Impact evaluation methods in Civitas for urban freight measures

Do evaluation methods and results for urban freight measures reach the objective of an appropriate impact evaluation?

Bachelor thesis at the | Chair of Integrated Traffic Planning
Fachgebiet Integrierte Verkehrsplanung
Faculty V
Technische Universität Berlin

written by: Arne Töpfer
Matrikelnummer: 321890

supervised by: Dr. K. Dziekan and
Dr. O. Schwedes

July 23rd, 2012

Eidesstattliche Erklärung

Die selbständige und eigenhändige Ausführung versichere ich an Eides statt.

Berlin, den 23.07.2012

Unterschrift

Contents

List of figures

List of tables

List of abbreviations

3PL	Third party logistics
BAU	Business as usual
CCI	Common Core Indicator
CI	Core indicator
CNG	Compressed natural gas
CSR	Corporate social responsibility
FER	Final Evaluation Report
GHG	Greenhouse gas
ICT	Information and communication technology
KPI	Key performance indicator
LC	Logistics center
LDV	Light delivery vehicle
LTZ	Limited Traffic Zone
MAC	Marginal abatement cost
MERT	Measure Evaluation Results Template
MSI	Measure Specific Indicator
PM	Particular matter
SCC	Social cost of carbon
tkm	Ton kilometer
UDC	Urban delivery (distribution) center
UGM	Urban Goods Movement
vkm	Vehicle kilometer
WP	Work package
WPCI	Work package core indicator

Abstract

In recent years, the development of fast growing cities, increasing goods traffic and sustainability objectives has required new concepts for urban logistics. Impact evaluation allows identifying successful concepts and is therefore the basis for future policies and decisions. This thesis analyzes impact evaluation within Civitas, an EU co-financed initiative aiming at implementing sustainable urban transport measures.

An assessment of impact evaluation methods and results of three measures identifies the evaluation of impacts on traffic and congestion as a major weakness. An empirical study shows that standardized common core indicators are unsuited to the purpose and rarely used. As individually used Measure Specific Indicators are inappropriate as well, this thesis introduces alternative approaches. As a conclusion, a set of 10 theme specific indicators for urban logistics measures is developed and given as a recommendation for future impact evaluation.

1 Introduction

1.1 Scope

This thesis analyzes the impact evaluation of Civitas measures concerning urban freight distribution and logistics. A Civitas measure is a project within the frame of Civitas, a European Union co-financed initiative, with the aim of developing sustainable transport in cities.

Since its introduction in 2002, there have been three Civitas phases: Civitas I (2002 – 2006), Civitas II (2005 – 2009) and Civitas Plus (2008 – 2012). Each phase consists of four or five Civitas projects. A project is a pool of up to six cities each hosting several measures. Civitas measures are divided into several topics with urban logistics being one of them. These topics are called work packages or clusters. An overview of Civitas projects, cities and work packages is given in Annexes I & II.

Impact evaluation describes quantifiable effects of a measure. In contrast to process evaluation, which focuses mainly on barriers and drivers before and during the measure's implementation, impact evaluation is based on measurable indicators, usually captured before, during and after the measure's implementation.

1.2 Structure

The framework for Civitas evaluation will be presented first. Chapter three then determines its strengths and weaknesses through a detailed assessment of three measures, one of each Civitas phase. In chapter four systematic issues and trends in impact evaluation of urban freight measures are identified. Here an empirical study of the use of indicators evaluates the compliance between framework documents and measure evaluation carried out in practice. While chapter three is more of a micro-analysis, chapter four is more of a macro-analysis. Chapter five introduces alternative approaches and ideas to overcome the identified weaknesses before conclusions will be drawn and recommendations can be given in chapter six. The recommendations given in chapter six are more detailed than usually expected within a conclusion due to their aim of providing comprehensible suggestions for future improvements.

The terminology used in this thesis refers to the definitions provided in the Civitas guidelines and framework documents. The terms 'urban logistics', 'urban freight' and 'urban goods distribution' are used synonymously. Furthermore, for ease of reading, names of indicators are put in inverted commas unless they occur in a table or otherwise separated from the sentences.

1.3 Characteristics of urban freight measures

Urban freight measures are characterized by the structure of their stakeholders. Unlike other measures, most urban freight measures require cooperation between the project team and several competing private logistics operators. Participating in a Civitas measure often means a change in a company's core business strategy, i.e. changing from direct deliveries to a hub strategy or from achieving economies of scale by using big trucks towards several small vehicles with route optimization. Measures from other work packages do not deal with competing companies. Metro and bus services are usually run by a monopolist. In contrast to the monopolist, freight operators fear shifts in market shares, losing customers due to alternating service quality and a change in their business strategy. This often results in a low number of participating operators, a complete failure or only partly implemented measures. Impact evaluation faces enormous difficulties in measuring evanescently small effects. As long as someone's core business is not affected, acceptance towards a Civitas measure is higher. This can be demonstrated by a freight measure in Toulouse, where shopkeepers related to the Civitas measure "[…] are ready to accept more freight delivery constraints to enable the use of only clean vehicles" (Mobilis, 2008: 24) whereas the Urban Delivery Center failed because carriers feared constraints in their business plans.

Urban logistics contains other characteristics which make the evaluation of these work package measures very challenging. Urban goods transport was not in the focus of politics and decision makers until the mid-1990s when it became more relevant as goods deliveries and pick-ups in city centers were identified as a driving force for urban congestion. Therefore, scientific surveys and data capturing of urban logistics have a relatively young history compared to passenger transport or extra-urban and long-distance logistics. Patier & Routhier consider a programme about urban logistics undertaken by French authorities in 1993 as the first important scientific data capturing to measure urban goods transport and its impacts (Patier & Routhier in: Bonnel et al., 2009). According to them, there is only little experience in data capturing of urban goods transport due to this young history. As the number of high quality surveys and data – the basis for every scientific impact evaluation – is low, work in this area is still rare. Thus, impact evaluation for past and current Civitas projects is challenged by the fact that standardized methods and indicators to measure urban logistics do not exist yet.

1.4 Objectives of urban freight measures

The overall objectives of Civitas urban freight measures are minimizing pollution, noise and congestion. While some evaluation plans and reports indicate three different objective levels (high-level objectives, strategic level objectives and measure specific objectives with concrete measure units), others simply contain high-level objectives. The definition of objectives and expected results are fundamental for impact evaluation.

There is a lot of conformity concerning the overall and high level objectives. Several other projects, support schemes and initiatives as well as the OECD agree with the Civitas overall objectives (OECD, 2012). Unlike the overall objectives, there is a high level of disagreement on measuring impacts in order to evaluate success.

Besides the Civitas Initiative, there are other urban logistics projects and thematic networks dedicated to the development of urban logistics solutions. Many of those projects are published on specific online platforms with the aim of sharing knowledge and experience of successfully implemented urban freight projects. BESTUFS (Best Urban Freight Solutions) is such a thematic network, sharing experience and recommendations about urban freight solutions through an online platform. Among the freight solutions being presented on their platform are also Civitas measures. Another network of public organizations, companies and European cities called START (Short Term Action to Re-organize Transport of Goods) also shares best practice projects. Both networks are European Union co-funded. Unlike Civitas projects, BESTUFS and START have no joint approach to impact evaluation. Their focus lies on the implementation process and the functionality of the different types of freight solutions.

2 Framework for evaluation in Civitas

2.1 Objectives

Evaluation in Civitas is based on global guideline documents named Framework for Evaluation. They are designed to ensure local and cross-site evaluation in order to help monitoring the Civitas Initiative and to formulate policy recommendations. The framework documents recognize the meaning of high-quality local evaluation by putting it at the beginning of a decision making process: local evaluation results are the base for a cross-site evaluation which can be taken to compare measures, determine a measure's effectiveness, to identify best practice and the potential for transferability. Conclusions can then be drawn and finally the European Commission can develop policy recommendations for sustainability in urban transport strategies (Pointer, 2009: 1).

2.2 Framework history

The baseline for Civitas evaluation is a European Community funded strategic initiative called Maestro. In 1999 this initiative developed a common European framework for design and evaluation of transport related pilot projects and demonstration measures including a list of 163 indicators referring to 76 impacts (Guard, 2006: 73-80). When the first framework document for Civitas I was created, only 28 indicators were taken. It was assumed that urban transport projects have a smaller range of impacts, hence less impacts, than transport projects in general. Since then, support initiatives, also named support actions, published new framework documents for each Civitas phase. Meteor, Guard and Pointer are the support actions for Civitas I, Civitas II and Civitas Plus.

The main ideas of Maestro are still part of the most recent framework document for Civitas Plus, but impact descriptions and common indicators were slightly changed. Guard and Pointer framework also contains some advice and lessons learned from its predecessor in order to improve impact evaluation constantly.

The main idea developed by Maestro was to categorize the great variety of possible impacts at several levels. The top level called evaluation area (also evaluation category) is divided into sub-categories, each containing several impacts. Each impact can be measured by one or more indicators. These indicators are called Common Core Indicators (CCI) because they are valid for the entire set of Civitas measures from all work packages. The complete lists of CCIs for each Civitas phase are given in Annexes III – V. Annex VI gives an overview of the modifications between the indicators for Civitas I and those for Civitas Plus: changes of names are symbolized by being crossed out, red symbolizes a new indicator. Two indicators were added between Civitas I and Civitas Plus, a few indicators such as 'quality of service' were renamed so that they can be applied not only to passenger transport and vehicle-based indicators for traffic level were replaced by traffic flow indicators. In general, only very few changes were made.

Impacts and indicators in Civitas evaluation framework documents are not mandatory and subject to flexibility. Each project team and Civitas city team shall only be given a baseline for their own evaluation plan. Evaluation teams are also encouraged to add own indicators if suitable.

2.3 Importance for urban goods transport

The Framework for Evaluation is a global document covering all Civitas work packages (WP). As different work packages are expected to have partly different impacts, likely impacts are listed for each work package. Among the likely impacts for urban freight measures, 14 have matching Common Core Indicators. While several specific impacts are listed for WPs referring to passenger transport, only one freight specific impact, freight movements, is given for urban logistics. This is a first hint that framework documents do not sufficiently provide support for urban logistics measures.

All 14 impacts have been measured for urban logistics. The history of Civitas evaluation shows that, although impacts in the sub-category transport system have the highest number of defined CCIs, measuring this sub-category causes most difficulties.

The analysis of three Civitas measure evaluations will show that evaluation for urban freight measures often uses other indicators than the CCIs, but struggles to find appropriate substitutes. The difficulties in evaluation of urban freight measures have not been taken into account when framework documents were revised.

3 Assessment of three measures

3.1 Civitas I: Stockholm 9.3

The measure *Consolidation of supplies in medieval Old Town / Restaurant Supply* contains the implementation of a logistics center for the consolidation of deliveries for restaurants in the Old Town of Stockholm. At the logistics center, goods are unloaded, consolidated and loaded again. Consolidation means that different goods are sorted based on the destination. The objectives are to reduce the number of small direct deliveries, to reduce congestion levels, to improve environment and living conditions as well as to reduce energy consumption and emissions. Originally, an electric vehicle should have been used for the consolidated deliveries to the restaurants. Concrete figures, i.e. a desired reduction of direct deliveries per day per restaurant from six to one, make the objectives precise and measurable. Due to an incident, the vehicle was destroyed before implementation. The measure was finally implemented with a diesel and later with a biogas vehicle. By the end of the measure, 35 over a total of 85 restaurants in the Old Town participated (Trendsetter, 2005: 36).

All nine indicators chosen for the evaluation according to the project evaluation plan were successfully measured. The method applied is the comparison between the situation before and after the measure's implementation. A BAU scenario as recommended in the Framework for Evaluation is missing. Figures in the following analysis are taken from the Trendsetter evaluation report for WP 9 (Trendsetter, 2005).

Environmental impacts are measured by four Common Core Indicators: CO_2 emissions, NOx emissions, PM emissions and noise levels. While no change in noise levels was achieved, emissions were reduced. CO_2 and NOx emissions decreased by around two percent[1] compared to the situation before. Although PM emissions decreased by 70 percent, no explanation for these huge differences in reductions is given. Differences among emission types are common, but usually do not differ in this range. The report's 'assumptions for calculations' provide additional emission data for the understanding of the results. According to these data, all three emission types are more than 80 percent[2] lower for a biogas vehicle than for a diesel vehicle. The huge differences between CO_2 reductions and PM reductions cannot be the result of the use of biogas. Since the achieved reductions in distance travelled should reduce all emissions similarly, it remains vague why such differences are measured. The evaluation report does not go beyond the general finding of decreased emissions and does not point out the wide range of reduction rates. Energy consumption, given through the CCI fuel mix, decreased as well.

[1] Own calculations based on absolute figures from the evaluation report (Trendsetter, 2005: 32-33).
[2] Own calculations based on absolute figures from the evaluation report (Trendsetter, 2005: 36).

Besides environment, the evaluation focus lies on the evaluation area of transport. Four indicators, including the CCIs 'vehicle kilometer' and 'number of goods vehicles moving in demo area', are calculated. Both indicators show reductions which are in a direct context to the quantitative objectives. As this is often not the case for other measures, a direct relation between an objective and an indicator is worth a mention at this point. In addition to the CCIs, two more transport indicators are measured. These two indicators are work package specific indicators, especially defined for the WP of urban logistics within Civitas Trendsetter. The subject of work package specific indicators will be discussed in chapter five. The indicator 'small deliveries' is also in direct relation to the objective whereas the last indicator, the 'vehicle load factor', proves efficiency of a consolidated delivery trip which is not part of the measure objective. However, proving a better load factor is common for measures aiming at optimizing urban logistics and therefore interesting for further comparisons.

The report states that congestion was reduced, although the CCI for congestion levels had not been calculated. It is not mentioned which specific indicator implies a reduction in congestion levels.

Overall, the impact evaluation of the measure *Restaurant Supply* in Stockholm represents a good example of Civitas impact evaluation. Economic and social impacts are not measured at all, but they are not among the objectives. Clear and precise objectives, measured by matching indicators build the basis for this successful impact evaluation.

3.2 Civitas II: Toulouse 10.1

The measure *Clean urban logistics and goods distribution platform in Toulouse* consists of two parts: the implementation of new freight regulations for the city center and the development of an Urban Delivery Center (UDC). Access control restrictions already existed before the implementation of this measure. The objectives can be summarized as defining, implementing, promoting and coordinating the modified city center access regulations and the UDC in order to "improve the transport and freight delivery in the Toulouse city hyper centre […]" (Mobilis, 2008: 1). Two years after the measure had started, the implementation of the UDC was abandoned. The new objective then was to show the benefits of a logistic platform run by Chronopost, the express delivery service of French postal service (Mobilis, 2008: 1-2).

The defined objectives are imprecise because it is not further explained what exactly is meant by 'improve' and 'optimize'. As the term 'improve' is highly dependent on the point of view, improving a situation for someone may worsen it for someone else. This can be demonstrated by the following example: if a two-lane city road were reduced to a single-lane road by transferring one car lane into a bicycle lane, cyclists might regard this as an improvement whereas it represents a deterioration for a car driver. It is therefore essential to explain what is meant by the term 'improve'. Unlike as in more general top-level documents for a Civitas phase or Civitas project, it is possible to define clear and detailed objectives at this point since it is about a specific measure.

Referring to the evaluation report's introduction, traffic congestion, a poor adaptation of existing access regulations to shopkeepers' needs and violation of existing access restrictions are major issues in Toulouse's freight transport (Mobilis, 2008: 1). Making the link between the issues and objectives allows identifying the potential impacts and most appropriate indicators. This link is clearly missing in the impact evaluation of this Civitas measure. It leads to the situation in which costs, benefits, emissions and acceptance were measured, while no efforts to monitor impacts on the major issues like rules violation and congestion were made.

Out of the five evaluation categories, three categories with the four impacts acceptance, emissions, costs and benefits were chosen for the impact evaluation of this measure.

Acceptance has been monitored in a very detailed way. Distinction was made between different stakeholders such as shopkeepers being affected by the access regulations, shopkeepers who are Chronopost clients and Chronopost deliverymen, so that separate results for the two sub-measures were possible. The survey goes beyond typical questions about satisfaction levels as it also concerns the relations between deliverymen and shopkeepers and the willingness of accepting constraints. The results for the acceptance of Chronopost's new delivery system are clear. One survey is enough, if someone is asked to evaluate a change. The respondent makes the comparison of two different situations. For the new access restrictions, people were only asked about the current situation (Mobilis, 2008: 23). Without any data of the situation before, conclusions about the acceptance cannot be drawn because there is no comparison between two situations. Although it is mentioned in the evaluation report that data were captured prior to the modifications, they cannot be found in the report (Mobilis, 2008: 18-23). However, measuring acceptance by the match between opening hours and unrestricted time slots for access is very useful. It is a way of identifying the ratio between the needs of affected people and necessary constraints to reach an objective. In combination with the precise and in-depth survey, the way acceptance was measured can be regarded as a very good example.

For this measure, economic impacts are calculated in a very detailed way. Many different types of costs were taken into account and an in-depth cost comparison to the baseline data shows the efforts made for this part of evaluation. Most interesting are the calculations of costs per vehicle type as a function of its operating range. The detection of savings in working hours for staff is just one parameter among the numerous parameters used to calculate additional costs as well as monetary savings.

Emissions were calculated by using ADEME's[3] software *Impact 2.0*. The software is designed to calculate emissions from road transport. The method behind the software is not subject of this analysis. Nevertheless, its results from running the software need to be interpreted carefully.

[3] ADEME (Agence de l'environnement et de la maitrise de l'énergie) is a public undertaking for environment protection and energy consumption control.

Focussing on CO_2 emissions due to their highest weight (Mobilis, 2008: 20) as done for this measure is inappropriate and misleading. The importance of a certain type of emission does not depend on its weight. Each emission has different proportions of the total amount of exhaust gas and has a different impact. Due to their different impacts, direct weight-based comparisons are meaningless. Usually, a distinction between greenhouse gas emissions causing climate change and toxic emissions causing severe damage to the ecosystem and humans is made. For GHG emissions, economic and environmental comparisons are common. The weight of GHG emissions can be transferred into CO_2 equivalents (CO_2e) by multiplying with the Global Warming Potential (GWP). One ton $C0_2$e equals a certain impact on climate change, regardless of the gas that was originally emitted. Economic comparisons are similar. Different types of emissions are transferred into the same unit, here into monetary units. A common method, as presented in the Stern Review (Stern, 2007), is calculating the Social Cost of Carbon (SCC) and Marginal Abatement Cost (MAC). While SCC refers to the monetary damage caused by emitting one ton of carbon, the MAC refers to the costs necessary to reduce one ton of emissions. Although SCC and MAC are more relevant for current emissions trading schemes and applied to high volumes, these units represent an excellent way of comparing emissions. Other emissions toxic for ecosystems and humans are usually not compared or transferred as above. The occupational exposure limit is an indicator for its toxicity for the human body. The different exposure limits show that the allowed concentration in workspace air is different for each toxic gas, but comparisons are meaningless.

As the implementation of the Chronopost logistic platform includes the introduction of new vehicles and new routes, the emission reductions were split into their origin. According to the evaluation report, approximately three quarters of the CO_2 emission reductions result from the use of electric and CNG vehicles (Mobilis, 2008: 19). It is an interesting fact that this is also the case for each emission type which decreased. Unfortunately, no conclusion is drawn from this fact. Instead, the finding that emissions can be reduced, especially CO_2 emissions, is underlined in the evaluation report. This fact can be found in emission tables of car manufacturers and environmental agencies. It seems more interesting that the use of clean vehicles contributes significantly more to the savings than the new route design when both are combined in the same experiment. This points out one of the weaknesses of the Civitas impact evaluation. The analysis of captured data is not focused on new findings, but often on repeating already proven facts. Thus, the real value of captured data is not identified and new conclusions are often missing. Impact evaluation stays behind its possibilities.

A very remarkable point of the evaluation of this measure is the fact that traffic congestion in the center of Toulouse is named as the main problem in evaluation reports at measure and project level (Mobilis, 2008; Mobilis, 2009), but traffic congestion had never been measured and none of the Common Core Indicators had been used. It therefore seems questionable to evaluate the goal to renew the freight management regulation and the city center access restriction as fully achieved

(Mobilis, 2008: 24). The implementation of rules itself is part of the process evaluation. Within an impact evaluation, the implementation of rules cannot be considered as a full achievement because there is no data proving it. Determining compliance by the rate of rules violation could be a valid basis for such a statement.

The results from two measurements made by the City of Toulouse in 2004 provide very detailed information for the baseline situation. Vehicle accesses and rules violation were monitored. Although these data are part of the evaluation report, there is neither a BAU scenario nor data of the situation after the implementation of the Civitas measure. It is mentioned that "[…] it would have been difficult to identify the origin of some impacts" (Mobilis, 2008: 16) as strong interrelations to the measures 6.1 *New Parking Management* and 6.2 *Public Space Redesign* are assumed. Thus, no further efforts to measure congestion were undertaken.

Looking at measure 6.2, there are not only interrelations but there is also a goal conflict. While measure 10.1 tries to reduce traffic congestion by removing barriers from the road such as delivery vehicles, measure 6.2 creates quasi-barriers by reducing the amount of available lanes and closing roads for traffic (Mobilis, 2009: 63).

The goal conflict for traffic congestion can be explained as follows: less or smaller delivery vehicles allow other cars to pass the road which leads to an increased average speed. Closing a lane means that the same number of vehicles share the remaining lanes. The traffic density thus escalates assuming no change in the car driver's behavior. According to the Fundamental Diagram of Traffic Flow, the average speed decreases when traffic density increases. As 'average speed' is the Common Core Indicator for congestion, the two measures are contradictory in their desired impact. It is impossible to determine the success of either one of the measures with this indicator. This goal conflict underlines the importance of well-defined indicators when the desired impact is strongly affected by another measure.

The overview of objectives and matching impacts and indicators given in Annex VII points out the poor overall quality of evaluation. Although environmental, social and economic impacts are measured in a detailed way, they are not related to the city's trouble with congestion. Transport related indicators are considered as irrelevant and therefore not given in the report even though the objective to optimize freight delivery requires their application.

3.3 Civitas Plus: Bologna 7.1

The measure titled *City Freight Delivery Plan for Bologna* relates to the access-restricted historical city center where a Limited Traffic Zone (LTZ) already existed. The measure consists of four parts: new access permission procedures and pricing policies to the LTZ; the introduction of low emission vehicles; the decrease of freight vehicle accesses to the LTZ through the Van Sharing System and improved road occupancy & deployment of dedicated loading areas.

According to the Local Evaluation Plan, the measure specific objectives are to

- "Demonstrate the impact that a regulatory action upon freight distribution can have on urban traffic congestion and pollution level […];
- Optimise and develop an effective integration between road pricing policies […] and technological tools;
- Contribute to decrease the number of kilometers travelled to provide the same services;
- Favour the completion of the 'City Freight Delivery Plan' and disseminate the new opportunities provided" (Mimosa, 2009: 65).

All measure specific objectives contribute to the strategic objective of energy efficient freight distribution and the high level objective of improving air quality. By 'technological tools' an IT tool calculating optimized delivery routes is meant. Together with the Van Sharing System, a virtual transit point shall be created. Such a virtual transit point has no physical transit points like a central warehouse but the network of existing physical infrastructure at each participating operator is seen as a virtual transit point.

It was originally planned to monitor emissions, operating costs and freight movements by a total of six indicators (Mimosa, 2009: 67). The measure evaluation report adds three more indicators for operating costs and freight movements. Comparing the objectives with the selected indicators, the following matches can be identified:

Desired objective	**Impact**	**Indicator**
Demonstrate impacts on congestion level	-	-
Demonstrate impacts on pollution	Emissions	Emissions of CO, CO_2, NOx, PM
Develop integration between road pricing and IT tools	Freight movements in LTZ	Vehicle accesses to LTZ, distance travelled, delivery time
	Operating cost	Operating cost
Decrease distance travelled	Freight movements in LTZ	Number of deliveries for vehicle, number of freight vehicle accesses

Table 3.1: Indicators for Measure BOL7.1 (Source: own representation)

Table 3.1 shows that four out of five desired impacts were measured. Like in Toulouse, congestion is named as an issue for the city, but neither the Evaluation Plan nor the Measure Evaluation Results Template (MERT) lists any indicators to measure the impacts on urban congestion.

Freight movements were measured with two different goals: to demonstrate the positive effects of combining road pricing with ICT and to demonstrate the efficiency of using shared resources through the Van Sharing System. The effects of combining road pricing with ICT were demonstrated by a software simulation which calculated the total distance and time travelled by delivery vehicles while executing 26 delivery orders in different ways. The simulation compares a situation of three companies performing independently to a situation in which vehicles from the

Van Sharing System are used. The Van Sharing System is also called Consortium. The results from two simulations with different parameters such as capacity constraints and number of delivery points are given in the MERT. While one simulation was successfully tested in reality, the second simulation has not been approved yet. The evaluation team draws the conclusion that a system structured as the Van Sharing System combined with IT tools would make deliveries more efficient by significantly reducing LTZ vehicle accesses, delivery times and total distance travelled. This conclusion is fully justified by the simulation results. It shows that the integration of road pricing policies, here the pay-to-access strategy of the LTZ, can be efficiently combined with IT tools (Mimosa, 2011: 4-5). The value of this result even goes beyond the measure objectives, because the successful simulation is also a key result for the Smartfreight Project, a European Union co-financed project aiming at integrating ICT solutions into urban freight management. However, it is important to mention that this does not imply a real reduction of freight movements since these routes were not applied to regular business.

In order to verify the benefits of the Van Sharing System, real performance data were used. The number of total freight vehicle accesses was compared to the number of accesses of the Van Sharing System vehicles. This comparison does not measure a change in vehicle accesses, but gives the ratio of accesses generated by the Consortium to all other freight vehicle accesses. The 'number of deliveries per vehicle', another indicator used for freight movements, is even smaller for the Consortium vehicles than for all other freight vehicles. No specific conclusion is drawn from this fact. It can be assumed that the statement "Only with an increasing participation and thus with more data, it will be possible to verify the benefits shown by the two simulations […]" (Mimosa, 2011: 14) relates to the poor results for the Consortium vehicles. One can also draw a complete different conclusion: as long as the Van Sharing System does not reach a critical size, it degrades the business as usual situation due to its poor performance. A critical size can be defined as the amount of participating companies in the Consortium necessary to reach a load factor equal to the vehicles of the so called own account delivery operators. As many Civitas measures showed, it is very difficult to convince freight operators to join an organization in which resources are shared. Talking about a critical size is therefore necessary to convince potential participants.

For the measurement of operating costs, data from the Consortium were taken. A comparison with a pre-Civitas situation or BAU scenario was not made because the transport companies were not willing to publish these data (Mimosa, 2011: 13). It can be assumed that the data are part of their critical business performance data. Companies do not want them being available to competitors and potential market entrants. This represents a common problem of the evaluation teams. Companies opposed to give away similar data in earlier cases of Civitas evaluation. New ways of data collecting will be needed. However, the evaluation team in Bologna decided to show the Consortium's economic development by calculating operating costs every year. The lack of pre-Civitas data has no relevance to this method. This seems to be the most appropriate solution without the BAU scenario data, but it measures something else than the defined objective. Even if the Consortium vehicles run with decreasing operating costs, it will not be possible to see if the Van Sharing System is financially more attractive than own account operators or 3PL providers.

Several units, also called sub-indicators, were defined to monitor operating costs in Bologna: costs per delivery, costs per trip, costs per packet, costs per kg, costs per cubic meter and costs per day (Mimosa, 2011: 13). Using these seven sub-indicators provides a high data quantity but a poor data quality. What does the sub-indicator 'costs per trip' imply? Raising costs per trip could be explained by a longer delivery route with more stops. Declining trip costs could be explained by an optimized route design. Both cases are positive developments but push the indicator into opposite directions. This sub-indicator allows no meaningful conclusions. As the goal is a reduction in operating costs for the transported goods, an appropriate sub-indicator needs to be directly related to the good. The customer pays for the transport of the good but neither for the distance travelled for a delivery trip nor for the number of delivery stops per trip. Making the delivery trip or one km distance travelled less expensive is not the goal. This could be achieved by shortening to one delivery stop per trip, taking a smaller vehicle or reducing the payload as an almost empty vehicle is less fuel consuming. According to the Civitas Plus Common Core Indicators, the unit of operating costs indicators is € / vkm. The meaning of this unit and the units mentioned above are questionable for the evaluation of urban freight measures. In the context of freight transport, a common unit is ton kilometer (tkm). Measuring operating costs related to tkm avoids ambivalent conclusions. The unit € / tkm is strongly focused on the objective of reducing operating costs for freight transport relative to transport volume.

This example shows another critical point in the impact evaluation: choosing an indicator and a unit that measures the impact in such a way that it is strongly related to the defined objective. In the case above, the chosen measurement units do measure the defined impact, but they are not related to the objective of demonstrating a positive economic effect. A positive economic effect can only be observed by decreasing average operating costs or marginal costs or other costs relative to business activity. The indicator unit needs to be independent from the parameters whose effects shall be demonstrated. For the indicator '€ / trip', a doubling of the trip length would double the indicator. If the number of trips were halved by this doubling, operating costs would not be changed since only half of the number of trips would be necessary. The amount of indicators does not compensate the poor value.

For this measure, the Copert software was used to calculate emissions. This software was designed to calculate emissions from road transport. As this is a standardized and well-accepted method of emission calculations, the method itself is not subject of this analysis. The parameters and the meaning of the results will be discussed instead.

The only parameter which varies over time is the total distance. Other relevant parameters, mostly engine related parameters, were taken from the LTZ access control system.

In 2010, the Consortium generated less than 0.015 percent of the daily deliveries inside the LTZ. This figure is defined as the penetration rate. Given that the reduced travel distance of the Consortium's vans is the only parameter changing the emissions, is it highly noticeable that CO_2 emissions were reduced by 0.48 percent during this year[4]. The reduction rate is more than 30 times higher than the penetration rate. Such results are possible if the replaced delivery trips were the highest contributors to CO_2 emissions but it seems very unlikely that such a small number of optimized trips can cause such a huge impact. There is no hint about a sanity-check comparing the indicators and putting the results into a context. In case of full plausibility, results should be further explained. Such findings need to be highlighted and analyzed.

For the evaluation of transport related impacts, definitions of several freight specific terms are given. Unfortunately, they are not clear and understandable. Referring to the definition, the terms packet, good and delivery all describe the same. Referring to the numeral results for the indicators, they describe something different. The lack of coherence between given definitions and their applications to the indicators even leads to wrong figures. The analysis of the figures also shows a strong dependence between indicators for economy and for transport system. According to the units, indicator 9A has to be the ratio of indicator 8.3 (number of deliveries in LTZ, daily average) to indicator 8.2 (number of vehicles entering LTZ, daily average). According to the numbers, indicator 9A is the ratio of indicator 8.3 to 8.1 (number of accesses in the LTZ) (Mimosa, 2011: 12).

$$\text{unit} - \text{wise:} \quad \frac{[\text{Ind}.\,8.3]}{[\text{Ind}.\,8.1]} = \frac{\text{deliveries}}{\text{accesses}} \neq \frac{\text{deliveries}}{\text{vehicles}} = \frac{[\text{Ind}.\,8.3]}{[\text{Ind}.\,8.2]} = [\text{Ind}.\,9\text{A}]$$

$$\text{numeral:} \quad \frac{\text{Ind}.\,8.3}{\text{Ind}.\,8.1} = \frac{20231}{1782} = 11.35 = \text{ Ind 9A}$$

This demonstrates that clear definitions as well as plausibility checks are required in order to avoid wrong figures.

In addition to the wrong figures, the following dependences between the indicators exist:

$$\text{Ind}.\,7 = \frac{\text{Ind}.\,5.1}{\text{Ind}.\,5.3}\,; \qquad \text{Ind}.\,8.3 = \frac{\text{Ind}.\,5.7}{\text{Ind}.\,5.1}\,; \qquad \text{Ind}.\,9\text{A} = \frac{\text{Ind}.\,5.2}{\text{Ind}.\,5.1}\,;$$

Dependence among sub-indicators is quite normal, but strong dependence between different indicators conflicts with the independence criteria for indicators (Pointer, 2009).

Summing up, impact evaluation for this measure makes great effort in monitoring the transport evaluation area, but suffers systematic weaknesses. Inappropriate and dependent indicators with conflicting definitions are the major weaknesses.

[4] Both figures are not part of the evaluation report. They are own calculations based on the MERT (Mimosa, 2011: 13-14).

3.4 Incoherence among evaluation levels

Measure results are the base of all high-level evaluation reports. Major findings and results shall be reflected in project and work package reports. However, a certain discrepancy can be observed.

The Mobilis FER summarizes for the work package ‘New concepts for the distribution of goods’ that “Most goals have been achieved or will be after the end of Mobilis project.” (Mobilis, 2009: 252). According to the MERT, only one out of four objectives of the freight measure in Toulouse was achieved and according to the Mobilis FER no results for the Venice freight measure were achieved. As the work package consists only of the two measures, it is obvious that most goals have not been achieved. The above statement must therefore refer to future achievements and is consequently a prediction. The statement is misleading and brightens the lack of a proof about the achievement of goals.

Another incoherence between different evaluation reports concerns the *Clean Urban Logistics* measure in Venice. While the success of implementation and outcome is rated as very successful in the Cluster Report (Guard, 2010 f: 4), no results for this measure were achieved according to the FER (Mobilis, 2009: 252).

Despite its positive rating of the Venice measure, the Cluster Report summarizes that “[…] cluster 4, Logistics and Goods Distribution related measures, is characterized by low success of the implementation process and low success of the outcome.” (Guard, 2010 f: 4). The detailed assessment of the three measures leads to the same result. The low outcome is also a result of poor impact evaluation as it does not succeed in visualizing and pointing out the most relevant measure results.

4 Investigation of systematic issues

4.1 A study on the use of indicators

4.1.1 Introduction

The choice of the right indicator is a key element for successful impact evaluation. Some impacts have standardized indicators, which are commonly used and well-established, e.g. carbon-dioxide emission indicators. Whenever GHG emissions were measured for Civitas, this indicator was used. Other impacts do not have standardized indicators. Although the Civitas support actions developed a list with Common Core Indicators, there are certain impacts, mainly transport system impacts, which were mostly measured by different non-Common Core Indicators. Some impacts are more difficult to monitor than others.

Within the frame of this bachelor thesis the following study on the use of indicators has been made.

4.1.2 Method of approach

This study identifies relevant characteristics of urban freight measures concerning the use of indicators. Data from evaluation reports of 106 Civitas I & II measures were analyzed for this study. Data were taken from intermediate evaluation levels, thus from work package, city and project evaluation reports (Guard, 2010 c - j; Miracles, 2006; Tellus, 2005 a, b; Tellus, 2006; Trendsetter, 2005; Vivaldi, 2005). This intermediate level assures comparability among the measures for the data aggregation. In-depth measure evaluation reports also often contain indicators for all evaluation categories even though no results were achieved (i.e. either ex-ante data or ex-post data were unavailable). Measures without any results, often caused by non-implementation, were excluded, i.e. they do not count for the base value. Attempts to monitor an indicator have not been counted. The categories energy and environment are not considered separately due to the structures of the evaluation reports. Fuel savings and emissions reductions were often directly related and presented as one single category. As impact evaluation for Civitas Plus is currently in process and it is not yet clear if the indicators chosen will be successfully measured, current measures are not part of the analysis.

4.1.3 Findings

The results of this study are as follows:

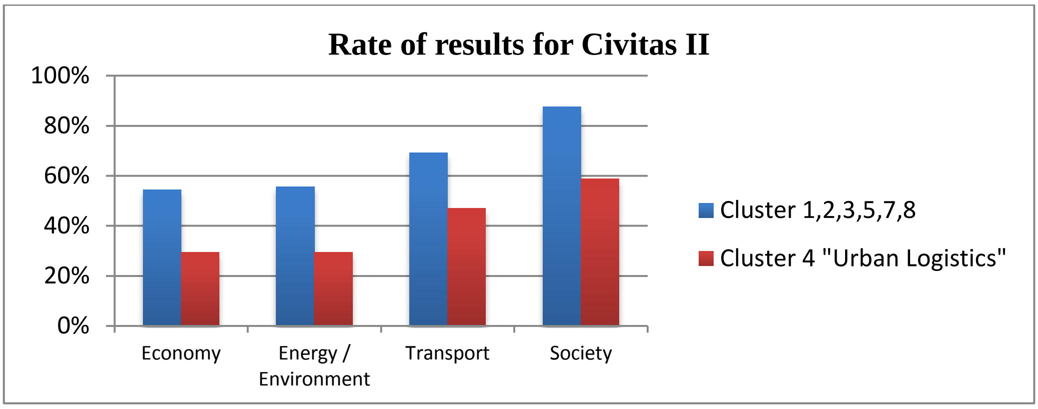

Figure 4.1: Rate of results per evaluation category (Source: own representation)

Figure 4.1 shows the rate of indicator-based results for Civitas II measures. Both CCI-based and non-CCI-based results are counted here. 97 Civitas II measures from all clusters except cluster six are taken into account. Cluster six ‘Traffic Management and Control’ could not be integrated into the figures due to its high specificity and the fact that its measures were only monitored in one evaluation category. For Cluster seven ‘Public Transport’ only measures about the public transport network could be taken into account due to the same reasons. Most significant in figure 4.1 are the low rates of results of urban logistics measures (Cluster 4) compared to the other clusters. Assuming that Cluster four measures had the same resources available for evaluation, the low rates demonstrate that measuring urban freight measures is more difficult than non-logistics measures. Different rates among the categories depend on where the evaluation focus had been set, but they do not cause the difference between freight and non-freight measures. The differentiation by cluster is usually slightly different from the differentiation by work package. In the case of urban freight measures, cluster four corresponds to WP 10 of Civitas II.

The study now focuses on the evaluation category of transport, in particular on the sub-category of transport system, since many difficulties in measuring transport related impacts have been identified through the assessment of measures in the previous chapter. The transport evaluation category is also of high relevance because the most important measure objectives refer to this category.

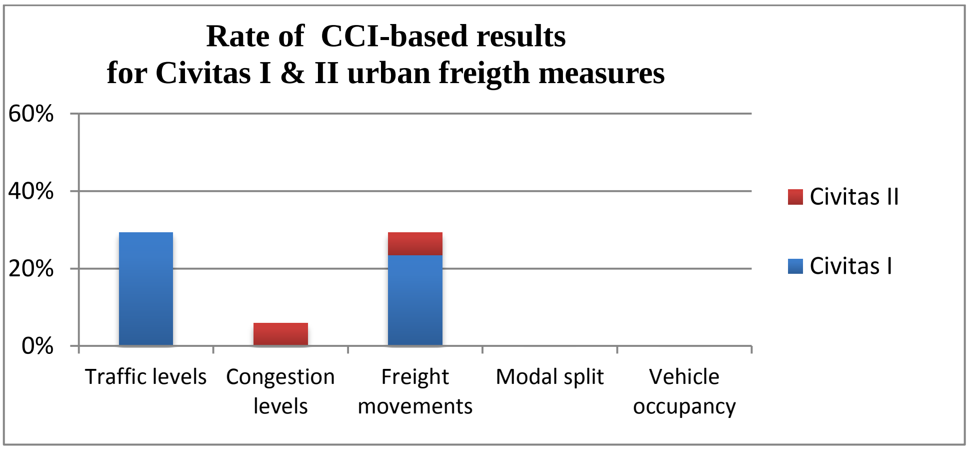

Figure 4.2: Rate of CCI-based results (Source: own representation)

Figure 4.2 shows the percentage of freight measures for which an impact was measured by the matching CCI. Changes of the CCIs between Meteor and Guard Framework for Evaluation are taken into account. A 'match' counts if the names and descriptions of indicators are identical; the indicator unit may be different. Out of 26 urban freight measures (nine for Civitas I, 17 for Civitas II), 65% measured transport related impacts. Figure 4.2 takes into account only the 65%, thus the 17 measures with results for the transport category. Thus, the base value for numbers in figure 4.2 equals 17. Furthermore, only the sub-category of transport system is shown because the other sub-categories security and service level are of no relevance to this WP. Security and service level are neither among the objectives nor among the results achieved. One can see that two out of five transport system related impacts were not measured by a CCI. 29% of the measures used the CCI to determine traffic levels. Also freight movements were measured by 29 %, congestion levels by 6%, thus one of 17.

It is highly remarkable that only very few evaluation teams used the CCIs, while almost all evaluation reports mention congestion level and traffic level as major issues in urban traffic. Worth mentioning are also the infinitesimally small rates for Civitas II measures compared to Civitas I measures. One single Civitas II measure used the CCIs (both for congestion levels and freight movements). The distinction between the contribution of Civitas I and II shows a development away from the CCIs. A significant discrepancy between the idea of Common Core Indicators and its actual application to measure evaluation can be observed. The Common Core Indicators for the transport system sub-category are obviously inappropriate for urban freight measures.

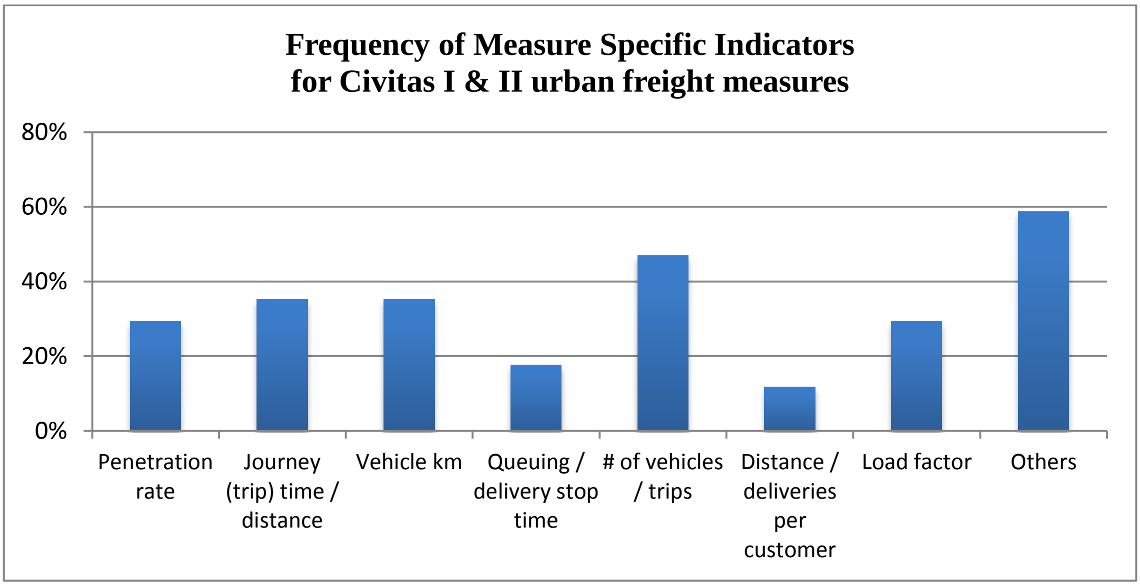

Figure 4.3: Frequency of Measure Specific Indicators (Source: own representation)

Instead of using the CCIs, almost all freight measures used one or several other indicators. The indicators most frequently used are presented in figure 4.3. The base value for figure 4.3 is the same as for figure 4.2: the 17 measures which determined results, CCI-based or non-CCI-based, for the transport category. Indicators which occurred once are summarized as 'others'. According to the measure reports, the indicators in figure 4.3 monitor congestion level and freight movements. One can see that individual indicators, hereafter called Measure Specific Indicators (MSI), in contrast to the CCIs, were used significantly more often. There are several combinations between an indicator and a matching impact in the evaluation reports. The high level of experimentation is proof of a lack of common understanding concerning traffic congestion and freight movements on the one hand, but also proof of the willingness to monitor impacts on the other hand.

As seen above, the CCIs have almost no meaning for measuring transport impacts. Some evaluation reports state that expected effects would be too small to be visible through the CCIs because too many external factors influence the results. The reports also mention the enormous effort it would take to monitor indicators which refer to the whole demonstration area. Measuring the average vehicle speed (CCI #23/24 for congestion level) seems to be too complex for some evaluation teams. Some reports doubt the CCIs' capability to identify the origin of a change in average speed due to interrelations to other measures. These shared concerns result in a trend towards indicators with a smaller observation unit, i.e. the freight vehicle itself or the parcel, instead of all vehicles moving in the demonstration area. This trend is an approach to reduce complexity and to use indicators which seem to be more tangible to local teams. Except for very small demonstration sites such as a construction site, there has been hardly any attempt to monitor hundreds of private and commercial vehicles moving in the demonstration area. This means that CCIs are more likely to be measured if the geographical area is small enough, thus capturing data seems less complex.

Among the indicators in figure 4.3, ‘trip time & distance’, ‘vehicle kilometers’ [5] and ‘number of vehicles or trips’ are popular. These indicators describe the characteristics of route design and business strategies, but they have low meaning for traffic congestion. If indicators showed a doubling in trip distances and a cut in the amount of vehicles, the total distance travelled could still be unchanged. Safe conclusions for the congestion level cannot be drawn. The penetration rate is the rate of trips or deliveries performed by Civitas-related vehicles, e.g. vehicles belonging to a van sharing system or delivery center, to total deliveries. This rate provides only valuable information about the measure’s acceptance by economic operators. Another indicator from the figure above is the ‘load rate’. It measures the occupancy of a vehicle: actual payload over maximal payload in terms of weight or volume. The queuing & delivery stop time per delivery measures the time for a certain delivery stop. It refers to the loading or unloading time as well as to the accessibility to the final delivery place. The distance between the parking spot and the goods handover is mainly relevant here. The queuing time refers to occupied and blocked delivery areas. The less time is needed, the sooner the delivery vehicle drives away and avoids blocking and hindering other cars. A direct effect to congestion can be expected without measuring congestion directly. In contrast to the others, this indicator allows drawing valid conclusions, thus being a suitable indicator.

Except for the ‘delivery stop time’ indicator, none of the indicators above allows drawing conclusions when considered separately. Especially the indicator ‘number of trips’ and the indicator ‘distance per trip’ have often been serving for positive conclusions. Such conclusions have little evidence. Only the combination of several indicators justifies safe conclusions.

The general lack of clarity between indicator and matching impact as well as the diffuse understanding of impacts are underlined by the statement “[r]educed congestion, based on fewer vehicle movements” (Trendsetter, 2005: 61). According to the framework documents, congestion and freight movements are both different impacts not directly related to each other. For the measure Göteborg 9.5, congestion level was measured by the ‘traffic flow’ indicator, which – according to the framework – measures traffic level instead of congestion.

Among the indicators summarized as ‘others’ in table 4.3, many refer to the economic activities of participating stakeholders, e.g. ‘total number of deliveries’, ‘number of deliveries per vehicle’ and ‘number of reusable transport devices’. The link to traffic characteristics is clearly missing, as these are more logistics performance indicators being interesting for delivery operators. Performance indicators are important for participating delivery operators and potentially new operators which still need to be convinced. For an impact evaluation, these indicators are rarely relevant.

This study shows that there are two general trends: the indicators measured refer to small observation units and tend to be business performance indicators. Chapter five will come back to

[5] The indicator ‘vehicle km’ used to be a CCI in Civitas I. Since Civitas II, it is no more a CCI. It was taken into account as a CCI for figure 4.2. However, it is listed again in figure 4.3 due to its relevance for several measures.

the idea of small observation units and demonstrate its meaning for alternative approaches measuring traffic impacts. The second trend, using performance indicators instead of impact indicators, is a negative trend. A failure in monitoring impacts cannot be compensated by establishing performance indicators. The risk of drawing wrong conclusions increases if only performance indicators are captured. Considering a pure increase in business activities as successful can be wrong if aspects like load factor, critical size and performance efficiency are not taken into account.

4.1.4 Criteria for Indicators

Based on eight criteria the Common Core Indicators were defined. The following breakdown of the transport system indicators shows that for evaluation of urban freight measures, the defined indicators do not meet these criteria. It has to be underlined at this point that this breakdown does not question the choice of CCIs for Civitas measures in general. The following breakdown only regards urban freight measures.

Each criterion will be discussed concerning the Common Core Indicators and the Measure Specific Indicators used by measure evaluation teams. For a better understanding, definitions according to Meteor, Guard and Pointer (Pointer, 2009: 22-23) are given for each criterion.

- Relevance: "each indicator should represent an assessment criterion, i.e. have a significant importance for the evaluation process";
  - CCIs: generally relevant to urban freight;
  - MSIs: partly resemble performance indicators; more relevant to freight operators, but not to impacts on urban traffic;
- Completeness: "the set of indicators should consider all aspects of the system / concept under evaluation";
  - CCIs: lack of efficiency indicators e.g. load factor, penetration rate;
  - MSIs: usually do not cover congestion;
- Availability: "readily available for entry into the monitoring system";
  - CCIs: low availability for traffic and congestion levels due to wide application areas
  - MSIs: medium availability for data of participating operators; low availability for comparative data i.e. data from non-participating companies and pre-Civitas data;
- Measurability: "the identified indicators should be capable of being measured objectively or subjectively";
  - CCIs: objectively measurable, but only with great time and effort due to the wide geographical area;
  - MSIs: only measurable by operators thus dependency on operators;
- Reliability: "clarity of definition and ease of aggregation";
  - CCIs: clear definitions are given;
  - MSIs: lack of clear and common definitions;

- Familiarity: “the indicators should be easy to understand”;
  - CCIs: numerous different combinations between impacts, indicators and conclusions are proof of no common understanding;
  - MSIs: mainly easy to understand;
- Non-redundancy: “indicators should not measure the same aspect of an assessment criterion”;
  - CCIs: non-redundant;
  - MSIs: mainly non-redundant;
- Independence: “small changes in the measurements of an indicator should not impact preferences assigned to other indicators of the evaluation model”;
  - CCIs: independent;
  - MSIs: not always independent (see section 3.3).

Additionally, indicators shall also have a temporal scope, a geographic scope, importance and permanence. As there are no significant differences between CCIs and non-CCIs, these characteristics are not considered any further. Neither CCIs nor the Measure Specific Indicators meet all the criteria. The above breakdown will be summarized in table 5.1 in the following chapter after alternative indicators will have been introduced.

“The indicators that need to be used will depend on the set of measures being implemented. The wide ranging nature of the measures will make some indicators irrelevant for certain cities, whereas in other cases some indicators may be hard to measure.“ (Pointer, 2009: 24). The study above shows that the transport system CCIs are irrelevant and hard to measure for most cities.

4.2 Definitions of freight specific terms

Often terms like access to a restricted area, trip, packet and delivery are not clearly defined. Sometimes several terms are used to describe the same activity, sometimes different activities are described by the same terms. As shown in the assessment of the measure evaluation for the *City Freight Delivery Plan for Bologna*, the lack of clear definitions for terms being used leads to misunderstandings in the counting methods. The analysis of other Civitas evaluation reports proved that there is a general lack of definitions for terms describing delivery system’s characteristics. As the terms ‘delivery’ and ‘delivery trip’ can be used to describe the same process while ‘delivery’ can describe two different processes, clear definitions are necessary. A delivery trip can be composed of several deliveries. Otherwise, understanding the description of different delivery route options becomes impossible. Private enterprises face the same difficulty when logistic operations need to be coordinated between different sites. Therefore, many enterprises include clear definitions in their logistic processes.

4.3 Up-scaling & transferability

Up-scaling is part of the impact evaluation. According to the framework documents, up-scaling means the “[…] estimation of the effects of a measure (or group of measures) if it / they were applied fully throughout the city." (Guard, 2006: 32). Empirical assessment with extrapolation is a recommended method. However, the common understanding of up-scaling, according to evaluation reports, is different. Instead of extrapolating effects, many reports describe whether or not an application to the whole city seems possible by listing potential barriers. The Guard Cluster Report for logistics is a good example of this common understanding of up-scaling (Guard, 2010 f: 19-20). Effects and quantitative measure results are generally not up-scaled. This understanding of up-scaling equals an estimation of transferability within the city.

The term transferability is used in many evaluation reports to describe a measure's potential application in another city. In the sense of evaluation reports, transferability is an up-scaling to other cities. However, both up-scaling as well as transferability estimations have not been used to estimate quantified results and therefore should not be part of impact evaluation. There is no value for impact evaluation if up-scaling refers to barriers and drivers. Both should clearly be separated, while up-scaling belongs to impact evaluation, transferability estimations should be part of process evaluation. The lack of clear distinction between these two terms becomes evident with the following statement which is the concluding sentence of the work package evaluation report: "All of these projects in WP 9 have a transferability possibility. They are performed in typically European cities and areas and for common purposes." (Trendsetter, 2005: 60). The statement refers to city characteristics, i.e. potential barriers or drivers, instead of referring to empirical assessment of impact results.

4.4 Dependency on private companies

Data for calculating indicators often come from external sources and cannot be measured by evaluation teams. Transport authorities, public authorities or NGOs normally provide data. As many urban freight measures are characterized by cooperation between Civitas and private freight operators, their data are necessary for impact evaluation. Usually, there are data generated during the measure's lifetime by participating operators. In order to put these data into a context and to draw conclusions, data from the situation before as well as a BAU scenario are necessary. As these data concern their key business, many companies are not willing to share confidential information and key data. Whether these are certain operating costs, load factors or route characteristics, evaluation teams struggle to have comparative data.

While focusing too much on indicators depending on private companies' performance and business, the risk of impact evaluation failure increases. Nevertheless, collecting performance data even if comparative data is unavailable might be useful in order to convince potential participants who are attracted by favorable figures. Either way, distinguishing between collecting data for impact evaluation and collecting data in order to convince other freight operators is necessary.

4.5 Coherence with support documents

Besides the Civitas framework document, several other mid- and top-level documents have been created. Their aim is to support and advise local Civitas teams as well as decision makers at a national and European level.

Policy Advice Notes documents summarize key findings and experiences for a work package made during a project phase. The Advice Notes for urban logistics for Civitas II (Guard, 2010 a) list impacts and benefits separately for the public, for freight companies and for consignees. A discrepancy between the measure results and the Advice Notes can be observed. The Advice Notes state reduced congestion due to a reduced number of heavy freight vehicles and a higher average vehicle speed as a key impact. Since only one of 17 Civitas freight measures monitored congestion (as defined in framework documents), there is very little basis for this statement. Cost savings and better loading rates are also listed among the observed impacts. Reading the Advice Notes, it seems that most measures proved this. In fact, loading rates were rarely compared to BAU loading rates. Captured data only proved a positive development of loading rates during the projects' lifetime, but data from the situation before were hardly available. The Policy Notes list some indicators for the transport category, but without any explanation and definition. Concerning environmental, social and other economic impacts, the Advice Notes are in line with local measure reports.

Neither any negative findings nor the difficulties in measuring transport system impacts, nor the low result rates are mentioned in this document. Four cities, including Toulouse, are even missing in the list of urban freight measures within Civitas II.

These Policy Advice Notes create a positive image of urban freight measures that does not reflect reality. A critical review about impacts and results is clearly missing.

Another support document is the Final Overview of Evaluation (Guard, 2010 b). It lists key findings, conclusions and recommendations for each work package. Again, the problems in impact evaluation are not mentioned. Conclusions and recommendations for the urban freight WP (called cluster in this document) focus on process evaluation. Two of the listed conclusions refer to economic and environmental impacts. The document admits that urban freight measures were less successful and assumes that city authorities are responsible for the failure because they did not meet the specific requirements (Guard, 2010 b: 32-33). A critical view on impact evaluation is again missing.

The two examples of support documents do not provide any valuable support to local teams for impact evaluation.

5 Alternative approaches

5.1 Introduction to research on urban logistics

Although scientific research on urban logistics has a young history of less than 20 years, there are several different approaches to measuring urban logistics. Patier & Routhier compared different methods aimed at measuring urban logistics. Their study showed that there are very few examples of well-executed surveys because numerous surveys failed to measure the important characteristics of urban logistics. For Patier & Routhier the success of such a survey is determined by its focus on "[…] the most important question about goods movement […]: road occupancy by the goods vehicles that are in competition with individual cars." (Patier & Routhier in: Bonnel et al., 2009: 269). They consider the term road occupancy to combine congestion and accessibility. As congestion is a key element of characterizing the quality of urban goods movement, their study is of high relevance for the impact evaluation for urban goods measures in Civitas. Many measure evaluation reports as well as mid- and top-level reports define congestion as a major issue. Although Patier & Routhier's study analyzed surveys to measure urban logistics in general instead of changes in urban logistics, both the mentioned surveys and the Civitas measures aim at quantifying congestion.

This chapter presents a best practice in measuring congestion but also general ideas for the improvement of impact evaluation.

5.2 Indicators of congestion level & system efficiency

Based on their definition of a successful method in measuring urban logistics, Patier & Routhier identified a survey carried out in Marseille, Bordeaux and Dijon named UGM Survey (Urban Goods Movement) as "[…] the most focused on capturing the entire scope of UGM and to the assessment of the relationships between the logistic behavior of the freight movement generators and the transport system." (Patier & Routhier in: Bonnel et al., 2009: 280). As the result of a research program launched by the French Ministry of Transport and the French Public Agency for Environment and Energy in 1993, the UGM Survey is not only one of the first surveys undertaken in urban logistics, but (in 2009) still considered to be the best.

The UGM Survey is an establishment-driver survey. Both, driver and establishment are observation units in this case. An establishment survey is carried out at the establishment such as a depot, stock or production facility containing data about the goods' origin and destination while the driver (or vehicle) survey is done in the vehicle, gathering information about the route, stops, needed time and distance travelled. The driver survey can be carried out by the driver himself / herself or by an interviewer accompanying the driver. The face-to-face survey by an interviewer assures the best data quality and is therefore most beneficial (Patier et al., 2004: 24-25). Based on the establishment-driver survey, it became possible to measure tour characteristics describing the efficiency of the tour itself (e.g. loading / unloading intensity, average load factor, average weight per kilometer) and more importantly those characterizing road occupancy e.g. total loading / unloading time, average unloading time, cause of delivery barriers such as parked vehicles or bicycle lanes, distance travelled per loading / unloading procedure. These tour characteristics are hereafter denoted as UGM Indicators. As the delivery vehicle itself occupies the road and therefore is a source of high congestion levels, its road occupancy given by the time spent parking or running on inner-city roads needs to be reduced.

List of UGM Indicators (Patier & Routhier in: Bonnel et al., 2009: 281-282)

- Number of loading / unloading (ratio, per time and employee)
- Loading / unloading density
- Loading / unloading intensity per activity
- Loading / unloading time
- Distance covered for loading / unloading
- Average length of the first leg from platform to delivery area
- Average distance travelled per collection / delivery
- Total distance travelled in urban area, per truck size
- Average time taken per delivery
- Average speed for round
- Average weight per kilometer (load factor)
- GHG / pollution emissions per kilometer

The indicator ‘average loading / unloading time’ is important for all measures aiming at optimizing the loading and unloading procedure through reduced delivery barriers, e.g. the implementation of delivery bays. The indicator ‘average distance travelled per delivery’ is highly relevant to all measures concerning freight bundling and consolidation, route optimization and implementing delivery platforms. Both are thus applicable to all urban freight measures concerning road transport. Besides these two indicators, the indicators for efficiency mentioned above can be used to measure the delivery system’s performance. They provide an excellent alternative to the performance indicators used by many local evaluation teams. The difference is that all UGM Indicators except the last one are relative figures, while the Measure Specific Indicators were mostly given in absolute figures. Absolute figures reflect the volume of business activity to a greater extent, while relative figures reflect efficiency. Although the meaning of so-called performance indicators for impact evaluation was questioned in chapter three, the UGM Indicators for efficiency are superior to Measure Specific Indicators, because they are relative figures and refer to impacts on congestion. In contrast to the Measure Specific Indicators, the UGM Indicators are given per relevant business activity unit. In this case, a relevant unit is the transport of the goods, that is the service or product which is sold. An irrelevant business activity unit is the vehicle itself. Therefore, the ‘distance travelled per vehicle’ is irrelevant, while the ‘distance travelled per delivery’ is highly relevant since it reflects the product. The objective can only be an increased efficiency for the product. Unlike the others, the indicator ‘average speed per round’ does refer to the round / delivery trip. In this case, it is not a performance indicator showing the delivery operator's efficiency, but an indicator referring to congestion levels. As the CCI for congestion was not used, this UGM indicator can replace the CCI and be regarded as a substitute which requires less time and effort. Concerns about a lack of comparability among cities turned out to be unfounded since similar results for the UGM Indicators were achieved. Relative figures are therefore appropriate for further evaluation. Not all UGM indicators are relevant for impact evaluation. Some of them are only relevant for the determination of a city's general logistics characteristics, e.g. 'loading intensity per activity'. Nevertheless, the majority is of high relevance for impact evaluation in Civitas.

Instead of choosing the freight vehicle as an observation unit, the road itself can also be an observation unit. Video surveillance of the road combined with digital image recognition can identify parked delivery vehicles, thus their parking time can be determined. It is possible to distinguish between legal and illegal double parking. Video surveillance can also be used to capture violations of access restrictions. The implementation of modern IT tools, video surveillance, and optical sensors offers an enormous amount of new possibilities. Although they will not be discussed in this thesis, mentioning these technologies aims at pointing out that indicators for road occupancy can also be captured by the use of technology. Both, human resources for traditional surveys and high tech resources can be used to monitor road occupancy.

The following decision matrix compares Common Core Indicators (only transport system sub-category), Measure Specific Indicators and UGM indicators by referring to the framework criteria which served to choose the CCIs. If a criterion is fully met, one point is awarded. A criterion partly met equals 0.5 point and no points are given if the criterion is not or rarely met. Taking into account the problems identified in past Civitas evaluations, two more criteria are added, coordination and insensitivity. Coordination refers to contrary impacts of other measures in the same demonstration area and insensitivity to disturbing external factors.

	Common Core Indicators	**Measure Specific Indicators**	**UGM Indicators**
Relevance	1	-	1
Completeness	0,5	-	0,5
Availability	0	0,5	0,5
Measurability	0,5	0,5	0,5
Reliability	1	0	1
Familiarity	0	1	1
Non-redundancy	1	1	1
Independence	1	0,5	1
sub-total for Pointer criteria	5	3,5	6,5
Coordination	0	1	1
Insensitivity	0	1	1
sub-total for own criteria	0	2	2
total	**5**	**5,5**	**8,5**

Table 5.1 Comparison of indicators (Source: own representation)

The UGM indicators rank first place for the Pointer criteria as well as for the total of criteria (see table 5.1). Measure Specific Indicators reach very few points for the Pointer criteria, but their strengths are coordination and insensitivity. The MSIs' strengths are the weaknesses of the CCIs.

In conclusion, Measure Specific Indicators are more adapted to urban freight requirements, but suffer from a lack of relevance and completeness. CCIs have a lack of familiarity and availability. Their ineligibility is shown by the low rates of results. The UGM indicators combine the advantages of CCIs and MSIs, thus representing a real alternative to the current way of monitoring urban logistics measures.

5.3 Scope and data availability

As the analysis has shown so far, the availability of information is one of the problems of impact evaluation. It occurred that no data of the situation before were measured or that data existed but were not available to the evaluation teams. A comparison between the situation before and after the measure's implementation became impossible. It also happened that pre-measure data were available, but no post-measure data were captured which made comparisons impossible as well. An interesting solution to the issue of data availability comes from COST, a European support program on Cooperation in Science and Technology. The COST action 350, titled *Integrated assessment of environmental impact of traffic and transport infrastructure*, suggests working with indicators depending on the availability of information (Goger et al. in: Calderón et al., 2009: 198). In case of low availability, a different indicator is used than in case of intermediate or high availability. The higher the availability, the more detailed is the indicator. The indicator of toxic emissions is the amount of emissions in case of low data availability. For intermediate availability, the indicator then refers to the risk of affecting numerous people and for a high data availability, the indicator refers to the number of people affected. The idea behind this example is that a certain amount of emissions does not always mean the same harm or negative effect when the scope is taken into consideration.

Copying this idea to the Civitas evaluation could lead to modified evaluation plans: indicators could be defined based on their data availability. If the availability level is unknown when the evaluation plans are created, up to three different indicators can be defined. In case a low availability is achieved, the least detailed indicator can still provide valuable results. In order to struggle difficulties in measuring urban congestion, another aspect is even more important for urban freight measures: the meaning of the scope. As shown above, measuring congestion levels is one of the biggest challenges. Defining a scope can be a solution here. Measurements become less difficult with a decreasing scope. The scope could be just one street instead of a whole area. One could also refer to another observation unit, which can be regarded as the continuation of narrowing the scope.

The project *TNT Express in Brussels - City Logistics Mobile Depot* is one of the Straighsol projects, a European Union-funded project focused on urban freight solutions. It aims at reducing emissions and congestion by implementing an inner-city depot for parcels. Trucks deliver parcels to this depot and electrically supported tricycles run the last mile. It is not attempted to measure congestion itself, but the project's "[…] impact on urban congestion […]" (Straighsol, 2012). The indicator chosen is 'truck kilometers'. Although measuring a reduction in vehicle kilometers is no new approach, here this indicator is used to measure an activity's impact on congestion instead of congestion itself. It seems to be only a slight difference, but especially since a lack of accuracy concerning impact and matching indicator has been identified, such a slight difference is highly important. Thus, this project can be regarded as an example of the use of indicators based on data availability.

5.4 Indices & relative indicators

Indicators given in relative figures called indices and KPIs represent another way of monitoring impacts. Indices and KPIs usually refer to a quantifiable unit of business activity and are more independent of changes in the level of activity. Such indices can be found in Corporate Social Responsibility (CSR) reports. The SCR report of Deutsche Post DHL contains CO_2 efficiency indices for its divisions. The reference figure for the Mail division is tkm, the one for the Supply Chain division is revenue (Deutsche Post DHL, 2012). The advantage of those indicators is their independence of business activity fluctuations.

5.5 The Trendsetter approach

Trendsetter developed a system of indicators at three levels: Trendsetter common indicators at the project level, WP common indicators at work package level and individual indicators at measure level (Trendsetter, 2004). The Trendsetter FEP assumes that "[…] many core indicators are possible to evaluate but are not relevant since the expected changes are insignificant." (Trendsetter, 2004: 14); an assumption which turns out to be true for the transport CCIs as shown in the study in section 4.1. The Trendsetter indicators, which were harmonized and coordinated with the Civitas I CCIs, were also divided into evaluation areas. Some Trendsetter indicators were simply copied from the CCIs. A new evaluation area called mobility was created including four Trendsetter common indicators: number of trips per mode, travel time per mode, quality of service per mode and acceptance per mode. According to the FEP they are all applicable to freight vehicles, thus to urban freight measures. However, only the first one, 'number of trips', became part of the WP common indicators. The rest of the Trendsetter common indicators for the areas energy and environment are identical with the CCIs. A list of 14 WP common indicators for urban freight measures (WP 9) was created as well.

Evaluation area	Work package CI	Matching CCI	Matching Trendsetter CI
Energy	Energy use	4	Energy use
Environment	CO_2 emissions	8	CO_2 emissions
Environment	NOx emissions	10	NOx emissions
Environment	PM emissions	11	PM emissions
Environment	Noise levels	12	Noise levels
Mobility	Number of trips	25	Number of trips
Society	Living conditions	-	-
Society	Working environment	-	-
Transport	Vehicle km	21/22	-
Transport	Vehicle load factor	-	-
Transport	Queuing time / stop time	-	-
Transport	Small deliveries	-	-
Transport	Vehicle fleet	-	-
Transport	Total distance	21/22	-

Table 5.2: Trendsetter WP 9 common indicators (Source: Trendsetter, with own modifications)

Table 5.2 shows the WP 9 common indicators with matching CCIs and Trendsetter CIs. There is a perfect match for energy, environment and mobility, but almost no match for the area of transport. This is in line with the findings in section 4.1. While the CCIs for some evaluation areas, especially energy, are suitable, CCIs are inappropriate to measure transport related impacts of urban freight measures. The transport related indicators are of such a high specificity that they are not part of the Trendsetter CIs.

The assessment of the three measures within WP 9 shows that each WPCI was used at least once, and that no indicators but the WPCIs were used. It was proved that a system of WP specific indicators works. However, the indicators themselves are only partly appropriate for monitoring impacts of urban logistics measures. As some WPCIs are identical to ‘individual indicators’, the same weaknesses occur. The total distance given in km/trip is meaningless since no clear conclusion referring to the objectives can be drawn.

The Trendsetter approach represents a best practice in terms of creating WP specific indicators and avoiding individual MSIs but not in terms of the indicators themselves.

6 Conclusion

The in-depth assessment of three Civitas urban freight measures as well as the study about more than 100 Civitas measures have identified strengths and weaknesses in Civitas' impact evaluation. There are huge differences among the evaluation areas. The impact evaluation of environmental and social impacts, in particular emissions and acceptance, is one of the very few strengths. Clear results and the use of Common Core Indicators characterize these areas. The major weakness is the poor quality of the evaluation area of transport. Ambiguous figures, a low rate of CCI-based results and ineffective experimentation in measuring traffic and congestions levels are some of its characteristic features.

Evaluating the subcategory of transport system is most important for impact evaluation since reducing traffic and congestion levels turned out to be the most important objective for urban freight measures. However, measuring traffic and congestions levels caused the most difficulties and represents the biggest challenge for future impact evaluation. Besides, an appropriate way of monitoring efficiency and improvements in freight deliveries, which is the third major objective, has not been found yet within Civitas evaluation.

Evaluation methods and results of urban freight measures have not yet reached the objective of an appropriate impact evaluation due to the failure in measuring the most significant impacts. Apart from issues specific to the evaluation area, several global issues were also identified. These include vague objectives, missing links between objective and indicator chosen, a lack of sanity checks and the repetition of well-known findings and facts already proven. Since the beginning of Civitas in 2002, there has hardly been any progress as proven by the micro-analysis in chapter three. Overall, impact evaluation for urban freight measures within the Civitas Initiative has not been successful.

In order to improve the overall quality of evaluation, precise and well-defined objectives which allow choosing more suitable indicators are required. Sanity checks are also recommended since they often point out new findings besides expected and generic findings. Conclusions in evaluation reports need to refer to the figures achieved and have to be drawn more carefully. The biggest improvement of impact evaluation for urban freight measures can be obtained in the evaluation area of transport. Since Common Core Indicators are inappropriate for the evaluation of congestion levels, traffic levels and efficiency of freight deliveries, and since individual approaches failed, common work package specific indicators can be regarded as a solution. Work package specific indicators were already used in Civitas Trendsetter, thus do not represent a totally new approach. Concerning the choice of indicators, some UGM Indicators combined with some individually used indicators are identified as most suitable to monitor impacts of urban freight measures.

The introduction of alternative indicators for the evaluation category of transport implies a differentiation based on Civitas work packages, thus a development towards theme specific indicators. While comparisons among measures of different work packages would still be possible in terms of economic, energetic, environmental, and social impacts, theme specific core indicators would raise the comparability inside the work package because no more individual indicators would be used. The objective of enabling comparability among measures through standardized indicators, i.e. CCIs, is currently not achieved. Replacing all transport CCIs by theme specific core indicators therefore does not conflict with the idea of comparability but would allow comparing results for future urban freight measures. Identifying best practice for fighting congestion, which is currently unattainable, will then be possible. The following table shows the urban logistics specific indicators called 'Work Package Core Indicators for urban logistics' which can be recommended for future impact evaluation. Indicators #2, #3, #4, #6 and #8 are modified UGM Indicators. These modifications only concern the terminology in order to have a common terminology for all indicators proposed.

#	Impact	Indicator	Description / Comment	Unit
1	Congestion: road occupancy & accessibility	Total distance travelled in demo area, per truck size	Monitors shifts from deliveries performed by big trucks towards those performed by small vans /LDVs (or inverse)	$\frac{km}{a}$
2		Average distance travelled per delivery / pick-up	Most important indicator for congestion as well as for delivery efficiency	km
3		Average speed for delivery tour	Ratio of total distance per delivery tour to total time per delivery tour; excluding delivery stop time	$\frac{km}{h}$
4		Traffic lane blocking time	Time of double parking per day; in truck equivalents	h
5	Freight delivery efficiency	Load factor	Ratio of actual pay load to max. pay load	%
6		Average time taken per delivery	Ratio of total time per delivery tour to number of deliveries per delivery tour	h
7		Average delivery stop time	Average parking time; includes both double parking and legal delivery bay parking	h
8		Average distance between LC and delivery point	Evaluates the location of logistics center	km
9	Success of implementa-tion	Penetration rate	Ratio of participants to potential participants; ratio of deliveries performed by participants to all deliveries	%
10		Violation rate	Number of violations per day	$\frac{1}{d}$

Table 6.1: Work package Core Indicators for urban logistics (Source: own representation)

In contrast to the CCIs which measure congestion levels, these WP core indicators only measure contributions to congestion. This turned out to be more appropriate for the limited resources available to Civitas evaluation. All indicators therefore refer only to freight vehicles instead of all vehicles including private cars. However, indicator #3 approximates the CCI for congestion as it also measures average speed. In order to obtain data of the baseline scenario, a sufficiently large number of freight vehicles should be taken into account while post-measure implementation data should focus on freight vehicles which are part of the measure. Unlike many individual indicators analyzed in previous chapters, the indicators recommended here refer to the product – the goods delivered – instead of the delivery tour. The traffic lane blocking time indicator recognizes the freight vehicle as one of the major obstacles for traffic flow. As small vehicles represent less disturbance to traffic, the indicator is measured in freight vehicle equivalents: a small pickup or light delivery vehicle (LDV) equals 0.5 equivalents, a mid-size delivery van equals one and a big truck equals two equivalents. This allows evaluating the impact of measures aiming at shifting deliveries from truck-based to LDV-based. Those measures usually show poorer results for indicators measuring any type of distance since more vehicles and delivery tours become necessary and total distance travelled increases.

Although it is difficult to draw the line between congestion indicators and efficiency indicators since impacts overlap, indicators are divided on the basis of their direct or indirect impact on congestion. In order to avoid uncertainty within local evaluation teams and to ensure a common understanding of the indicators, definitions of logistics specific terms are given in Annex VIII.

Altogether, the set of WP core indicators covers all the important aspects of all major types of measures within this work package. For measures concerning the implementation of logistics centers or any other type of UDC, the first nine indicators are very important. For route optimizing measures indicators #1, #2, #3, #4 and #6 are highly relevant. The success of access and delivery restrictions can be measured by using indicator #10. For the fourth type of measures, the optimization of vehicle access to the point of delivery and improved parking during a delivery stop, indicator #4 and #7 are essential. Since many Civitas measures combine two or more of these types, it can be assumed that many future measures will use the entire set of indicators. Regarding the data capturing methods, all necessary data can be gathered through establishment-driver surveys or road segment monitoring. The establishment-driver survey can use data from IT tools such as route planning software, but this also requires accompanying the driver on his or her delivery tour. Although this type of survey is quite labor intensive for evaluation teams, it assures high-quality data and avoids dependency on internal business data of participants. Road segment monitoring, mainly applicable to indicators #4, #7 and #10, can be simplified through video surveillance in combination with IT tools avoiding on-site manual counting.

A more critical reflection of difficulties in past evaluation processes as well as weaknesses of current methods in support documents and advice notes can also contribute to increasing the quality of impact evaluation. The set of indicators recommended combines the ability of UGM indicators to measure impacts on traffic and congestion, the increased comparability through standardized work package specific indicators of the Trendsetter project and the independency from business activity fluctuations of KPIs and UGM Indicators. It also takes into account the limited data availability by reflecting figuratively the idea of smaller indicator scopes. In conclusion, the 'Work Package Core Indicators for urban logistics' unite the advantages of the alternative approaches introduced in chapter five, while they meet all requirements and indicator criteria as defined by Civitas support actions. The deployment of the recommendations developed in this thesis, in particular the deployment of standardized Work Package Core Indicators, can result in a successful and appropriate impact evaluation of urban freight measures within the Civitas Initiative.

German Summary / Deutsche Zusammenfassung

Die Bachelorarbeit mit dem Titel 'Wirksamkeitsbewertung urbaner Logistikprojekte der Civitas Initiative' beschäftigt sich mit den Bewertungsmethoden und -ergebnissen von Civitas-Maßnahmen. Civitas ist eine durch die Europäische Union geförderte Initiative europäischer Städte zur Umsetzung von Nachhaltigkeitsprojekten im Stadtverkehr. Logistikprojekte stellen dabei einen der Themenschwerpunkte dar. Die Wirksamkeitsbewertung (*impact evaluation*), das Erfassen und Auswerten von quantitativen Auswirkungen einer Maßnahme ist dabei ein zentraler Aspekt der Civitas-Initiative. Sie dient neben der Überprüfung des Erfolges einer Maßnahme vor allem zum Identifizieren der erfolgreichsten Maßnahmen als Entscheidungsgrundlage für nachfolgende Projektphasen. Mittels einer Analyse dreier Ergebnisberichte von Maßnahmen aus den bisherigen drei Civitas-Phasen zeigt die vorliegende Arbeit Stärken und Schwächen der *impact evaluation* auf. Es zeigt sich, dass die Bewertung von ökologischen und gesellschaftlichen Auswirkungen zu deren Stärken gehört, während die Bewertung von Auswirkungen auf den Verkehr, besonders auf die Verkehrsbelastung und den Verkehrsfluss, bisher wenig erfolgreich ist. Da das primäre Ziel der meisten Civitas-Logistikmaßnahmen jedoch die Reduzierung der Verkehrsbelastung durch Liefer- und Warenverkehr ist, ist die Bewertung der verkehrlichen Auswirkungen essentiell für eine erfolgreiche Wirksamkeitsbewertung. Die Analyseergebnisse dieser Arbeit sehen in der Verwendung unpassender Indikatoren die Hauptursache der mangelhaften Evaluation verkehrlicher Auswirkungen.

Eine im Rahmen dieser Bachelorarbeit durchgeführte Studie zeigt außerdem, dass die in Civitas-Richtlinien vorgegebenen, standardisierten Indikatoren nur äußerst selten angewendet werden. Gleichzeitig ist eine häufige Verwendung individueller Indikatoren zu beobachten. Die Arbeit kommt zu dem Ergebnis, dass weder die standardisierten noch die individuellen Indikatoren geeignet sind. Die vorgegebenen Indikatoren sind zu allgemein, die individuellen Indikatoren nicht zielführend. Die Titelfrage, ob Bewertungsmethoden und -ergebnisse das Ziel einer geeigneten Wirksamkeitsbewertung erreichen, wird deshalb verneint.

Die Suche nach alternativen Methoden und Indikatoren identifiziert Indikatoren einer französischen Studie über urbane Logistik als möglichen Lösungsweg für eine verbesserte *impact evaluation*, ergibt aber auch, dass bisher wenig Forschungsergebnisse und Erfahrungen in diesem Bereich vorliegen.

Die Arbeit gibt im Schlussteil detaillierte Verbesserungsvorschläge für zukünftige Wirksamkeitsbewertungen. Die Einführung von standardisierten, themenspezifischen Indikatoren für urbane Logistikmaßnahmen, darunter auch einige der in Frankreich entwickelten Indikatoren, wird empfohlen. Das bisher verfehlte Ziel einer geeigneten Wirksamkeitsbewertung für urbane Logistikprojekte kann, so die abschließende Annahme dieser Bachelorarbeit, durch die Umsetzung der genannten Verbesserungsvorschläge für zukünftige Maßnahmen erreicht werden.

References

- **Deutsche Post DHL, 2012:** Bericht zur Unternehmensverantwortung 2011: Wir transportieren Verantwortung. Bonn. Deutsche Post DHL. PDF document available at http://www.dp-dhl.com/de/verantwortung/umweltschutz.html; last visit 27/6/12.
- **Goger et al. in: Calderón et al., 2009:** Integrated assessment of environmental impact of traffic and transport infrastructure. Calderón, E. J., Pronello, C., Goger, T. (Editors). Madrid: Universidad Politécnica de Madrid.
- **Guard, 2006:** Framework for Evaluation. Deliverable 2.1, version 5.2; internal document. Civitas Guard.
- **Guard, 2010 a:** Policy Advice Notes 05: Logistics and freight distribution. Civitas Guard. Available at the Civitas online resource center.*
- **Guard, 2010 b:** Final Overview of Evaluation. Deliverable D2.2, final version. Civitas Guard. Available at the Civitas online resource center.*
- **Guard, 2010 c:** Cluster Report 1: Alternative car use. Deliverable D2.2, final version. Civitas Guard. Available at the Civitas online resource center.*
- **Guard, 2010 d:** Cluster Report 2: Clean vehicles and fuels. Deliverable D2.2, final version. Civitas Guard. Available at the Civitas online resource center.*
- **Guard, 2010 e:** Cluster Report 3: Cycling and walking. Deliverable D2.2, final version. Civitas Guard. Available at the Civitas online resource center.*
- **Guard, 2010 f:** Cluster Report 4: Logistics and goods distribution. Deliverable D2.2, final version. Civitas Guard. Available at the Civitas online resource center.*
- **Guard, 2010 g:** Cluster Report 5: Mobility management. Deliverable D2.2, final version. Civitas Guard. Available at the Civitas online resource center.*
- **Guard, 2010 h:** Cluster Report 6: Traffic management and control. Deliverable D2.2, final version. Civitas Guard. Available at the Civitas online resource center.*
- **Guard, 2010 i:** Cluster Report 7: Public transport. Deliverable D2.2, final version. Civitas Guard. Available at the Civitas online resource center.*
- **Guard, 2010 j:** Cluster Report 8: Access and parking management. Deliverable D2.2, final version. Civitas Guard. Available at the Civitas online resource center.*
- **Mimosa, 2009:** 9.1 Local Evaluation Plan Bologna. Version 6; internal document. Civitas Mimosa.
- **Mimosa, 2011:** Measure Evaluation Results Template Bologna 7.1: City Freight Delivery Plan. Version of July 27th 2011; internal draft. Civitas Mimosa.
- **Miracles, 2006:** Report on evaluation Results. Deliverable D4.2, version 6.0. Civitas Miracles. Available at the Civitas online resource center.*
- **Mobilis, 2008:** Measure Evaluation Results Template TLS10.1: Clean Urban Logistics and goods distribution platform in Toulouse. Internal document. Civitas Mobilis.
- **Mobilis, 2009:** Evaluation Report. Project Final Evaluation Report. Deliverable 3.2, version1-1; internal document. Civitas Mobilis.

- **OECD, 2012:** Directorate for Science, Technology and Industry: Urban Freight Logistics. Paris. Organization for Economic Co-operation and Development. Online resource: http://www.oecd.org; last visit 11/6/12.
- **Patier & Routhier in: Bonnel et al., 2009:** Bonnel, P., Lee-Gosselin, M., Zmud, J., Madre, J.-L. (Editors): Transport Survey Methods: Keeping up with a changing world. Bingley (UK): Emerald Group.
- **Patier et al., 2004:** Transports des Marchandises en Ville: quelles spécificités, quelles méthodes. Méthodologies des enquêtes tournées dans les agglomérations urbaines. Rapport Final. Patier, D., Cholez, C., Routhier, J-L., Ambrosini, C. Laboratoire d'Economie des Transports, Université Lumière Lyon 2, Ecole Nationale des Travaux Public de L'Etat. PDF document available at http://hal.archives-ouvertes.fr/; last visit 29/6/12.
- **Pointer, 2009:** Framework for Evaluation in Pointer. Deliverable 2.3.1, final version; internal document. Civitas Pointer.
- **Stern, 2007:** The economics of climate change: The Stern Review. Cambridge. Cambridge University Press.
- **Straighsol, 2012:** Project Demonstrations: TNT Express in Brussels. Online resource: http://www.straightsol.eu/demonstrations.htm; last visit 11/6/12.
- **Tellus, 2005 a:** Final Evaluation Report Berlin. Version of November 2005. Civitas Tellus. Available at the Civitas online resource center.*
- **Tellus, 2005 b:** Final Evaluation Report Göteborg. Version of November 2005. Civitas Tellus. Available at the Civitas online resource center.*
- **Tellus, 2006:** Final Evaluation Report Rotterdam. Civitas Tellus. Available at the Civitas online resource center.*
- **Trendsetter, 2004:** Trendsetter Final Evaluation Plan. External Deliverable D4.1, version of February 2004. Civitas Trendsetter. Available at the Civitas online resource center.*
- **Trendsetter, 2005:** Evaluation Report - New concepts for the distribution of goods (WP9). External deliverable 4.3e, version of November 2005; internal document. Civitas Trendsetter.
- **Vivaldi, 2005**: Evaluation Results Report Vol. 1: Final Evaluation Report. Deliverable 9, version 1. Civitas Vivaldi. Available at the Civitas online resource center.*

* The Civitas online resource center is available at http://www.civitas.eu/index.php?id=32; last visit 16/7/12.

Annex I: Projects & cities in Civitas I, II and Plus

Civitas I 2002 - 2006		**Civitas II 2005 - 2009**		**Civitas Plus 2008 - 2012**	
Project	City	Project	City	Project	City
MIRACLES	Barcelona	CARAVEL	Genoa	ARCHIMEDES	Aalborg
	Cork		Burgos		Donostia San Sebastian
	Winchester		Krakow		Brighton & Hove
	Rome		Stuttgart		Iasi
TELLUS	Rotterdam	MOBILIS	Toulouse		Monza
	Berlin		Debrecen		Usti
	Göteborg		Ljubljana	ELAN	Brno
	Gdynia		Venice		Gent
	Bukarest		Odense		Ljubljana
VIVALDI	Nantes	SMILE	Malmö		Porto
	Bristol		Norwich		Zagreb
	Bremen		Potenza	MIMOSA	Bologna
	Kaunas		Suceava		Funchal
	Alborg		Tallinn		Gdansk
TRENDSETTER	Lille	SUCCESS	La Rochelle		Tallinn
	Prague		Preston		Utrecht
	Graz		Ploiesti	MODERN	Brescia
	Stockholm				Coimbra
	Pecs				Craiova
					Vitoria-Gasteiz
				RENAISSANCE	Bath & N.E.Somerset
					Gorna Oryahovitsa
					Perugia
					Skopje
					Szczecinek

Source: Civitas Pointer

Annex II: Measure themes & work packages in Civitas I, II and Plus

CIVITAS I	**CIVITAS II**	**CIVITAS Plus**
WP12: Clean private and public fleets	WP5: Clean vehicles and alternative fuels	1: Alternaive fuels & clean, energy-efficient vehicles
WP7: Stimulation of PT modes	WP8: Stimulation of PT modes	2: High quality energy-efficient Passenger Transport
WP6: Integrated pricing strategies	WP7: Integrated pricing strategies	3: Economic based Demand Management strategies
WP5: Access restriction	WP6: Access management	4: Mobility Management, Communication & Education
WP10: Innovative soft measures	WP11: Innovative soft measures	5: Safety and Security
WP8: New forms of vehicle use and ownership	WP9: New forms of vehicle use and ownership	6: Mobility services for energy-efficient vehicle use
WP9: New concepts for goods distribution	WP10: New concepts for goods distribution	7: Energy-efficient freight distribution
WP11: Integration of traffic management systems	WP12: Telematics	8: Innovative Transport Telematics systems

Source: Civitas Pointer

Annex III: Civitas I Common Core Indicators

NO.	EVALUATION CATEGORY	EVALUATION SUB-CATEGORY	IMPACT	INDICATOR	DESCRIPTION	DATA /UNITS
	ECONOMY					
1		**Benefits**	Operating Revenues	Operating revenues	Revenues per PT pkm	Euros/pkm, quantitative, derived or measurement
2		**Costs**	Operating Costs	Operating costs	Costs per PT pkm	Euros/pkm, quantitative, derived or measurement
	ENERGY					
3		**Energy Consumption**	Fuel Consumption	Vehicle fuel efficiency	Fuel used per vkm, per vehicle type	MJ/vkm, quantitative, derived or measurement
4				Fuel mix	Energy used per type of fuel, per vehicle type	MJ, quantitative, derived or measurement
	ENVIRONMENT					
5		**Pollution/Nuisance**	Air Quality	CO levels	CO concentration	Ppm or g/m3, quantitative, measurement
6				NOx levels	NOx concentration	Ppm or g/m3, quantitative, measurement
7				Particulate levels	Particulate (pm10) concentration	Ppm or g/m3, quantitative, measurement
8			Emissions	CO2 emissions	CO2 per vkm	G/vkm, quantitative, derived
9				CO emissions	CO per vkm	G/vkm, quantitative, derived
10				NOx emissions	NOx per vkm	G/vkm, quantitative, derived
11				Small particulate emissions	Pm10 per vkm	G/vkm, quantitative, derived
12			Noise	Noise perception	Perception of noise	Index, qualitative, collected, survey
	SOCIETY					
13		**Acceptance**	Awareness	Awareness level	Degree to which the awareness of the policies/measures has changed	Index, qualitative, collected, survey
14			Acceptance	Acceptance level	Attitude survey of current acceptance with the measure	Index, qualitative, collected, survey
15		**Accessibility**	Spatial Accessibility	Perception of PT accessibility	Attitude survey of perception of physical accessibility of PT network (distance to nearest PT stops)	Index, qualitative, collected, survey
16			Economic Accessibility	PT services relative cost	Cost of PT related to average personal income […]	Index, quantitative, measurement
17		**Security**	Security	Perception of PT security	Perception of security when using PT options	Index, qualitative, collected, survey
	TRANSPORT					
18		**Quality of Service**	Service reliability	Accuracy of PT timekeeping	Percentage of services arriving/departing on time compared to timetables (each city should fix the interval of time considered as a delay compared with timetable)	%, quantitative, collected, measurement
19			Quality of service	Quality of PT service	Perception of quality of PT services	Index, qualitative, collected, survey
20		**Safety**	Transport Safety	No. of injuries and deaths caused by accidents	General transport accident no. within the city causing injured and deaths	Quantitative, measurement
21		**Transport System**	Traffic Levels	Vkm by vehicle type - peak	Total trips length per vehicle per day	Vkm per day, quantitative, measured
22				Vkm by vehicle type -off peak	Total trip length per vehicle per day	Vkm per day, quantitative, measured
23			Congestion Levels	Average vehicle speed - peak	Average vehicle speed over total network	Km/hr, quantitative, derived
24				Average vehicle speed - off peak	Average vehicle speed over total network	Km/hr, quantitative, derived
25			Freight Movements	Total no. of goods vehicles moving in demo areas	Assessment of whether the daily no. of goods vehicles accessing city centre changes as a result of the demonstrations	Quantitative, derived or measurement
26			Modal split	Average modal split-PAX	Percentage of pkm for each mode	%, quantitative, derived
27				Average modal split-vehicles	Percentage of vkm for each mode	%, quantitative, derived
28			Vehicle Occupancy	Average occupancy	Mean no. persons per vehicle/day	Persons/vehicle, quantitative, derived, measurement

Source: Civitas Pointer

Annex IV: Civitas II Common Core Indicators

NO.	EVALUATION CATEGORY	EVALUATION SUB-CATEGORY	IMPACT	INDICATOR	DESCRIPTION	DATA /UNITS
	ECONOMY					
1		**Benefits**	Operating Revenues	Operating revenues	Revenues per pkm	Euros/pkm, quantitative, derived or measurement
2		**Costs**	Operating Costs	Operating costs	Costs per pkm	Euros/pkm, quantitative, derived or measurement
	ENERGY					
3		**Energy Consumption**	Fuel Consumption	Vehicle fuel efficiency	Fuel used per vkm, per vehicle type	MJ/vkm, quantitative, derived or measurement
4				Fuel mix	Percentage of fuel used by type	Percentage, quantitative, derived or measurement
	ENVIRONMENT					
5		**Pollution/Nuisance**	Air Quality	CO levels	CO concentration	Ppm or g/m3, quantitative, measurement
6				NOx levels	NOx concentration	Ppm or g/m3, quantitative, measurement
7				Particulate levels	Particulate PM10 and/or PM2.5 concentration	Ppm or g/m3, quantitative, measurement
8			Emissions	CO2 emissions	CO2 per vkm by type	G/vkm, quantitative, derived
9				CO emissions	CO per vkm by type	G/vkm, quantitative, derived
10				NOx emissions	NOx per vkm by type	G/vkm, quantitative, derived
11				Particulate emissions	PM10 and/or PM2.5 per vkm by type	G/vkm, quantitative, derived
12			Noise	Noise perception	Perception of noise	Index (%), qualitative, collected, survey
	SOCIETY					
13		**Acceptance**	Awareness	Awareness level	Awareness of the policies/measures	Index (%), qualitative, collected, survey
14			Acceptance	Acceptance level	Attitude survey of current acceptance of the measure	Index (%), qualitative, collected, survey
15		**Accessibility**	Spatial Accessibility	Perception of accessibility	Perception of physical accessibility of service	Index(%), qualitative, collected, survey
16			Economic Accessibility	Relative cost of service	Cost of service relative to average personal income	Index(%), quantitative, measurement
17		**Security**	Security	Perception of security	Perception of security when using service	Index, qualitative, collected, survey
	TRANSPORT					
18		**Quality of Service**	Service reliability	Accuracy of timekeeping	Number and percentage of services arriving / departing on time	No and %, quantitative, collected, measurement
19			Quality of service	Quality of service	Perception of quality of service	Index, qualitative, collected, survey
20		**Safety**	Transport Safety	Injuries and deaths caused by transport accidents	Number of accidents, fatalities and casualties caused by transport accidents	No, Quantitative, measurement
21		**Transport System**	Traffic Levels	Traffic flow by vehicle type - peak	Average vehicles per hour by vehicle type - peak	Veh per hour, quantitative, measured
22				Traffic flow by vehicle type - off peak	Average vehicles per hour by vehicle type – off peak	Veh per hour, quantitative, measured
23			Congestion Levels	Average vehicle speed - peak	Average vehicle speed over total network	Km/hr, quantitative, derived
24				Average vehicle speed - off peak	Average vehicle speed over total network	Km/hr, quantitative, derived
25			Freight Movements	Goods vehicles moving in demo areas	Daily number of goods vehicles moving in area	No, Quantitative, derived or measurement
26			Modal split	Average modal split-passengers	Percentage of passenger-km for each mode	%, quantitative, derived
27				Average modal split-vehicles	Percentage of vehicle-km for each mode	%, quantitative, derived
28			Vehicle Occupancy	Average occupancy	Mean no. persons per vehicle/day	Persons/vehicle, quantitative, derived, measurement
29			Modal split	Average modal split-trips	Percentage of trips for each mode	%, quantitative, derived

Source: Civitas Pointer

Annex V: Civitas Plus Common Core Indicators

NO.	EVALUATION CATEGORY & SUB-CATEGORY	IMPACT	INDICATOR	DESCRIPTION	DATA /UNITS
	ECONOMY				
1	**Benefits**	Operating Revenues	Operating revenues	Revenues per pkm or vkm	Euros/pkm or Euros/vkm, quantitative, derived or measured
2A	**Costs**	Costs	Capital costs	Capital cost per system or unit	Euros, quantitative
2B			Operating costs	Costs per pkm or vkm	Euros/pkm or Euros/vkm, quantitative, derived or measured
	ENERGY				
3	**Energy Consumption**	Fuel Consumption	Vehicle fuel efficiency	Fuel used per vkm, per vehicle type	MJ/vkm, quantitative, derived or measured
4			Fuel mix	Percentage of fuel used by type	Percentage, quantitative, derived or measured
	ENVIRONMENT				
5	**Pollution and Nuisance**	Air Quality	CO levels	CO concentration	Ppm or g/m3, quantitative, measured
6			NOx levels	NOx concentration	Ppm or g/m3, quantitative, measured
7			Particulate levels	Particulate PM10 and/or PM2.5 concentration	Ppm or g/m3, quantitative, measured
8		Emissions	CO2 emissions	CO2 per vkm by type	G/vkm, quantitative, derived
9			CO emissions	CO per vkm by type	G/vkm, quantitative, derived
10			NOx emissions	NOx per vkm by type	G/vkm, quantitative, derived
11			Particulate emissions	PM10 and/or PM2.5 per vkm by type	G/vkm, quantitative, derived
12		Noise	Noise perception	Perception of noise	Index (%), qualitative, collected, survey
	SOCIETY				
13	**Acceptance**	Awareness	Awareness level	Awareness of the policies/measures	Index (%), qualitative, collected, survey
14		Acceptance	Acceptance level	Attitude survey of current acceptance of the measure	Index (%), qualitative, collected, survey
15	**Accessibility**	Spatial Accessibility	Perception of accessibility	Perception of physical accessibility of service	Index(%), qualitative, collected, survey
16		Economic Accessibility	Relative cost of service	Cost of service relative to average personal income	Index(%), quantitative, measured
17	**Security**	Security	Perception of security	Perception of security when using service	Index, qualitative, collected, survey
	TRANSPORT				
18	**Quality of Service**	Service reliability	Accuracy of timekeeping	Number and percentage of services arriving / departing on time	No and %, quantitative, collected, measured
19		Quality of service	Quality of service	Perception of quality of service	Index, qualitative, collected, survey
20	**Safety**	Transport Safety	Injuries and deaths caused by transport accidents	Numbers of accidents, fatalities and casualties caused by transport accidents	No, quantitative, measured
21	**Transport System**	Traffic Levels	Traffic flow by vehicle type - peak	Average vehicles per hour by vehicle type - peak	Veh per hour, quantitative, measured
22			Traffic flow by vehicle type - off peak	Average vehicles per hour by vehicle type – off peak	Veh per hour, quantitative, measured
23		Congestion Levels	Average vehicle speed - peak	Average vehicle speed over total network	Km/hr, quantitative, derived
24			Average vehicle speed - off peak	Average vehicle speed over total network	Km/hr, quantitative, derived
25		Freight Movements	Goods vehicles moving in demo areas	Daily number of goods vehicles moving in area	No, quantitative, derived or measured
26		Modal split	Average modal split-passengers	Percentage of passenger-km for each mode	%, quantitative, derived
27			Average modal split-vehicles	Percentage of vehicle-km for each mode	%, quantitative, derived
28			Average modal split-trips	Percentage of trips for each mode	%, quantitative, derived
29		Vehicle Occupancy	Average occupancy	Mean no. persons per vehicle/day	Persons/vehicle, quantitative, derived, measured

Source: Civitas Pointer

Annex VI: Changes of CCIs between Civitas I and Civitas Plus

	Evaluation Area	Evaluation sub-category	Impact	Indicator
1	Economy	Benefits	Operating revenues	Operating revenues
2		Costs	~~Operating~~ costs	Operating costs
2A				***Capital costs***
3	Energy	Energy consumption	Fuel consumption	Vehicle fuel efficiency
4				Fuel mix
5	Environ-ment	Pollution & nuisance	Air quality	CO levels
6				NOx levels
7				PM levels
8			Emissions	CO_2 emissions
9				CO emissions
10				NOx emissions
11				PM emissions
12			Noise	Noise perception
13	Society	Acceptance	Awareness	Awareness level
14			Acceptance	Acceptance level
15		Accessibility	Spatial accessibility	Perception of ~~PT~~ accessibility
16			Economic accessibility	~~PT services relative cost~~ relative cost of service
17		Security	Security	Perception of ~~PT~~ security
18	Transport	Quality of service	Service reliability	Accuracy of ~~PT~~ timekeeping
19			Quality of service	Quality of ~~PT~~ service
20		Safety	Transport safety	No of injuries and deaths caused by accidents
21		Transport system	Traffic levels	~~Vkm by vehicle type - peak~~ traffic flow by vehicle type (Veh/h)
22				~~Vkm by vehicle type - off peak~~ traffic flow by vehicle type - off peak (Veh/h)
23			Congestion levels	Average vehicle speed - peak
24				Average vehicle speed - off peak
25			Freight movements	Total number of goods vehicles moving in demo area
26			Modal split	Average modal split - PAX
27				Average modal split - vehicle
28				***Average modal split - trips***
~~28~~ 29			Vehicle occupancy	average occupancy

Source: own representation

Annex VII: Matches between impacts and indicators for TLS10.1

Overall objective	Action	Sub-objective[6]	Evaluation area	Impact	Indicator	Measured for TLS 10.1
Improve & optimize freight delivery in Toulouse	Imple-mentation of new access regulation	Reducing congestion	Transport system	Congestion level	Average speed	no
		Reducing pollution	Environment	Pollution	CO_2 / CO / NOx / PM emissions	no
		Enforcing access restrictions	Society	Acceptance	non-respect through rules violation	no
		Raising acceptance	Society	Acceptance	Acceptance level	yes
			Society	Convenience level	Match between stakeholders' needs and restrictions	yes
	Chrono-post delivery platform	Reducing congestion	Transport system	Congestion level	Average speed	no
		Reducing pollution	Environment	Emissions	CO_2 / CO / NOx / PM emissions	yes
		Demonstrating economic viability	Economy	Costs	Operating costs	yes
		Raising acceptance	Society	Acceptance	Acceptance level	yes

Source: own representation

[6] 'Sub-objectives' are not given in the MERT; they are own additions, taking into account the issues named in the MERT. Thereby, a clear overview of the measured impacts becomes possible.

Annex VIII: Definitions of logistics specific terms for WP core indicators

• **Delivery**	a delivery consists of the handover of one or more goods to one receiver (a shop / restaurant etc.) at a certain time;
• **Delivery stop**	time while vehicle is parked; it includes loading and unloading time;
• **Delivery tour**	consists of one or more deliveries, it starts at the operator's facility / LC / UDC and ends when the vehicle returns to the operator's facility / LC / UDC;
• **Loading**	the process of moving goods from a vehicle's load floor to the agreed location (under the receiver's area of responsibility);
• **Unloading**	the process of moving goods from the place of dispatch (under the sender's area of responsibility) onto a vehicle's load floor;

Source: own representation