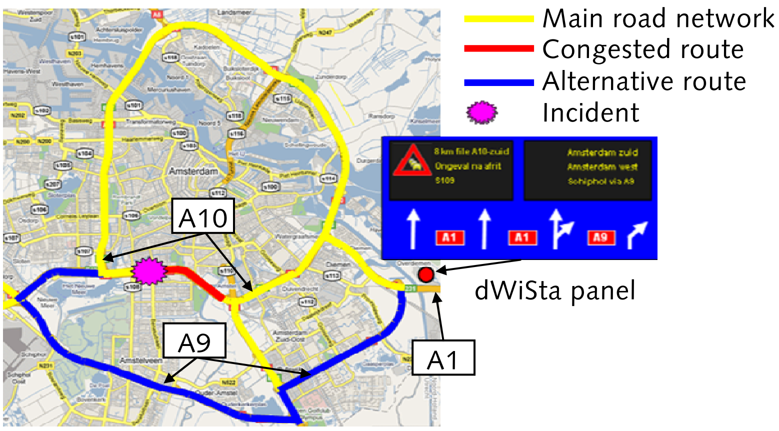


Figure 5.2: Illustration of the control strategy “rerouting at a distance of the ring road”
Including example of a message displayed on a dWiSta panel

Drawing the logic model
The development of the logic model for this control strategy starts with defining required resources and expected impacts using project documents. Resources that are part of the “Field Test”, but that are not used for this control strategy, are be included in the logic model. Some resources that are included may however also be part of resources required for other control strategies. In this case study focus is on resources required for regular congestion. Using DTM for non-regular congestion may require additional resources to be deployed such as arrangements between road authorities and emergency services and incident detection systems.

Table 5.1 lists elements required for the logic model that could be identified from project documents. Information about deployment of resources and traffic responses is missing, since assumptions made in earlier design phases were not explicitly documented.

[35] In Dutch “BermDRIPs”. ShoulderDRIPs are Variable Message Signs placed on shoulders of roads, as opposed to DRIPs, which are placed over road surfaces on gantries.

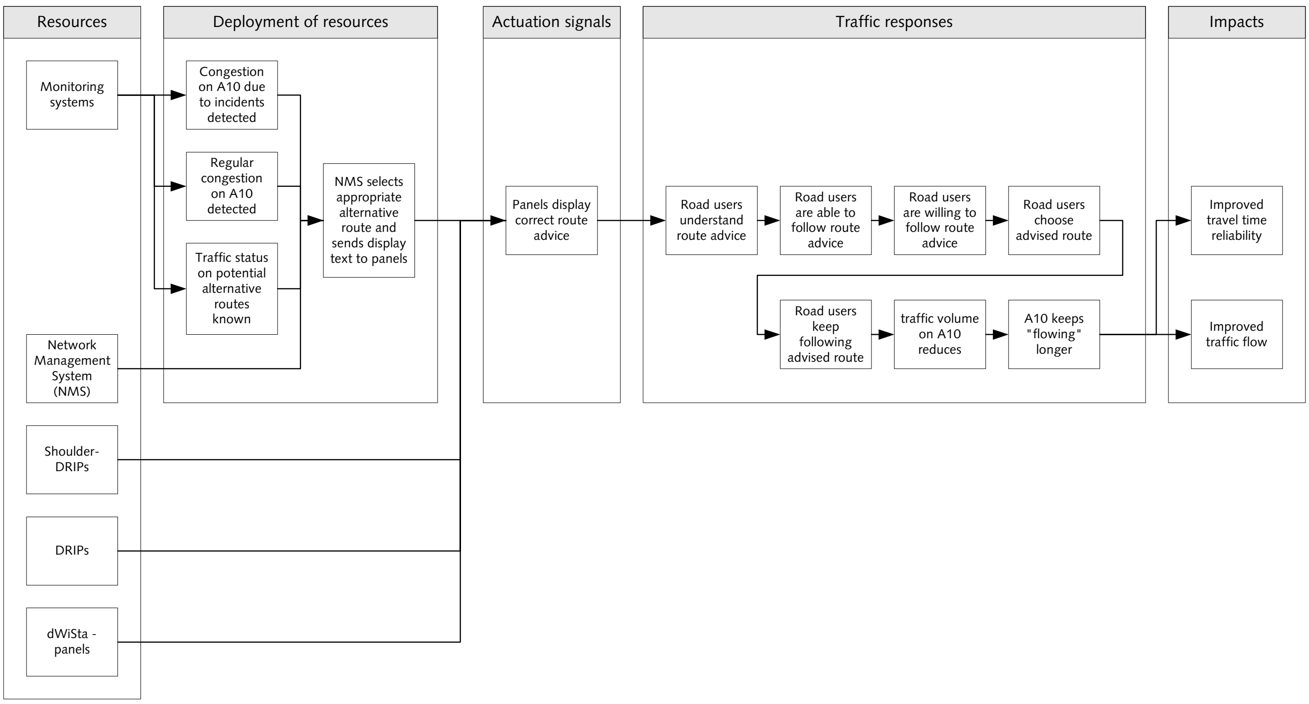
Table 5.1 Elements for the logic model, based on project documents
Sources: (Rijkswaterstaat, 2009; Gemeente Amsterdam et al, Not dated; Arane Adviseurs in Verkeer en Vervoer, 2008)

Resources	Deployment of resources	Actuation signals	Traffic responses	Impacts
• Monitoring systems • Network Management System (NMS) • dWiSta panels • Dynamic Route Information Panels (DRIPs) • ShoulderDRIPs	• Reroute on a distance of the ring road	• Panels on regional "decision points" display route advice to use A10 or other motorway	-	• More reliable travel times south side Amsterdam • Improved traffic flow south side Amsterdam

To be able to construct the logic model, information missing in Table 5.1 was reconstructed. This was done in a workshop with project designers and various experts, by interviewing experts and "brainstorming" by the author.

The resulting logic model is displayed in Figure 5.3. In Appendix D the logic model is displayed on a larger scale. The sequence of "understand – able – willing" used in the logic model is a common sequence used in user response evaluation (Lambers, 2008).

Figure 5.3: Logic model for the control strategy "rerouting at a distance from the ring road" of the "Field Test"



Note that the logic model is based on assumptions, which need to be tested in evaluation. In reality DTM measures may not function as assumed in this logic model. Assumptions made in the logic model in Figure 5.3 can be "read" as follows:

- **If** monitoring systems are functioning correctly, **then 1)** congestion due to incidents on the A10 is detected **or 2)** regular congestion on the A10 is detected **and 3)** traffic status on potential alternative routes is known.
- **If** congestion due to incidents on the A10 is detected **or if** regular congestion on the A10 is detected **and if** traffic status on potential alternative routes is known, **then** the Network Management System selects an appropriate alternative route via other motorways and transmits route advice to be displayed to panels.
- **If** the NMS sends route advice to be displayed on panels, **then 1)** DRIPs will display this route advice correctly **and 2)** shoulder DRIPs will display this route advice correctly **and 3)** the dWiSta panels will display this route advice correctly.
- **If** panels display route advice correctly, **then** road users understand displayed route advice.
- **If** road users understand displayed route advice, **then** they are able to follow it.
- **If** road users are able to follow displayed route advice, **then** they are willing to follow displayed route advice.
- **If** road users are willing follow displayed route advice, **then** they will actually follow displayed route advice.
- **If** road users are following advised routes, **then** they will keep following advised routes and will not deviate from it.
- **If** road users keep following advised routes, **then** traffic volume on congested parts of the A10 will reduce.
- **If** traffic volume on congested parts of the A10 reduces, **then** the A10 will keep "flowing" longer.
- **If** the A10 keeps "flowing" longer, **then 1)** travel time reliability in the network will improve **and 2)** traffic flow in the network will improve.

The logic model in Figure 5.3 should be seen as an intermediary product and not as the "final" logic model to be used for evaluation. The following remarks can be made about the developed logic model:

- At the time of writing the way monitoring systems detect congestion, determine the status of the network and how the Network Management System (NMS) is expected to select alternative routes is not yet known. "Black boxes" are therefore for the functioning of these resources. These "black boxes" will need to be clarified once more information on expected functioning of monitoring systems and the NMS is available.
- Panel design, combinations of panels used and "text strategies" influence whether road users understand displayed route advice or not and are able and willing to follow route advice. These are not yet known at the time of writing.

- The assumption that "keeping the A10 flowing longer" will lead to improved travel time reliability and reduced congestion is one of the main thoughts of the "Network vision North-Holland[36]". In order to properly evaluate this assumption, various sub-assumptions underlying this assumption need to be made explicit.
- Assumptions could not yet be formulated in such a way that they can be used to formulate evaluation questions sharply enough. The final version of the logic model should result in evaluation questions that cannot be interpreted in various ways, thus for which an unequivocal answer is possible.

Once missing information is known, the logic model can be adapted. The general structure of the logic model will not change because of this. The logic model will become more detailed, since sub-assumptions used in "black boxes" included in the logic model are made explicit. The logic model will however become more suitable as a basis for evaluation.

Adding external influences
Next, external influences are identified and added to the logic model. Identifying external influences was done based on external influences mentioned in the current evaluation methodology of Rijkswaterstaat, information gathered in interviews and insights of the author. Note that the list of external influences identified in this case study is likely not complete and should be updated as the project evolves.

Three external influences directly affect processes of the control strategy of "rerouting at a distance of the ring road". Below an explanation of these external influences will be given.

- Static signposts may be conflicting with route advice displayed by panels. This may influence whether road users understand route advices and continue following advised routes once they have chosen to follow them. At the time of writing it was not yet known whether adaptations of static signposting would be included in the scope of the project. Therefore, it is included as an external influence.
- The information in-car navigation systems provide may be conflicting with route advices displayed on panels. This may thus influence whether road users are willing to follow advised routes.

[36] In Dutch: "Netwerkvisie Noord-Holland"

Other external influences indirectly affect processes of the studied control strategy. These external influences affect whether, where and when road users decide to travel and via which route, thus affecting traffic volumes on the A10 (and the rest of the network):

- The moment in time;
- Price or mobility policies;
- Construction works;
- Incidents;
- Events;
- Economic developments (for example the credit crisis, changes in gas prices, increased unemployment)
- Changes in origin-destination patterns (due to spatial-economic developments such as the opening of a new business park);

Weather conditions have both direct and indirect effects on processes within the logic model:

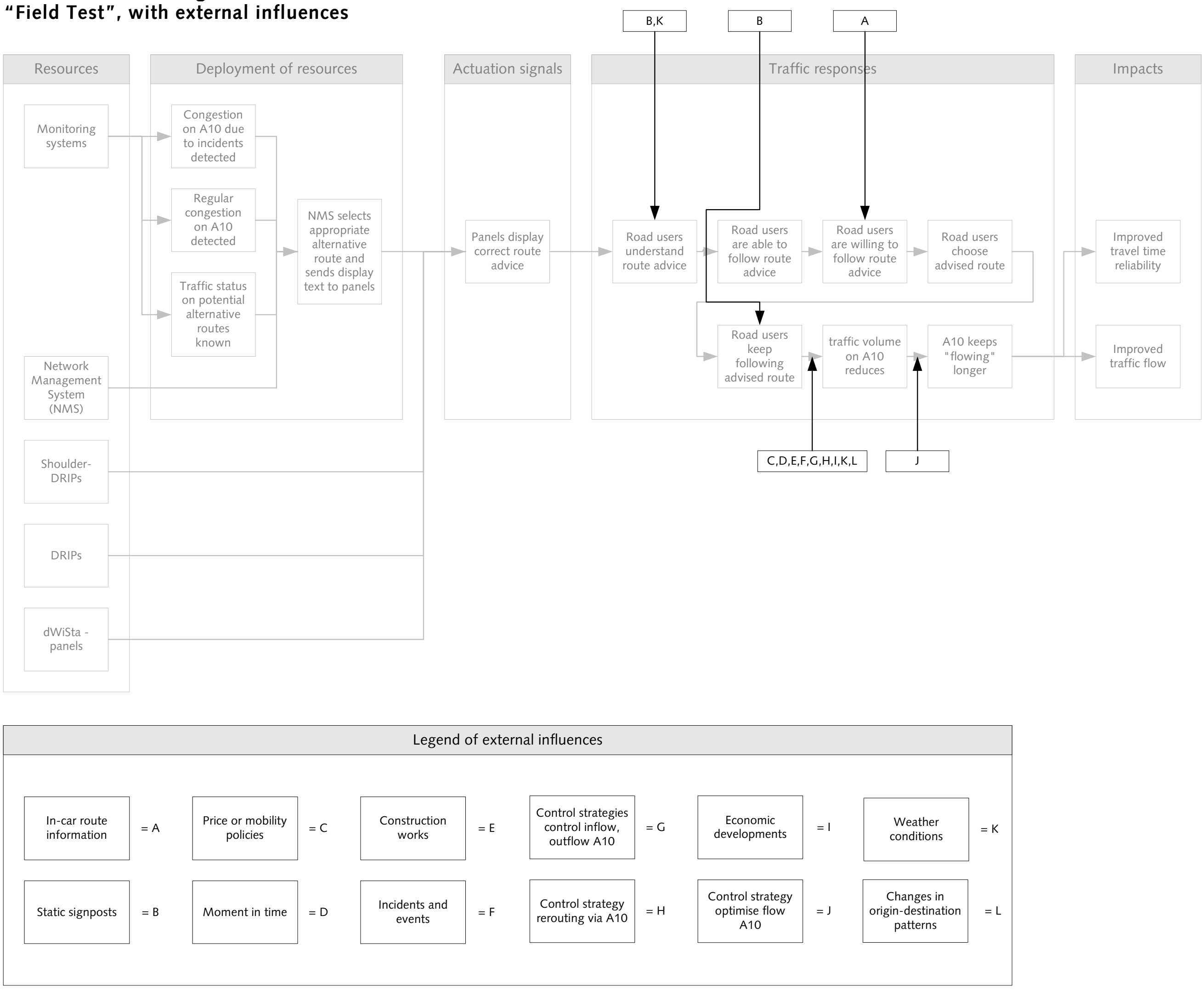
- The direct effect consists of influencing whether road users are able to read and thus understand route advices displayed on panels.
- The indirect effects consist of reductions of capacity of the network and impacts on traffic volumes within the network. Weather conditions thus affect infrastructure supply and demand.

Finally, other control strategies of the "Field Test" can affect traffic volume on the A10 or the traffic flow on the A10. For the logic model of the control strategy "rerouting at a distance of the ring road", these are included as "external" influences. In reality, these external influences are part of the "Field Test" and interactions between them need to be evaluated. In the reflection on the scope of the case study in Section 5.5 this is discussed. The control strategies that also influence traffic volume or traffic flow on the A10 are:

- Rerouting of traffic via non-congested sections of the A10 ring road;
- Optimising traffic flow on the A10;
- Improving outflow of traffic from the A10 to other motorways;
- Improving outflow of traffic from the A10 to local roads, N201 and off ramps A9, A4;
- Controlling inflow of traffic from local roads (except the four main axes of Amsterdam) to the A10;
- Controlling inflow of traffic from the four main axes of Amsterdam to the A10;
- Controlling inflow of traffic from the A1, A2, A4 and A9 to the A10.

The resulting logic model with external influences is shown in Figure 5.4. Like in the example of the previous chapter, external influences that affect processes directly are placed "on the boxes" and above the diagram and external influences that affect subsequent processes indirectly "on the arrows" and below the diagram. A legend is used for external influences. The logic model is shown on a larger scale in Appendix E.

Figure 5.4: Logic model for the control strategy “rerouting at a distance from the ring road” of the “Field Test”, with external influences



5.4 Developing an evaluation approach

In this section, the developed logic model is used to create an evaluation approach. First of all, evaluation questions are derived directly from the logic model shown in Figure 5.4. Hereafter, a possible selection of evaluation methods and indicators and an experimentation set-up is made.

Formulating evaluation questions

Possible evaluation questions are formulated using the logic model developed in the previous section. This results in an overview of evaluation questions relating to technical, user-response and impact aspects. This overview allows a prioritisation of evaluation questions to be made by evaluators, project designers and decision-makers. Which evaluation questions should be selected for actual evaluation is not discussed in this report. This depends on preferences of evaluators, project designers and decision-makers.

The following evaluation questions are derived directly from the logic model of Figure 5.4:

1. Do monitoring systems correctly detect regular congestion? And congestion due to incidents? And do monitoring systems correctly determine traffic status of potential alternative routes?
2. If regular or non-regular congestion is detected and the status of potential alternative routes is determined by monitoring systems, does the Network Management System select appropriate alternative routes?
3. If the Network Management System selects appropriate alternative routes, do DRIPs, ShoulderDRIPs and dWiSta panels display correct route advices?
4. If panels display correct route advices, are road users able to understand it? Is this influenced by static signposting? And by weather conditions?
5. If road users are able to understand displayed route advices, are they able and willing to follow them? Is this influenced by in-car route information?
6. If road users are able and willing to follow route advices, do they follow advised routes?
7. If road users have chosen to follow advised routes, do they keep following these routes? Is this influenced by static signposting?
8. If road users keep following advised routes, does this lead to reductions of traffic volume on congested parts of the A10?
9. Does the A10 “flow” longer due to reductions of traffic volume on congested parts the A10?
10. Is travel time reliability in the network improved due to the longer “flowing” of the A10?
11. Is congestion in the network reduced due to the longer “flowing” of the A10?

Selecting evaluation methods and experimentation set-up

The selection of an evaluation approach will depend on many factors, including:

- Selected evaluation questions (evaluation focus);
- Realisation planning;
- Available evaluation resources.

A possible evaluation approach is sketched below. This approach is meant to illustrate how the logic model helps in selecting appropriate evaluation methods and experimentation set-up. For evaluation questions 8 to 11 indicators will be selected. It will be demonstrated how logic models help in this selection.

Focus of the evaluation approach is on evaluation for regular congestion, during or after realisation of the "Field Test". Section 5.5 reflects on the case study, including the applicability of the proposed evaluation methodology for non-regular situations and ex-ante evaluation.

The logic model shows that in evaluation questions 1 to 7 not many external influences need to be taken into account. Functioning of monitoring systems, the Network Management System and panels and road user responses to resulting actuation signals can be evaluated using methods common in DTM evaluation. A possible approach to answer (groups of) evaluation questions is explained below.

<u>Questions 1 to 3:</u>
Diagnostic tools can be used to detect abnormalities in loggings. These tools can be combined with visual observations to determine the functioning of monitoring systems, panels and the Network Management System.

<u>Question 4:</u>
Focus groups and questionnaires can be used to answer this question. Results should be compared for days with good and poor visibility due to weather conditions. Specific questions should be aimed at the comprehensibility of route advices displayed on panels in combination with static signposts.

<u>Question 5:</u>
A stimulus-response study can be used to determine percentages of road users that decide to follow displayed route advices. Measuring fractions of road users that choose directions corresponding to displayed route advices can do this. After this a comparison between these fractions before and after panels start can be made. Focus groups and questionnaires can be used to verify results. Specific questions should be aimed at impacts in-car route information has on choices road users make.

<u>Questions 6 to 7:</u>
Again, questionnaires and focus groups can be used to answer these questions. The possibility exists that road users follow route advices but deviate from advised routes further on. Therefore, specific questions should be aimed at determining if this is the case. Questions should be formulated about whether static signposts cause such deviations. Furthermore, traffic flows on advised alternative routes might be measured to determine whether these routes are used more frequently.

Once evaluation questions 1 to 7 are answered, assumptions about technical functioning of DTM measures and road user responses are tested. The result of this testing is that it is known to what extent part of project theory (the "left" side of the logic model) is correct. How to test assumptions about impacts of the "Field Test" has not been described yet. Evaluation questions 8 to 11 need to be answered to determine these impacts.

Questions 8 to 11

The logic model shows that from evaluation question 8 onwards, many external influences have to be taken into account in evaluation. This can be seen in Figure 5.4: just before "Traffic volume A10 decreases" many external influences "enter" the logic model. From this point on the logic model moves from local effects towards network-wide impacts.

The logic model thus indicates that that an evaluation approach for evaluation questions 8 to 11 should be selected "with care". There is a risk that changes in indicators caused by external influences are attributed to the "Field Test". Answering these evaluation questions using measurements before and after realisation of all measures should not be preferred.

A possible approach that both answers evaluation questions 8 to 11 and takes external influences into account is explained below. As a part of this approach possible indicators that can be used to answer evaluation questions 8 to 11 are listed. Hereafter, an experimentation set-up to determine these indicators is given.

Indicators that may be used to answer evaluation questions 8 to 11 are:
Question 8:
- Total inflow of vehicles to A10 from other motorways;

Question 9:
- A performance measure related to the Macroscopic Fundamental Diagram[37] of sections of the A10;
- Vehicle delay hours on sections of the A10;
- Starting and ending hours of congestion;

Question 10:
- Standard deviations of travel times for selection of origin-destination relations within the network;
- Skewness of travel times for selection of origin-destination relations within the network;

Question 11:
- A performance measure related to the Macroscopic Fundamental Diagram for (sub) networks;
- Vehicle delay hours on the network;
- Traffic flow at a selection of links within the network.

[37] Research on such a performance measure is being done as part of the Field Test Amsterdam

To reduce impacts of external influences, time between before and after study is reduced as much as possible. To achieve this reduction, before measurements are started when:

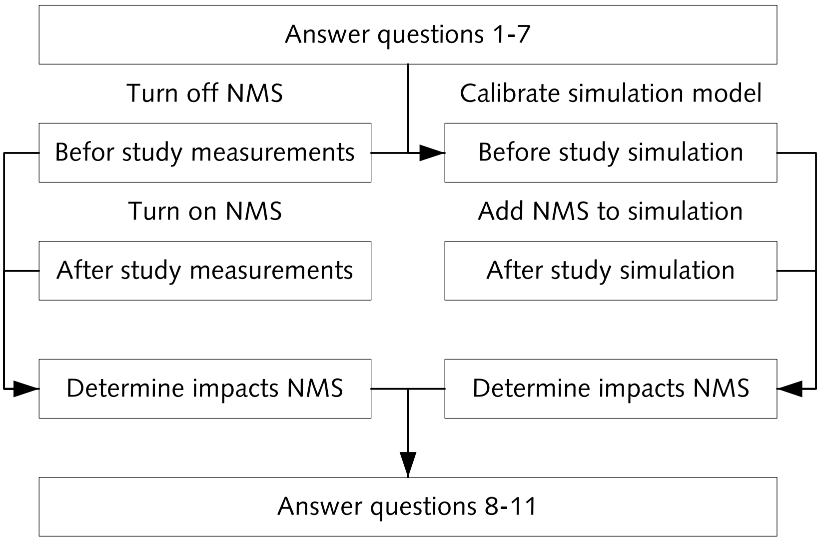
- All DTM measures are realised;
- All DTM measures, including the Network Management System function according to specifications;
- Evaluation questions 1 to 7 are answered;
- The Network Management System is "switched off" and panels no longer display route advices.[38]

After sufficient data[39] is gathered, the Network Management System is turned on again to gather after study data.
The "matching" approach used in the current evaluation methodology (see Section 2.3) can then be used to determine impacts of the "Field Test" based on before and after study.

A simulation study is used to further reduce the likelihood that changes in indicators caused by external influences are attributed to the "Field Test". The simulation model is (preferably) calibrated on data of the performed before study. After this, effects of DTM measures are modelled based on answers to evaluation questions 1 to 7. An advantage of simulation models is that (unknown) changes in impacts of external influences do not affect simulation results.

After simulation study, results of both studies are then compared. If both the before and after study and the simulation study show similar results, answers to evaluation questions 8 to 11 can be given. If this is the case, (unknown) changes in external influences have not impacted before and after study. The approach is illustrated in Figure 5.5.

Figure 5.5: Possible evaluation approach to answer questions 8-11



[38] "Switching off" DTM measures after they have been in operation will require strict arrangements between road authorities and clear communication to road users. In this report it is assumed this is not an issue.

[39] See Section 4.5 for some considerations to determine how many measurements are "sufficient"

5.5 Reflection on case study

This section reflects on the case study of the “Field Test”. First a reflection on the application of the proposed evaluation methodology is presented. Observations about strengths of the methodology and encountered difficulties in applying the methodology are made here. After this a reflection on the scope of the case study is given.

Reflection on application of evaluation methodology
First of all, the case study shows that it is possible to apply the proposed evaluation methodology to part of a real-life DTM project.

The logic model and evaluation approach developed in this case study should be seen as an intermediate result of the methodology. “Black boxes” currently exist within the logic model, which need to be made more explicit. The formulated evaluation questions cannot be answered unambiguously and are open to various interpretations. Therefore further work to make assumptions explicit, define indicators, update the logic model and reformulate the evaluation questions is required.

Three causes can be given for the fact that no “final” result could be produced in the duration of this research:

- The chosen project is still in its design phase at the time of writing; the logic model should evolve as project design evolves.
- Assumptions made in previous design phases have not been explicitly documented. This causes various expectations about the functioning of DTM measures to exist. By integrating the development of the logic model into the “Sustainable traffic management[40]” process this might have been prevented.
- Creating logic models is complex and experience with creating it needs to be gained.

The case study shows that for a complex, network-wide DTM project a structured approach towards creating evaluation questions is required. During the case study a workshop with project designers was organised to develop a logic model for the “Field Test”. The structured approach of Chapter 4 was not yet developed at the time of this workshop. The following difficulties were encountered during this workshop:

- Various views existed on the approach to develop the logic model: Should it be done “bottom up” (reasoning forward from the resources) or “top down” (reasoning backward from the expected impact)? Should specific scenarios be used to reconstruct assumptions made in previous project phases or should the discussion remain general in nature? The methodology of Chapter 4 might have prevented this difficulty from appearing.
- Various views existed on how resources should be deployed and what traffic responses and impacts are expected. This is caused by the fact that project objectives and assumptions made in previous design phases were not explicitly documented.

[40] In Dutch: “GebiedsGericht Benutten (GGB)”

The proposed evaluation methodology promotes an integral approach to evaluation. The developed logic model shows that various evaluation types (technical evaluation, user response evaluation, impact evaluation) are interrelated. Evaluation question 1 to 3 can be considered part of technical evaluation, questions 4 to 7 part of user-response evaluation and questions 8 to 11 part of impact evaluation. The logic model shows the relations between these evaluation types.

During the case study the logic model was a helpful tool in the selection of appropriate evaluation methods. The logic model helped to identify external influences that may not have been considered using the current evaluation methodology, such as in-car route information and static signposting. The logic model shows which external influences are relevant for specific evaluation questions. In this report a possible evaluation approach was developed using this information.

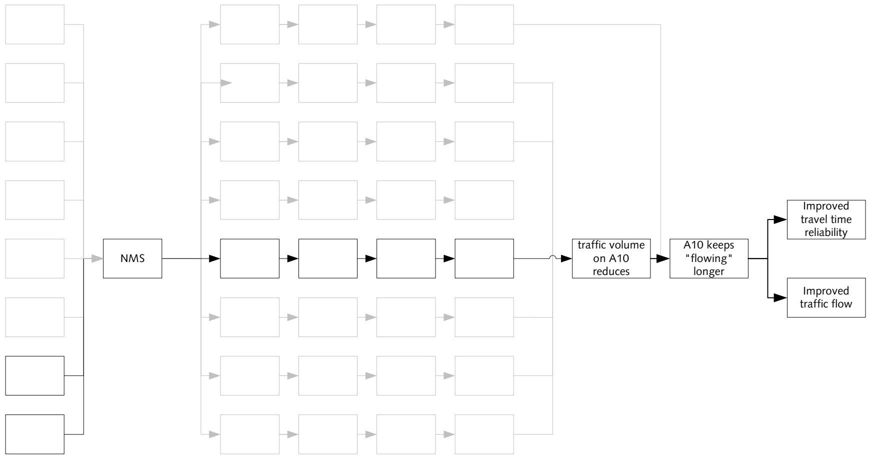
The logic model also proved to be a useful tool for evaluation planning and communication between various disciplines. Two examples are given:

- The logic model helps to communicate about the selected evaluation approach. For example, it helps to explain how evaluation questions were chosen and why a "traditional" before and after study is not recommendable to answer evaluation questions 8 to 11.
- The logic model can be used to determine interactions between evaluation types. Often evaluations are divided over separate evaluation teams. If this is the case, logic models can be used to determine interactions between these valuation teams. For example, the combined simulation/measurements approach used to answer questions 8 to 11 (impact evaluation) cannot be performed without results of questions 1 to 7 (technical and user response evaluation). Drawing boundaries between evaluation parts in the logic model helps to identify these necessary information flows.

Reflection on scope of case study
For the case study, evaluation scope was reduced. Possible implications if evaluation scope is expanded are discussed below.

The case study focused on one control strategy of the "Field Test". If all control strategies are taken into account, the resulting logic model will become more complex. The logic model developed in this case study forms part of the complete logic model for all eight control strategies. The logic model for "Rerouting at a distance of the ring road" will not change if it is placed within the complete logic model. The complete logic model will "diverge" into various "branches" (=control strategies) "after" the Network Management System. Since each control strategy contributes to achievement of the same impacts, the logic model "converges" again at the right side of the logic model. See Figure 5.6 for a sketch of the complete logic model.

. . . .
Figure 5.6: Sketch of expected shape of logic model once all eight control strategies are included
A sketch of the logic model developed in the case study is shown in black. Expected shape of logic model not in scope of the case study is shown in grey. After the NMS the logic model "diverges" into eight control strategies, one of which is studied in this chapter



Expected interrelations between control strategies were included in this case study by including other control strategies as "external" influences. At the location where these "external" influences enter the logic model for "Rerouting at a distance of the ring road", the complete logic model will "converge", see Figure 5.6. From this point forward, separating influences of the various control strategies in evaluation will be difficult, if not impossible. However, the evaluation approach sketched in this case study for evaluation questions 8 to 11 is applicable to determine the *combined* impacts of all eight control strategies, using the same indicators.

Within the case study, the methodology was applied only for evaluation for regular situations. At the time of writing it was uncertain whether the "Field Test Amsterdam" would be evaluated for non-regular situations. The proposed evaluation methodology is however also applicable to non-regular situations, since logic models can be developed non-regular situations as well as regular situations. For the "Field Test" it may be convenient to develop separate logic models for regular and non-regular situations. Non-regular situations may require extra resources and different evaluation methods. Simulation models may be required to develop reference situations for evaluation.

This case study focused on evaluation during or after realisation of the "Field Test". In order to check assumptions made in the logic model, project theory could also be tested in ex-ante evaluation. For example:

- Functioning of panels and road user responses can be tested in a driving simulator or on test sites outside the actual project area.
- Functioning of monitoring systems can be evaluated before operation of panels commences.
- Simulation study can help to formulate expectations about impacts of the "Field Test".

In this manner, ex-ante evaluation can increase certainty about correctness of project theory, as it is documented in the logic model.

5.6 Conclusions

In this chapter a case study of the "Field Test" was used to answer the following question:

Research question of Chapter 5

Is it possible to apply the proposed evaluation methodology to a real-life, network-wide DTM project?

The case study has shown that it is possible to apply the approach to real-life, network-wide DTM projects. It should be noted that the approach was only applied to one control strategy of the "Field Test". Whether it is possible to apply the approach to the entire project cannot be concluded with certainty. Furthermore, the "Field Test" is due to start in 2010 and will be evaluated in 2011, so feedback on the proposed methodology cannot be given yet. However, the methodology was applied to a complex control strategy and therefore it is expected it is possible to apply the methodology for the complete project.

The case study showed the added value of the proposed evaluation methodology:
- The proposed evaluation methodology promotes an integral approach to evaluation;
- The logic model proved to be a useful tool for evaluation planning and communication between various disciplines;
- A structured approach to developing evaluation questions is required in complex, network-wide DTM projects, which the proposed evaluation methodology provides;
- The logic model helped to identify external influences that might not have been considered using the current evaluation methodology
- The logic model proved to be a helpful tool in the selection of appropriate evaluation methods, indicators and experimental set-up, taking identified external influences explicitly into account.

The logic model and evaluation approach developed in the case study should be seen as an intermediate result of the methodology. Three causes can be given for the fact that no "final" result could be produced in the duration of this research:
- The chosen project is still in its design phase at the time of writing; the logic model should evolve as project design evolves.
- Assumptions made in previous design phases had not explicitly been documented. This causes various expectations about the functioning of DTM measures to exist.
- Creating logic models is a complex process and experience with creating it needs to be gained.

6.SWOT-analysis proposed evaluation methodology

6.1 Introduction

In Chapter 2 the following objective of ex-post evaluation was defined:

Objective of ex-post evaluation for DTM projects

Assessing the functioning and impacts of DTM measures and providing information on the use and allocation of public resources or the efficiency of DTM measures.

In Chapter 4 an evaluation methodology for ex-post evaluation of DTM projects was proposed as an alternative to the current evaluation methodology. In this chapter, a SWOT analysis for this proposed evaluation methodology is performed. This results in:

- Strengths of the proposed evaluation methodology that are helpful to achieve the objective of ex-post evaluation;
- Weaknesses of the proposed evaluation methodology that are harmful to achieve the objective of ex-post evaluation;
- Opportunities for the proposed evaluation methodology that are helpful to achieve the objective of ex-post evaluation;
- Threats to the proposed evaluation methodology that are harmful to achieve the objective of ex-post evaluation.

The following question is thus answered in this chapter:

Research question of Chapter 6

What are strengths, weaknesses, opportunities and threats of the proposed evaluation methodology?

The SWOT-analysis is performed using literature on theory-based evaluation and logic models and findings of the case study of Chapter 5. In Section 6.2 strengths of the proposed evaluation methodology are identified and in Section 6.3 its weaknesses. Next, opportunities and threats of the proposed methodology are discussed in Sections 6.4 and 6.5. Each section starts with an overview of the SWOTs, followed by an explanation. SWOTs of the proposed evaluation methodology are summarised in the conclusions of this chapter. In the next chapter these SWOTs will be compared with SWOTs of the current evaluation methodology to come to the final conclusions and recommendations of this report.

6.2 Strengths of proposed evaluation methodology

The SWOT analysis starts with a description of strengths of the proposed evaluation methodology in this section. When possible, strengths will be illustrated with the case study of the previous chapter.

Strengths of the proposed evaluation methodology, that are helpful to achieve the objective of ex-post evaluation, are that it:

- Is flexible to take any impact of DTM into account;
- Links the various evaluation types common in DTM evaluation to each other and links evaluation to other project phases;
- Explicitly evaluates the functioning of (individual) DTM measures and road user response to them by evaluating assumptions made in DTM design;
- Explicitly deals with causality issues and uses results of technical and user-response evaluation to support conclusions of impact evaluation;
- Is intuitive and convincing.

Below an explanation is given.

Flexible to take any impact of DTM into account
The methodology is flexible to take any impact of DTM into account. The methodology starts with expected project impacts and allows evaluators to design evaluation plans specific for these impacts. Furthermore, logic models are also applicable to non-regular situations.

Links evaluation types and project phases
The methodology integrates various evaluation types common in DTM projects and various project phases into one evaluation. The creation of logic models requires input of project designers and decision-makers, therefore firmly connecting evaluation design to project objectives and project design. Applying the evaluation methodology will automatically result in an overview of relevant evaluation questions and evaluation types, which can then be subdivided into various evaluation disciplines (for example technical evaluation, user-response evaluation and impact evaluation). This promotes an integral view on the evaluation process. Trade-offs between *learning* aspects of evaluation and *accountability* aspects are made explicit.

This strength of the methodology was observed in the case study; logic models was a useful communication tool and promoted a structured, integral approach towards project realisation planning, project design and evaluation planning and design.

Evaluates the functioning of DTM measures and road user-responses
The functioning of (individual) DTM measures and road user response to them are explicitly evaluated. Therefore, by using the proposed evaluation methodology one will learn more about what works, when it works and why. Since the methodology is based on project theory the methodology facilitates aggregation of evaluation results into a broader base of theoretical and project knowledge (Weiss, 1995; Pawson and Tilley, not dated).

Causality issues dealt with explicitly
Analysing project theory reduces the likelihood that measured changes in indicators, that are in fact not caused by DTM measures, are attributed to them. This likelihood is reduced first of all by verifying underlying assumptions DTM design. If one discovers that underlying assumptions are correct, ascribing (at least part of) measured changes in indicators to DTM measures becomes more plausible. Secondly, the creation of logic models and identification of external influences on DTM projects makes evaluators aware which conclusions about project impacts should be made with caution. In the case study for example, logic models helped to identify external influences on the project that may not have been considered using the current evaluation methodology.

Intuitive and convincing
The approach used is convincing and intuitive. Using assumptions behind project design as the basis of evaluation is intuitive. Evaluations that address theoretical assumptions embedded in projects may have substantial influence on both policy and popular opinion and can therefore be convincing (Pawson and Tilley, not dated; Weiss, 1995).

6.3 Weaknesses of proposed evaluation methodology

The SWOT analysis will continue with a description of weaknesses of the proposed evaluation methodology in this section. Again weaknesses will be illustrated with the case study of the previous chapter when possible.

Weaknesses of the proposed evaluation methodology, that are harmful to achieve the objective of ex-post evaluation, are that it:
- Does not (yet) include socio-economic evaluation;
- Does not result in "black or white" answers;
- Is complex and time-consuming, especially during evaluation design.

Below an explanation is given.

No socio-economic evaluation
The proposed evaluation methodology currently does not include socio-economic evaluation. The methodology therefore does not yet result in an evaluation that provides information on the use and allocation of public resources or the efficiency of DTM measures. The objective of ex-post evaluation cannot (yet) be fully met with the proposed evaluation methodology. A socio-economic evaluation can however be integrated into the methodology in a straightforward and logical manner, see Section 6.4. Further research is required for this.

No "black or white" answers
An evaluation using the proposed methodology may result in finding that parts of project theory are correct and other parts are incorrect. Some DTM measures may function as expected, some may not. Road user response to some DTM measures may be different than expected. Although this is realistic and useful information, decision-makers who need to make "black or white", decisions may not be happy with such "grey" answer (Pawson and Tilley, not dated; Weiss, 1998).

Complex and time-consuming
Applying the methodology can be complex and time-consuming and no experience has been gained in applying it yet. Especially during evaluation design, the methodology requires extensive cooperation and coordination between disciplines. This difficulty was noticed in the case study. The methodology requires underlying assumptions to be made explicit, which is not always common to do. In the case study the situation was encountered that assumptions made in previous design phases were not explicitly documented and needed to be reconstructed. Incorporating the approach into the "Sustainable traffic management[41]" process may reduce this weakness, see Section 6.4.

This weakness causes two pitfalls to be avoided when applying the methodology:
- Evaluators may feel tempted to focus solely on evaluation questions that are relatively easy to answer. These questions will generally focus on microscopic levels. Evaluation questions that are harder to answer may be ignored, those that focus on large scales. The *accountability* aspect of ex-post evaluation is then neglected and resulting evaluations may not be useful to decision makers.
- The second pitfall to be avoided is to get entangled in details while developing logic models. This may hinder progress of development of evaluation plans. Instead, "black boxes" and alternative paths can be used to keep development of logic models on track, see Section 6.4. This pitfall was noted when applying the methodology for the case study in a workgroup setting.

6.4 Opportunities for proposed evaluation methodology

This section lists opportunities for the proposed evaluation methodology, which are helpful to achieve the objective of ex-post evaluation.

[41] In Dutch: "GebiedsGericht Benutten (GGB)"

Opportunities for the proposed evaluation methodology, that are helpful to achieve the objective of ex-post evaluation, are that it:

- Can use "meta logic models" to reduce its complexity;
- Can be expanded with socio-economic evaluation in a logical and straightforward manner;
- Is suitable for expansion to ex ante and non-regular situations evaluation;
- Is complementary to the current evaluation methodology of Rijkswaterstaat;
- Is complementary to the "Sustainable traffic management[42]" process;
- Can include the various timescales over which impacts of DTM evolve.

Below an explanation is given.

"Meta logic models" reduce complexity

Part of the complexity of applying the methodology may be avoided by creating one "meta logic model" which contains various black boxes. For each black box a "micro logic model" can then be created. See Figure 6.1 for an illustration. If inputs and outputs of each "micro logic model" are defined using a "meta logic model" each "micro logic model" can be created with a smaller group of experts, reducing complexity of the methodology.

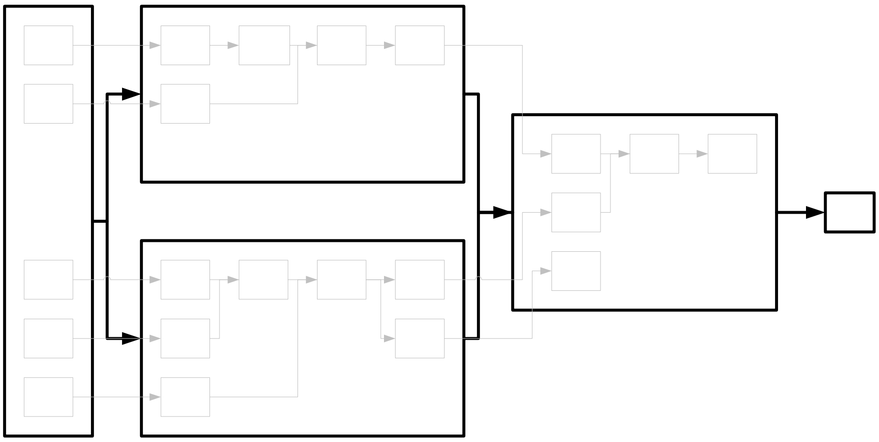


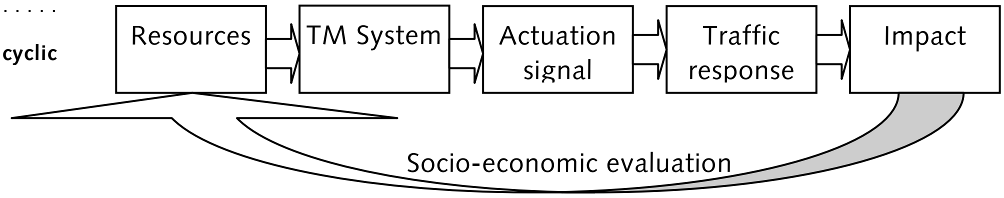
Figure 6.1: "Meta logic model" with several "micro logic models"
Thick black lines indicate "meta logic model", thin grey lines "micro logic models"

Socio-economic evaluation can be included in a logical manner

Socio economic evaluation can be added to the methodology in a logical manner. In Section 6.3 the absence of a socio-economic evaluation was identified as a weakness of the methodology as it is currently. So far logic models have been presented as a sequential models, but adding socio economic evaluation would make them cyclic. This can be seen in Figure 6.2. Socio-economic evaluation would compare final project impacts to invested resources (money, labour, time, materials, etc.)

[42] In Dutch: "GebiedsGericht Benutten (GGB)"

Figure 6.2: Socio-economic evaluation makes logic models cyclic



Suitable for expansion to ex ante and non-regular situations evaluation
The methodology proposed in this research focuses on ex-post evaluation and evaluation under regular conditions. The methodology may also be used for ex-ante evaluation and evaluation of DTM projects under non-regular situations. In the reflection of the case study (Section 5.5) some examples are mentioned.

Complementary to current evaluation methodology
An opportunity to the proposed evaluation methodology is that it is complementary to the current evaluation methodology of Rijkswaterstaat. If used with proper caution, the current evaluation methodology is a useful tool to answer evaluation questions formulated with the proposed evaluation methodology. Experience gained in applying the current methodology can be used within the proposed methodology.

Complementary to "Sustainable traffic management" process
The development of project logic models can be included into the "Sustainable traffic management" process. On the one hand, this would provide a solid basis for theory-based evaluation. On the other hand logic models will serve as a useful communication tool between decision-makers, stakeholders, project designers and evaluators. Logic models ask these involved parties to make their assumptions explicit and to reach consensus with their colleagues about what they are trying to do and why (McLaughlin and Jordan, 1999; Weiss, 1995; W.K. Kellog Foundation, 1998). Besides evaluation itself DTM design may also improve due to this.

Various timescales impacts DTM can be included
The various timescales over which impacts of DTM evolve can be incorporated into the methodology. As noted in Section 6.3, the methodology currently bases "final" project impacts on project objectives and sets this as the right boundary of logic models. This may however not be the actual "final" project impact, as discussed in Section 2.5. The right boundary of logic models would then be different from project objectives.

6.5 Threats to proposed evaluation methodology

Describing threats to the proposed evaluation methodology in this section completes the SWOT analysis.

Threats to the proposed evaluation methodology, that are harmful to achieve the objective of ex-post evaluation, are that:
- The utility of applying the (complex) methodology may be questioned;
- Logic models developed for evaluation may become a source of debate, hindering evaluation;
- It may be interpreted in a too flexible manner.

Below an explanation is given.

Utility may be questioned
Due to the complexity of applying it, the utility of the methodology may be questioned. Project designers may question the utility of unravelling assumptions and sub-assumptions underlying their design. Evaluators may feel tempted to choose an alternative "one-size, fits-all" approach (Pawson and Tilley, not dated; Weiss, 1998).

Logic model may become source of debate
Logic models themselves may become a source of debate. Logic models ask involved parties to make their assumptions explicit and to reach consensus with their colleagues about what they are trying to do and why (McLaughlin and Jordan, 1999; Weiss, 1995; W.K. Kellog Foundation, 1998). This may reveal conflicting interests and views. Although resulting discussions will generally be a very relevant, they may also cause delays to evaluation (McLaughlin and Jordan, 1999; Weiss, 1998). This threat was apparent in the workshop of the case study.

Flexibility may result in difficulty to perform meta-evaluation
The methodology does not specifically prescribe which indicators should be measured. A result could be that it is hard to perform meta-evaluation by comparing results of several evaluation studies. A "middle road" should therefore be found between prescribing specific indicators and evaluation methods and prescribing a "way of thinking" which allows the evaluator to develop a project-specific evaluation plan.

6.6 Conclusions

In this chapter a SWOT analysis was performed on the evaluation methodology proposed in Chapter 4. SWOTs are helpful or harmful to achieve the objective of ex-post evaluation defined in Chapter 2:

Objective of ex-post evaluation for DTM projects

Assessing the functioning and impacts of DTM measures and providing information on the use and allocation of public resources or the efficiency of DTM measures.

The following question was answered in this chapter:

Research question of Chapter 6

What are strengths, weaknesses, opportunities and threats of the proposed evaluation methodology?

SWOT analysis was performed using literature study on theory-based evaluation and logic models and the case study described in Chapter 5. Table 6.1 summarises SWOTs of the proposed evaluation methodology.

Table 6.1: SWOT analysis of proposed evaluation methodology
Based on (Wikimedia Foundation Inc., 2008)

	Helpful to achieve the objective of ex-post evaluation	**Harmful** to achieve the objective of ex-post evaluation
Internal origin (attributes of the methodology)	S • Evaluates functioning DTM measures and user-responses • Can evaluate any impact • Links evaluation types and project phases • Causality issues and trade-offs dealt with explicitly • Convincing, intuitive	W • Further development on socio-economic evaluation needed • No "black or white" answers • Complex, time-consuming, no experience
External origin (attributes of the environment)	O • "Meta logic models" • Suitable for expansion with socio-economic, ex-ante and non-regular evaluation. • Include timescales impacts • Complementary to current evaluation methodology • Incorporate into "Sustainable traffic Management" process	T • Utility may be questioned • Logic model may be debated, hindering evaluation • Possibly too flexible resulting in difficulties for meta-evaluation

The SWOT analysis results in a promising picture. Threats and weaknesses of the methodology mainly originate from the fact that the methodology requires further development. By using the listed opportunities, these weaknesses and threats can be reduced. Therefore, further development of the methodology is required. In the next chapter, SWOTs of the proposed evaluation methodology will be compared with those of the current evaluation methodology.

7. Conclusions and recommendations

7.1 Introduction

The objective of this research is to identify strengths, weaknesses, opportunities and threats of the current ex-post evaluation methodology for DTM of Rijkswaterstaat and to propose a new evaluation methodology

To achieve this objective, the following research questions were answered in this report:

- What are strengths, weaknesses, opportunities and threats of the current evaluation methodology?
- For which of the identified weaknesses do alternative evaluation methodologies not offer an improvement?
- What potential improvement offered in evaluation-related literature for these remaining weaknesses is most promising?
- How can this most promising improvement be incorporated into a proposal for a new evaluation methodology?
- What are strengths, weaknesses, opportunities and threats of the proposed evaluation methodology?

To be able to answer these research questions, the objective of ex-post evaluation has to be defined. In this report the following definition was chosen:

Objective of ex-post evaluation for DTM projects

Assessing the functioning and impacts of DTM measures and providing information on the use and allocation of public resources or the efficiency of DTM measures.

In Section 7.2 of this final chapter a summary of answers to the research questions stated above will be given. Using this summary of answers the final conclusions of this report can be drawn in Section 7.3, by answering the following central research question of this report:

Central research question

How do strengths, weaknesses, opportunities and threats of the proposed evaluation methodology relate to those of the current evaluation methodology of Rijkswaterstaat?

Based on the conclusions of Section 7.3, recommendations will be made in Section 7.4.

7.2 Main research findings

In this section the main findings of this research are summarised. This is done by answering each of the research questions recounted above. The findings will be used to support the answer to the main research question in Section 7.3.

Research question of Chapter 2

What are strengths, weaknesses, opportunities and threats of the current evaluation methodology?

The current evaluation methodology focuses on a before and after study experimental design to determine (direct) impacts of DTM projects on traffic flow. External influences are removed from the before and after study as much as possible, thus "matching" conditions of before and after study.

Strengths of the current evaluation methodology, that are helpful to achieve the objective of ex-post evaluation, are that it:

- Is based on a straightforward and intuitive approach;
- Is applicable to determine impacts on traffic flow of various projects without much adaptation;
- Has been applied frequently; experience is gained with applying the methodology;
- Connects evaluation to the policy cycle.

The current evaluation methodology has four weaknesses, however. These weaknesses, that are harmful to achieve the defined objective of ex-post evaluation, are that it:

- Does not analyse the functioning of DTM measures and road user responses to them (technical and user-response evaluation);
- Does not provide information on the use and allocation of public resources (socio-economic evaluation);
- Does not focus on all relevant DTM impacts;
- Focuses on before and after study to determine impacts of DTM. This experimental design is weak at proving causality between DTM measures and measured changes of indicators.

An opportunity for the current evaluation methodology is that the approach used is applicable to other types of impacts as well. By defining indicators for traffic safety, emissions and road user behaviour for example, the before and after study approach used can be applied to determine these impacts.

Three trends in DTM were identified as threats to the current evaluation methodology. These trends are:

- Diversification of objectives of DTM;
- Increasingly network-wide approach to DTM design;
- Increasing coordination between DTM measures.

These trends in DTM will make identified weaknesses of the current evaluation methodology more apparent in the future. Only improving the current evaluation methodology can mitigate these threats.

Research question 1 of Chapter 3

For which of the identified weaknesses do alternative evaluation methodologies not offer an improvement?

Three alternative evaluation methodologies considered in this research offer improvements for the identified weaknesses to some extent:

- The use and allocation of public resources is analysed in socio-economic evaluation;
- Although only in regular conditions, a wider range of impacts is taken into account;
- The functioning of DTM measures and road user response to them is analysed using technical and user-response evaluation.

Two weaknesses remain however:

- Proving causality between DTM measures and measured changes of indicators remains a weakness. Two reasons for this can be given:
    - The approach to impact evaluation is a similar before and after study approach as used in the current evaluation methodology.
    - Information of technical and user-response information is not explicitly used to support conclusions of impact evaluation. Each evaluation type is considered separately and is not explicitly linked to other evaluation types.
- The latter also causes the second weakness: the potential to learn from evaluation is not fully used. By treating evaluation types separately it is harder to learn what works, when it works and why.

Research question 2 of Chapter 3

What potential improvement offered in evaluation-related literature for these remaining weaknesses is most promising?

Potential improvements to these remaining weaknesses have been studied. The experimental design of impact evaluation can be adapted. However, before and after study with control group, trend or timeline analysis and alternating evaluation each have drawbacks or practical limitations. A hybrid simulation/modelling approach and data mining techniques are promising methods to improve confidence about causality between DTM measures and measured changes of indicators. Nevertheless, the potential to learn from evaluation is not improved if underlying assumptions are not evaluated.

Theory-based evaluation was selected as the most promising potential improvement, since it:

- Improves confidence about causality between DTM measures and measured changes of indicators;
- Improves the potential to learn from evaluation;
- Is in principle applicable to any DTM project.

The concept of theory-based evaluation is that project design is based on theory and that this theory should be the basis of evaluation. This project theory explains how projects are expected to achieve their desired final impacts. By making these assumptions explicit and focusing evaluation on these assumptions, evaluation findings will be more reliable and the potential to learn from evaluation is increased.

Research question of Chapter 4

How can this most promising improvement be incorporated into a proposal for a new evaluation methodology?

To be able to apply theory-based evaluation to DTM projects, project theory first needs to be made explicit. Logic models were used for this purpose and adapted to DTM projects. These are graphical depictions of chains of causal assumptions linking project resources (human resources, hardware and software) to expected impacts. A step-by step approach is proposed to create logic models for DTM projects.

Based on these logic models, relevant evaluation questions are formulated in a structured manner. The formulated evaluation questions and logic models are then used to select appropriate evaluation methods, indicators and experimentation set-ups. External influences on evaluation are explicitly taken into account in these selections.

In the scope of this report the methodology ends with the selection of an evaluation approach and experimentation set-up. Further work is needed to complete the proposed evaluation methodology.

Research question of Chapter 5

Is it possible to apply the proposed evaluation methodology to a real-life, network-wide DTM project?

The case study showed that it is possible to apply the methodology to real-life, network-wide DTM projects. The logic model and evaluation approach developed in the case study should be seen as an intermediate result of the methodology. Three causes can be given for the fact that no “final” result could be produced in the duration of this research:

- The chosen project is still in its design phase at the time of writing; the logic model should evolve as project design evolves.
- Assumptions made in previous design phases have not been explicitly documented. This causes various expectations about the functioning of DTM measures to exist.
- Creating logic models is a complex process and experience with creating it needs to be gained.

Research question of Chapter 6

What are strengths, weaknesses, opportunities and threats of the proposed evaluation methodology?

Strengths of the proposed evaluation methodology are that it:

- Is flexible to take any impact of DTM into account;
- Links the various evaluation types common in DTM evaluation to each other and links evaluation to other project phases. Trade-offs between various priorities within evaluation are made explicit;
- Explicitly evaluates the functioning of (individual) DTM measures and road user response to them by evaluating assumptions made in DTM design;
- Explicitly deals with causality issues and uses results of technical and user-response to support conclusions of impact evaluation;
- Is intuitive and convincing.

Weaknesses of the proposed evaluation methodology are that it:
- Does not (yet) include socio-economic evaluation;
- Does not result in "black or white" answers;
- Is complex and time-consuming, especially during evaluation design.

Opportunities for the proposed evaluation methodology are that it:
- Can use "meta logic models" to reduce its complexity;
- Can be expanded with socio-economic evaluation in a logical and straightforward manner;
- Is suitable for expansion to ex ante and non-regular situations evaluation;
- Is complementary to the current evaluation methodology of Rijkswaterstaat;
- Is complementary to the "Sustainable traffic management[43]" process;
- Can include the various timescales over which impacts of DTM evolve.

Threats to the proposed evaluation methodology are that:
- The utility of applying the (complex) methodology may be questioned;
- Logic models developed for evaluation may become a source of debate;
- It may be interpreted in a too flexible manner, resulting in difficulties for performing meta-evaluation.

The SWOT analysis results in a promising picture. Threats and weaknesses of the proposed methodology mainly originate from the fact that the methodology requires further development. Using the listed opportunities, these weaknesses and threats can be reduced. Further development of the methodology is therefore required.

7.3 Conclusions

Using the summary of answers to the research questions of the previous section, the central research question of this report can be answered:

Central research question

How do strengths, weaknesses, opportunities and threats of the proposed evaluation methodology relate to those of the current evaluation methodology of Rijkswaterstaat?

Complexity of the current and proposed methodology
The current evaluation methodology is a straightforward and intuitive approach and experience has been gained in applying it. The proposed evaluation methodology is on the one hand intuitive and convincing, but on the other hand complex and time-consuming and no experience has been gained in applying it.

[43] In Dutch: "GebiedsGericht Benutten (GGB)"

Suitability of current evaluation methodology
The current evaluation methodology is suitable to evaluate DTM projects if *each* of the following conditions is met:

- The project is affected by external influences which:
    - are known and of which the expected impact is small compared to impacts of DTM measures, or
    - of which the impact can be determined precisely,

  since the current evaluation methodology is weak at proving causality between DTM measures and measured changes of indicators;
- The project uses proven technology (i.e. of which the functioning has been evaluated), since the current evaluation methodology does not evaluatethe functioning of DTM measures and (sometimes unpredictable) road user response to them;
- The project only has an impact on traffic flow.

The current evaluation methodology is also suitable as an evaluation tool to be used within the proposed evaluation methodology to answer evaluation questions that meet these conditions.

Suitability of proposed evaluation methodology
The proposed evaluation methodology is suitable to evaluate DTM projects if *any* of the following conditions are met:

- The project is affected by numerous external influences with unknown or unmeasurable impact. The proposed evaluation methodology explicitly deals with causality issues. Causal attribution of effects to DTM measures is supported by evaluating assumptions behind DTM design;
- The project uses unproven technology (i.e. of which the functioning has never been evaluated). The proposed evaluation methodology explicitly evaluates the functioning of DTM measures and road user response to them;
- The project has impacts other than on traffic flow. The proposed evaluation methodology is flexible to take any impact into account.

The proposed evaluation methodology is better suited to cope with current trends in DTM than the current evaluation methodology. These trends are towards an increasing coordination of DTM measures, a network-wide approach to designing them and diversification of objectives of DTM projects. Further development of the proposed methodology is required however.

The proposed evaluation methodology promotes communication and cooperation between decision-makers, project designers and evaluators and explicitly links (ex-post) evaluation to previous project phases (definition of project objectives and project design). The proposed evaluation methodology is therefore suitable to be incorporated into the "Sustainable traffic management" process. Besides improving evaluation quality this can improve DTM design quality.

Both methodologies require expansion with socio-economic evaluation, however the proposed evaluation methodology can be expanded with socio-economic evaluation in a logical and straightforward manner.

7.4 Recommendations

Within the "National evaluation program[44]", a new evaluation guideline for DTM projects for Rijkswaterstaat will be drawn up. This report and the proposed evaluation methodology provide a basis for this new evaluation guideline.

This report ends with recommendations, which are based on the conclusions drawn in Section 7.3. These recommendations cover three subjects. Firstly recommendations will be made on how to evaluate DTM projects until the new evaluation guideline has been drawn up. Next, recommendations for the contents of the new evaluation guideline are made. Finally, to develop and improve this new evaluation guideline, further research is required. Recommendations to this end are presented at the end of this chapter.

DTM evaluation until a new evaluation guideline is drawn up
Below recommendations will be made on how to evaluate DTM projects until a new evaluation guideline has been drawn up:

- Before each evaluation project, determine whether the conditions mentioned under the first conclusion in Section 7.3 are met.
    - If these criteria are met, the current evaluation methodology can be applied to evaluate the project.
    - If these criteria are not met, follow the proposed evaluation methodology.
- Document the experiences gained with the application of the proposed evaluation methodology. This could include documenting the difficulties encountered when applying it, interviewing evaluators on their experiences after applying the proposed evaluation methodology and asking for potential improvements of the proposed evaluation methodology.

[44] In Dutch: "Landelijk evaluatieprogramma"

Development of a new evaluation guideline
Within the "National evaluation program", a new guideline for the evaluation of DTM projects will be drawn up. Below recommendations for the contents of this new guideline will be made.

The guideline should balance prescribing a flexible "way of thinking" with strictly prescribing an evaluation approach. The advantage of prescribing a flexible "way of thinking" is that it allows a project-specific evaluation plan to be drawn up. A result could however be that chosen indicators, definitions given to these indicators and approaches to determining them differ for each evaluation. Meta-evaluation, evaluation across several evaluation studies, would then be difficult to perform. Strictly prescribing which indicators to be measured, how these are defined and how these should be measured makes meta-evaluation easier. However, a risk is that this results in an evaluation approach that does not take project-specific aspects into account.

Below recommendations for the contents of the new guideline are made on how to achieve this balance:

- The proposed evaluation methodology forms the basis of the new evaluation guideline.
- A checklist of relevant impacts of DTM measures and possible road user responses to them is included in the evaluation guideline. This checklist can be updated as knowledge about DTM increases. The Finnish ITS evaluation methodology (Finnish Ministry of Transport and Communications, 2002) includes a listing of relevant DTM impacts.
- After logic model development and evaluation question formulation using the proposed evaluation methodology, the selected evaluation questions are subdivided into several evaluation types (technical, user-response, impact and socio-economic evaluation). The interdependencies and required information flows between these evaluation types are made explicit.
- For each evaluation type, a set of indicators that can be chosen to answer evaluation questions is prescribed, including a strict definition for each indicator. A useful list of indicators is included in the Finnish ITS evaluation methodology.
- For each evaluation type a set of evaluation methods that can be used to answer evaluation questions is included, including advantages and disadvantages of each method.
- The before and after study approach used in the current evaluation methodology is included into the guideline as one of the available methods. Strengths and weaknesses of the approach identified in this report are included.
- Conditions under which (not) to apply each evaluation method are explicitly prescribed. These conditions for the before and after study approach used in the current evaluation methodology are given in Section 7.3.
- The guideline promotes the use of a combination of methods to answer evaluation questions if projects are affected by many external influences at the corresponding locations of their logic models.

Recommendations for further research

To draw up the new evaluation guideline and improve it, further research is required. Below recommendations for further research are made.

Promising evaluation methods that are recommended for further research are data mining methods such as Principal Component Analysis (discussed in Chapter 3). A threat to the proposed evaluation methodology is that impacts of DTM projects that were not expected based on project theory may be missed in evaluation. Data mining techniques allow the uncovering of patterns in data, also those that were not expected. Such techniques can therefore be extremely useful as a means of exploratory data analysis and for revealing relationships that have not been anticipated. Further research is needed to answer the following questions:

- What data mining techniques are available?
- Which of these techniques can be applied to network-wide DTM projects?
- What are strengths and weaknesses of these techniques?
- When should these techniques be applied? And when not?
- What pitfalls should be avoided?

A second promising group of evaluation methods that is recommended for further research are hybrid simulations/measurements approaches. Two variants of such approaches are possible:

- A simulation model is calibrated on data of before study measurements. Microscopic effects of DTM measures are then added to the simulation model, based on actual measurements gathered in earlier phases of evaluation. Simulation results are then compared with after study measurements. The advantage is that *changes in* impacts of external influences do not affect simulation results, whilst these may affect before and after study based on measurements.
- A simulation model is calibrated on data of after study and microscopic effects of DTM measures, based on actual measurements gathered in earlier phases of evaluation. The effects of DTM measures are "removed" from the simulation model to recreate a reference situation. The advantage is that this allows an estimate of the DTM impacts in non-regular situations to be made, such as incidents. In such situations it is hard if, not impossible, to estimate DTM impacts using a before and after study approach.

Questions to be answered in further research are:

- What inputs are required for simulation models in such an approach?
- What requirements should simulation models meet?
- What are strengths and weaknesses of such an approach?
- When should such an approach be applied? And when not?
- What pitfalls should be avoided?

Socio-economic evaluation of DTM projects is another subject that is recommended for further research. DTM projects produce impacts that are hard to measure. It is even harder to quantify and attach a monetary value to these impacts. It is recommended to perform further research to be able to decide which socio-economic evaluation method should be prescribed under which conditions within the new evaluation guideline. Questions to be answered are:

- Which methods for socio-economic evaluation are available?
- What are strengths and weaknesses of these methods?
- When should which method be applied?
- How can socio-economic evaluation be integrated into the proposed evaluation methodology?

As a starting point for this research a discussion of three possible methods for socio-economic evaluation and their applicability to DTM projects is presented in Appendix A.

How to include the various timescales over which impacts of DTM evolve into the new evaluation guideline also requires further research. Direct DTM impacts lead to indirect impacts, since road users adapt to new situations. These indirect impacts consist for example of changes of destination, route, departure time and mode choices of road users. A new equilibrium is formed. In the new equilibrium congestion in networks can be at the same level as before implementation of DTM measures. However, more trips were made possible, which have an economic value. Questions to be answered are:

- How do impacts of DTM evolve over time?
- Should the proposed evaluation methodology take these timescales of impacts into account?
- If yes, how can the proposed evaluation methodology be adapted?
- If yes, how should socio-economic evaluation be adapted?

Fourth subject for further research is how to integrate logic model development into the DTM design process. Integrating logic model development and "Sustainable traffic management" process could do this. Logic models are useful tools that can improve communication between the various phases of this process and between the various parties involved (stakeholders, decision-makers, project designers, project evaluators). This would improve evaluation quality as well as DTM design quality. Questions to be answered are:

- What extra steps are needed in the "Sustainable traffic management" process and the proposed evaluation methodology to be able to integrate them?
- How can the information generated by the proposed evaluation methodology be fed back to "Sustainable traffic management"?
- What are advantages and disadvantages of integrating the proposed evaluation methodology and "Sustainable traffic management"?

These recommendations will result in an evaluation guideline that, if properly applied, produces evaluations that are more useful, convincing and reliable than using the current evaluation methodology.

Bibliography

Arane Adviseurs in Verkeer en Vervoer (2008). *Praktijkproef Amsterdam - GGB gebiedsgerichte uitwerking.*

B&A Group. *Effectevaluatie.* Retrieved december 1, 2008 from Centrum voor Beleidsevaluatie: http://www.bagroep.nl/index.php?id=145

Bouyssou, D. (1990). Building criteria: A prerequisite for MCDA. In C. Costa, *Readings in Multiple Criteria Decision Aid* (pp. 58-80). Berlin: Springer-Verlag.

Brand, D., Parody, T. E., Orban, J. E., and Brown, V. E. (2004). A Benefit/Cost Analysis of the Commercial Vehicle Information Systems and Networks (CVISN) Program. *Research in Transportation Economics , 8*, 379-401.

Brans, J., Vincke, P., and Mareschal, B. (1986). How to select and rank projects: The PROMTHEE method. *European Journal of Operational Research , 24*, 228-238.

Bruzon, V., and Mudge, R. (2007). *Development of a project evaluation methodology framework for Canadian Intelligent Transportation Systems.*

Chen, H-T and Rossi, P.H. (1987).The theory-driven approach to validity. *Evaluation and program planning*, 95-103.

Coëmet, M. (2003). *Kosteneffectiviteit benuttingsmaatregelen.* Adviesdienst Verkeer en Vervoer.

De Brucker, K., Verbeke, A., and Macharis, C. (2004). The applicability of multicriteria-analysis to the evaluation of Intelligent Transportation Systems (ITS). *Research in Transportation Economics , 8*, 151-179.

Dhevi Baskar, L., De Schutter, B., and Hellendoorn, H. (2008). Dynamic Speed Limits and On-Ramp Metering for IVHS using Model Predictive Control. *Proceedings of the 11th International IEEE Conference on Intelligent Transportation Systems*, (pp. 821-826). Beijing.

Finnish Ministry of Transport and Communications. (2002). *Guidelines for the evaluation of ITS projects.* Helsinki: VTT Building and Transport.

Gemeente Amsterdam, Provincie Noord-Holland, Rijkswaterstaat, Stadsregio Amsterdam and Regionaal Orgaan Amsterdam. *Netwerkvisie Noord-Holland - Op weg naar een optimaal gebruik van het Noord-Hollandse wegennet.*

Giorgi, L., and Tandon, A. (2002). Introduction: the theory and practice of evaluation. In L. Giorgi, and A. Pearman, *Project and policy evaluation in transport* (pp. 1-13). Aldershot: Ashgate Publishing Limited.

Haight, F. (2002). Evaluation of projects and programmes: principles and examples. In L. Giorgi, and A. Pearman, *Project and policy evaluation in transport* (pp. 181-208). Aldershot: Ashgate Publishing Limited.

Hall, R. W., Miller, M., and Khatak, A. (1996). Evaluating the effectiveness of integrated traffic corridors: theory and practice. *Journal of Intelligent Transportation Systems , 3:1*, 49-67.

Haynes, K. E., and Li, M. (2004). Analytical alternatives in intelligent transportation system (ITS) evaluation. *Research in Transportation Economics , 8*, 127-149

Hoogendoorn, S., and Van Zuylen, H. (2008). *AMICI: Advanced Multi-agent Information and Control for Integrated multi-class traffic networks.* Retrieved September 25, 2008, from Nederlandse Organisatie voor Wetenschappelijk onderzoek: www.tsl.tudelft.nl/tsl_platos_results.ppt

Kamel, B., Benasser, A., and Jolly, D. (2008). Flatness Based Control of Traffic Flow for Coordination of Ramp Metering and Variable Speed Limits. *Proceedings of the 11th International IEEE Conference on Intelligent Transportation Systems*, (pp. 838-843). Beijing.

Kee, J. E. (1994). Benefit-Cost Analysis in program evaluation. In J. S. Wholey, H. P. Hatry, and K. E. Newcomer, *Handbook of practical program evaluation* (pp. 456-488). San Francisco: Jossey-Bass.

Kotsialos, A., Papageorgiou, M., and Middelham, F. (2005). Local and optimal coordinated ramp metering for freeway networks. *Journal of Intelligent Transportation Systems, 9:4*, 187-203.

Lambers, M. (2008) *Human factors and human error as part of the system: A psychological approach.* Rotterdam: Rijkswaterstaat Centre for Transport and Navigation.

Lint, H. van (not dated). Traffic theory for intelligent transport systems for road transport. Version 25. Delft: TU Delft.

Macharis, C., Verbeke, A., and De Brucker, K. (2004). The strategic evaluation of new technologies through multcriteria analysis: the ADVISORS case. *Research in Transportation Economics,* 8, 443-462.

McLaughlin, J.A. and Jordan, G.B. (1999). Logic models: a tool for telling your program's performance story. *Evaluation and program planning,* 22, 65-72

Ministry of Public Works and Water Management. (2006). *Nota Mobiliteit deel IV - Na parlementaire behandeling vastgestelde PKB.* Moore, T., and Pozdena, R. (2004). Framework for an economic evaluation of transportation investments. *Research in Transportation Economics, 8*, 17-45.

Ministry of Transport, Public Works and Water Management, 2008. *Beleidskader Benutten. Eén van de pijlers voor een betere bereikbaarheid. Achtergronddocument.* DGP.

Newman-Askins, R., Ferreira, L., and Bunker, J. M. (2003). Intelligent Transport Systems evaluation: from theory to practice. *21st ARRB and 11th REAA Conference.* Cairns.

Ozbay, K., and Bartin, B. (2004). Estimation of economic impact of VMS route guidance using microsimulation. *Research in transportation economics , 8*, 215-241.

Palumbo, D. (1987). *The politics of program evaluation.* Newbury Park: Sage.

Panou, M. and Bekiaris, E. (2004). ITS clustering and terminology: one concept with many meanings. *Research in transportation economics, 8*, 49-67.

Pawson, R. and Tilley, N. (not dated). *Realist evaluation*

Rijkswaterstaat (2002). *Gebiedsgericht benutten*. Retrieved November 21, 2008, from http://www.avb-bureau.nl/GebiedsGerichtBenutten/index.html

Rijkswaterstaat (2008a). *Projectplan landelijk evaluatieprogramma (Concept).* Ministerie van Verkeer en Waterstaat, Dienst Verkeer en Scheepvaart. Delft: Rijkswaterstaat.

Rijskwaterstaat (2008b). *Maatregelencatalogus Benutten*. Retrieved September 25, 2008, from http://www.maatregelencatalogus.nl/

Rijkswaterstaat (2009). *Projectplan Realisatiefase Praktijkproef Verkeersmanagement Amsterdam. Concept, versie 1.0.* Not published.

Rijkswaterstaat Adviesdienst Verkeer en Vervoer (2002). *Leidraad model- en evaluatiestudies benuttingsmaatregelen.*

Rijkswaterstaat Dienst Verkeer en Scheepvaart (2008). *DVS agenda 2012. We stomen door.*

Rijkswaterstaat Traffic Centre the Netherlands (2008). *Evaluatiemethodiek - Richtlijn voor de ex-post evaluatie van benuttingsmaatregelen op het hoofdwegennet.*

Rijskwaterstaat Directie Noord Holland (2007). *Verbetering doorstroming A10 - Eindrapport.*

Rogers, P.J. (2000). Causal models in program theory evaluation. *New directions for evaluation,* 87, 47-55. San Francisco: Jossey-Bass

Rossi, P.H., Lipsey, M.W. and Freeman, H.E. (2004). *Evaluation – A systematic approach.* Newbury Park: Sage.

Roy, B. (1968). Classement et choix en presence de points de vue multiples - la methode Electre. *Revue francaise d'Informatique et de Recherche Operationelle , 8*, 57-75.

Saaty, T. (1988). *The Analytic Hierarchy Process.* New York: McGraw-Hill.

Schwarz, G., Van Schie, C., and Verroen, E. J. (2007). *Verkenning Praktijkproef Verkeersmanagement.* Amersfoort: Twynstra Gudde Adviseurs en Managers.

Stevens, A. (2004). The application and limitations of Cost-Benefit Analysis (CBA) for Intelligent Transport Systems. *Research in Transportation Economics , 8*, 91-111.

Taale, H., Schuurman, H. and Bootsma, G. (2002). *Evaluatie groot onderhoud aan de A10-West*. Adviesdienst Verkeer en Vervoer

Thill, J.-C., Rogova, G., and Yan, J. (2004). Evaluating benefits and costs of Intelligent Transportation Systems elements from a planning perspective. *Research in Transportation Economics , 8*, 571-603.

Tsekeris, T. and Stathopolous, A. (2006). Measuring variability in urban traffic flow by use of Principal Component Analysis. *Journal of Transportation and Statistics*, 9(1), 49-62.

Van Berkum, E., Kooten, J. v., Wee, B. v., Immers, B., Meurs, H., and Schuurman, H. (2008, April 14). Effecten van Netwerkmanagement. *Netwerk Management Magazine*.

Van Leusden, W., Bakker, M., Hepp, L., and Vreeswijk, J. (2008, July 16). Luchtkwaliteit: van randvoorwaarde naar doelstelling. *Netwerk Management magazine* , pp. 24-27.

Van Luling, B., Nuijten, M.J., Baggen, J.H. (2004). Bestuurlijke processen rond dynamisch verkeersmanagement. Organisatie en aanpak van netwerkbrede DVM-projecten. *Colloquium Vervoersplanologisch Speurwerk 2004.*

Verma, S. and Burnett, M. (1999). Adressing the attribution question in extension. *Annual conference of the American Evaluation Association*, Orlando, Florida, November 3-6.

Weiss, C.H. (1995). Nothing as practical as good theory: exploring theory-based evaluation for Comprehensive Community Initiatives for children and families. *New approaches to evaluating Community Initiatives. Concepts, methods and contexts,* 65-92.

Wholey, J. S., Hatry, H. P., and Newcomer, K. E. (1994). *Handbook of practical program evaluation.* San Francisco: Josey-Bass Publishers.

Wholey J.S. (1987). Evaluability Assessment: Developing Program Theory. *New Directions for Program Evaluation*, 33, 77-92.

Wikimedia Foundation Inc. (2008). *SWOT Analysis.* Retrieved November 27, 2008 from Wikipedia, the free encyclopedia: http://en.wikipedia.org/wiki/SWOT_analysis

W.K. Kellogg Foundation (1998). *Evaluation handbook.* Battle Creek: W.K. Kellogg Foundation.

W.K. Kellogg Foundation (2004). *Logic Model development guide. Using Logic Models to bring together planning, evaluation and action.* Battle Creek: W.K. Kellogg Foundation.

Zhang, X., Kompfner, P., White, C., and Sexton, B. (1998). *Guidebook for assessment of transport telematics applications: updated version. Converge Project deliverable D2.3.1.* ERTICO.

Appendix A: Socio-economic evaluation of DTM projects

In Section 3.2 it was concluded that each of the three alternative evaluation methodologies studied include some form of socio-economic evaluation, but there is no consensus about which method for socio-economic evaluation is preferred. This Appendix will therefore determine which methods for socio-economic evaluation are most applicable for DTM projects.

In Chapter 2 it was determined that an ex-post socio-economic evaluation is required to show to what extent the impacts generated by the project compensate for the invested public resources. Methods to achieve this can be subdivided into three types: Cost-Benefit Analysis (CBA), Cost-Effectiveness Analysis (CEA) and Multi-Criteria Analysis (MCA). A short description of each of these methods will be given below, followed by a discussion of the applicability of each method to (network-wide) DTM projects.

Multi-Criteria Analysis
Multi-Criteria Analysis has its origins in operations research. It allows the comparison of a number of actions of alternatives in terms of specific criteria, which represent the objectives and sub objectives of the decision-makers (De Brucker, Verbeke and Macharis, 2004). Due to its suitability for comparing alternatives it is mainly applied in ex-ante evaluation, but it could also be useful in ex-post evaluation.

Many variants of MCA have been developed over time, but all of them take approximately the same steps. These steps are:

1. Generation of the alternatives for the evaluation: the situation with DTM measures and the "zero-alternative" the situation without DTM measures;
2. Generation of a set of criteria;
3. Determining the score per criterion for both alternatives;
4. Aggregating the scores at each criterion to gain the final score per alternative. This involves attaching a "weight" to each criterion;
5. Integrating the results of the MCA into the decision-making process.

The variants of MCA mainly differ in the way the total score per alternative is calculated, that is the way in which the aggregate the scores per criterion. Many methods have been developed to achieve this aggregation. See (De Brucker et al, 2004) for an overview of aggregation methods. Much-used methods are the Analytical Hierarchy Process (Saaty, 1988), Electre (Roy, 1968) en PROMETHEE (Brans, Vincke, and Mareschal, 1986); variants have been developed for each of these.

The criteria for the MCA can be generated in several manners. Two possible approaches are the "top-down" approach, in which evaluation criteria are extracted from the project objectives, and the "bottom-up" approach, in which the criteria are based on the expected effects of the project. Combining these two methods will yield the most complete set of relevant criteria (Bouyssou, 1990).

Cost-Benefit Analysis
A second method for socio-economic evaluation is Cost-Benefit Analysis (CBA). In contrast to MCA, which has its origins in operations research, CBA is based on welfare-economic principles. In the case of the government welfare-economics assumes that it strives to maximize social welfare (Stevens, 2004). This means the government strives to use its scarce public resources for the projects that yield the biggest increase of social welfare, that is the projects which have the highest benefit-cost ratio.

CBA thus strives to express the costs and benefits of an entire project into a single monetary sum. To determine this sum in a proper way, the principles of good Cost-Benefit Analysis need to be taken into account (Moore and Pozdena, 2004):

1. The sum of costs and benefits should be determined based on a comparison of the situation *with* DTM measures and *without* DTM measures (the "zero" alternative). The latter is not necessarily the situation in which "nothing" has happened.
2. The perspective of the evaluation needs to be clear. The government will usually take the perspective of society as a whole and decide on a geographic delineation and a delineation in time.
3. All significant costs and benefits within the delineation need to be determined. This includes costs and benefits that are notoriously hard to quantify, such as safety improvements and reductions of emissions (Stevens, 2004). Transfers of costs or benefits from one party to another (such as tax revenues) do not increase social welfare and should thus be included (Brand, Parody, Orban and Brown, 2004).
4. The distribution of impacts may be as important as their totals and should not be neglected in the CBA.
5. Costs and benefits in the future have less value than those in the present. All costs and benefits need to be discounted to a chosen basis. To this end the discount rate of the project needs to be determined: this percentage indicates the yearly decrease of value of costs and benefits due to inflation, opportunity costs[45] and risk.

[45] The public resources invested in the project cannot be used for another project. This alternative project would have yielded an increase of social welfare that now cannot be achieved. The fact that one cannot achieve this increase of social welfare is called "opportunity costs".

CBA can be seen as a specific type of MCA (De Brucker et al, 2004). The monetary value attached to each of the costs and benefits can be seen as the weight for each criterion. By adding all costs and benefits, CBA implicitly assumes compensation over the criteria is possible. For example, an increase of travel costs can be compensated by an improvement of air quality and vice versa.

Cost-Effectiveness Analysis
The third method for socio-economic evaluation is Cost-Effectiveness Analysis (CEA). In CEA the effectiveness of the project is expressed in a ratio between the monetary costs of the project and a single, quantifiable indicator for the extent in which the projects achieved its objectives (Newman-Askins et al, 2003). In the case of DTM the Cost-Effectiveness can for example be expressed in the number of reduced number of vehicle hours of delay[46] per invested euro. An example of a CEA is the study "Cost-effectiveness of utilisation measures[47]" (Coëmet, 2003).

Suitability for application to DTM
Even though CEA is a simple method, since it doesn't require the evaluator to quantify all relevant costs and benefits, it has two important drawbacks (Kee, 1994). Firstly, cost effectiveness is not suitable for a project with multiple costs and benefits. The interpretation of a CEA with multiple costs and benefits is highly subjective, unless weights are attached to each benefit to reach a common denominator. Secondly CEA does not indicate whether the investments into DTM measures were justified, whereas a positive benefit-cost ratio in CBA does give this justification.

CBA is more extensive and complete than CEA and may thus solve these disadvantages. (Stevens, 2004) mentions three possible approaches to CBA in DTM projects:

- CBA as a stand-alone evaluation method. Examples of this approach are the CBA framework for ITS projects developed by Moore and Pozdena (Moore and Pozdena, 2004), the Canadian guideline for the evaluation of ITS projects (Bruzon and Mudge, 2007) and the work of Ozbay and Martin (Ozbay and Martin, 2004).
- CBA within a quantitative and qualitative evaluation framework. The Fimnish guideline for DTM evaluation (Finnish Ministry of Transport and Communications, 2002) for example advises the use of CBA as part of an evaluation framework which includes qualitative criteria.
- CBA as an element of a MCA. The difference with the previous approach is that some form of aggregation of the scores per evaluation criterion is made to come to a final judgment about the project. An example is the ADVISORS evaluation framework (Macharis et al, 2004).

[46] In Dutch: "Voertuigverliesuren"

[47] In Dutch: "Kosteneffectiviteit benuttingsmaatregelen"

Stevens (2004) concludes after a study into the applicability of CBA to DTM that, although CBA is a useful and necessary tool in the evaluation of DTM projects, it is insufficient to base decision-making completely on it. DTM projects generate effects which cannot be captured in a CBA, but which may be important to involved stakeholders. Stevens is thus in favour of an approach in which CBA is part of a MCA. Several authors and existing evaluation guidelines (Finnish Ministry of Transport and Communications, 2002; Haynes and Li, 2004; Newman-Askins et al, 2003; Zhang et al, 1998) advise to adopt a similar approach in which CBA is seen as part of a larger socio-economic evaluation framework.

In (De Brucker et al, 2004) it is argued that MCA is more suitable as the central evaluation methodology for DTM projects than CBA. They provide the following arguments for this:

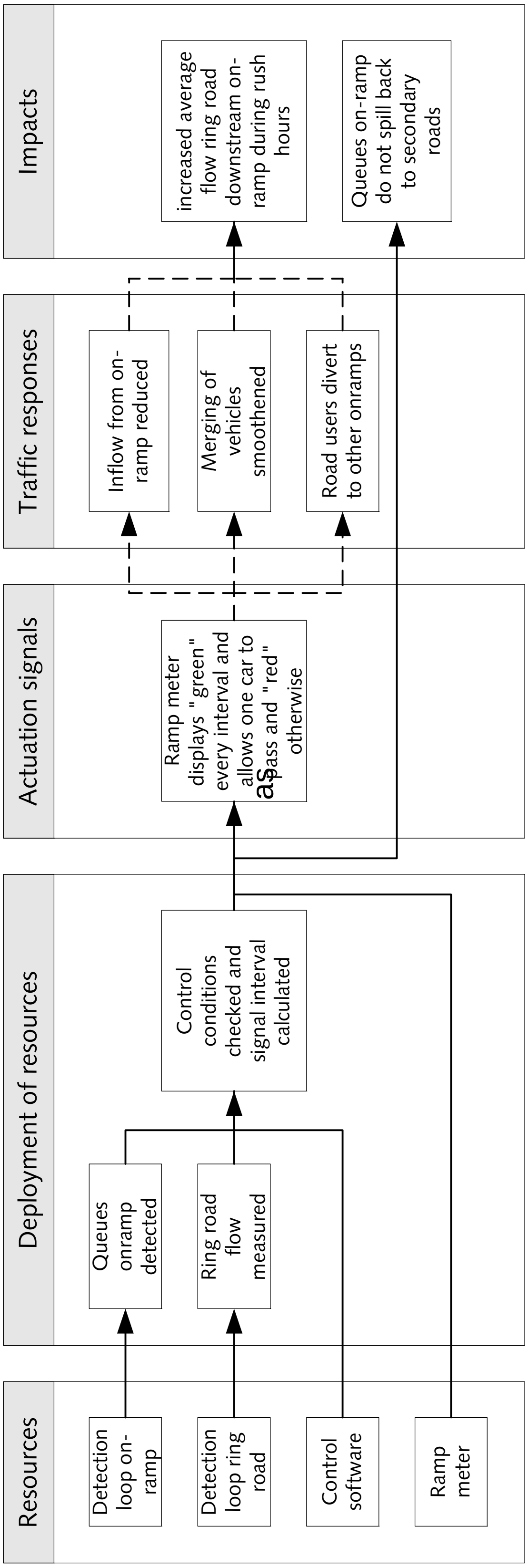
- It is hard to attach a monetary value to many of the effects of DTM, in some cases it is even not possible to quantify the effects.
- CBA implicitly allows compensation on the scores of all criteria without any limitations by aggregating all costs and benefits into a single benefit-cost ratio. A pitfall in CBA is thus the "black box syndrome" (Kee, 1994), in which assumptions and uncertainties underlying the CBA remain hidden in the CBA by the focus on the final benefit-cost ratio. This benefit-cost ratio may give a false sense of certainty about the underlying approach. In MCA the process aggregation of scores is made explicit and is open to debate.
- MCA allows the stakeholders involved to determine their own objectives and criteria, as opposed to CBA (Macharis, Verbeke and De Brucker, 2004).
- The aggregation of costs and benefits into a single benefit-cost ratio in CBA neglects the distribution of effects. This argument can also be found in (Newman-Askins et al, 2003) and (Stevens, 2004). Moore and Pozdena argue however that in CBA it is also possible to show the distribution of effects by drawing up a CBA per area or per stakeholder (Moore and Pozdena, 2004).

A disadvantage of MCA is that it does not provide a final judgment of whether the investments in the DTM project were justified or not. CBA does this by calculating the benefit-cost ratio. (De Brucker et al, 2004) argue however that a postive benefit-cost ratio only indicates that the winners have lost more than the losers have lost, which is not acceptable per se.

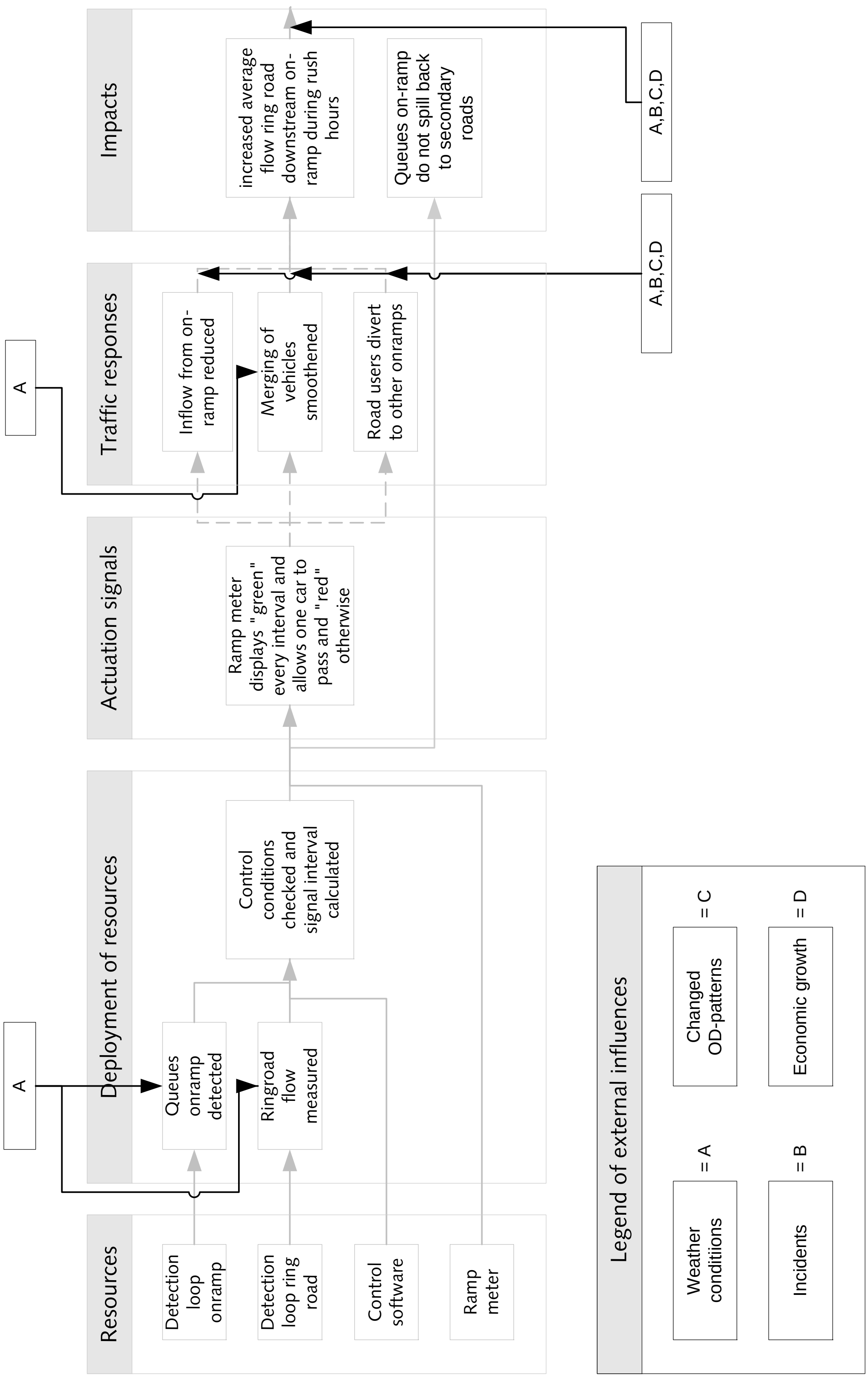
(Giorgi and Tandon, 2002) point out that both MCA and CBA have disadvantages. “One major problem with Cost-Benefit Analysis is its theoretical reliance on market values and by extension on shadow prices; on the other hand, the more flexible design of multi-criteria analysis can increase the likelihood of double counting. Both methods face problems with the specification of weights to apply to different criteria, albeit in different ways: cost-benefit analysis in view of the difficulties involved in measuring reliably the “willingness-to-pay”; multi-criteria analysis in adopting a “subjectivist” approach to this and relying on the decision-maker or a round of experts to determine how important any particular type of good or impact is for social welfare” (Giorgi and Tandon, 2002).

Conclusions about methods for socio economic evaluation
Three methods for socio-economic evaluation were studied, namely Multi-Criteria Analysis (MCA), Cost-Benefit Analysis (CBA) and Cost-Effectiveness Analysis (CEA). Cost-Effectiveness is a straightforward method, which should be preferred in DTM projects with a single objective and low evaluation resources. For more complex projects, involving multiple objectives and/or many actors a combination of CBA within a qualitative evaluation framework should be preferred. A complete CBA is usually not feasible for a DTM project, due to the number of impacts that are hard to quantify and even harder to give a monetary value. The qualitative evaluation framework, possibly an MCA, can be used to evaluate these impacts in a qualitative manner.

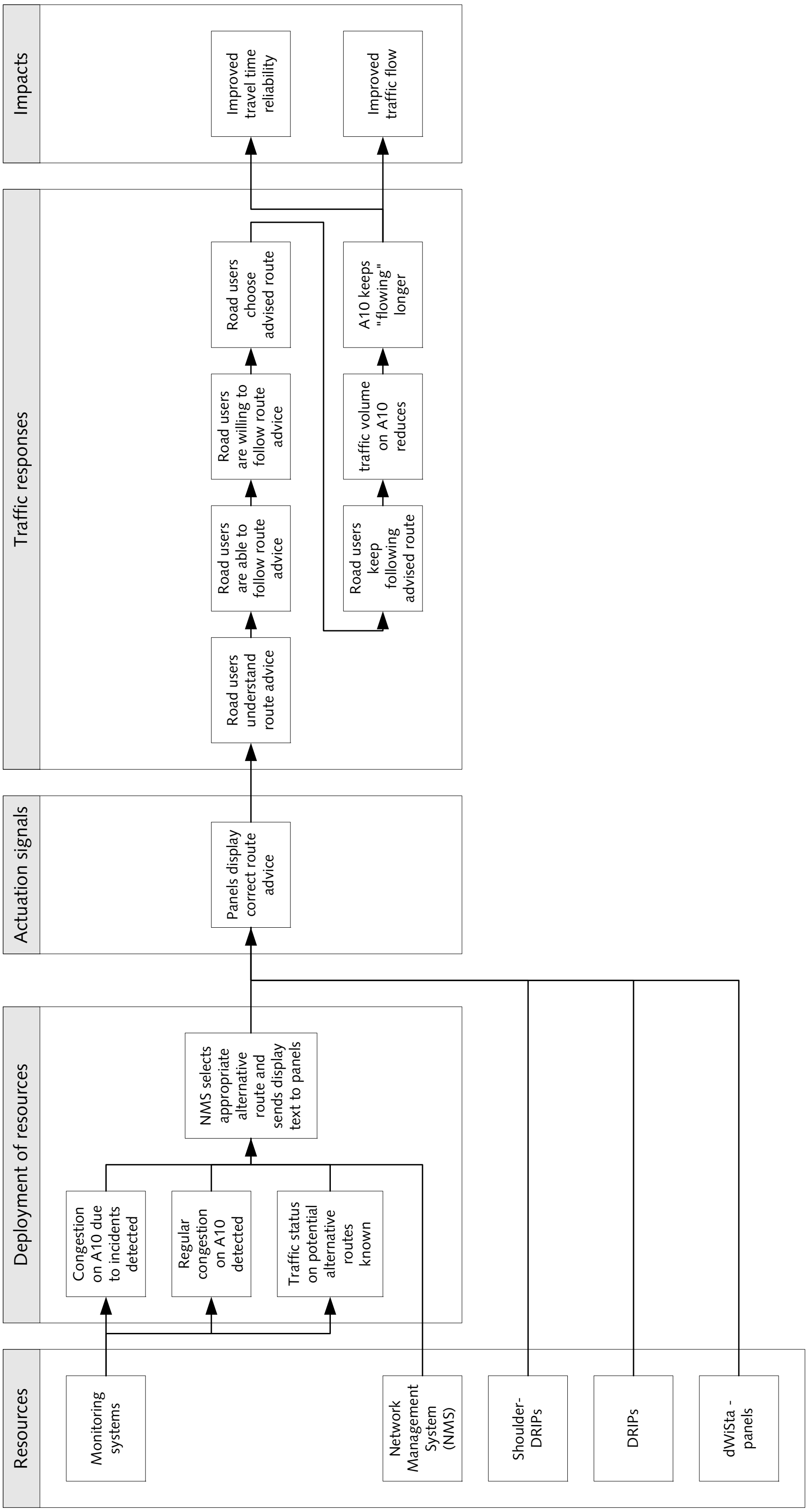
Appendix B : Logic model of example DTM project



Appendix C: Logic model of example DTM project with external influences



Appendix D: Logic model of "Field Test Traffic Management"



Appendix E : Logic model of “Field Test Traffic Management” with external influences