Local4Global

SYSTEM-OF-SYSTEMS THAT ACT LOCALLY FOR OPTIMIZING GLOBALLY

611538, FP7-ICT-2013.3.4

Deliverable D 6.1.2:

Evaluation Plan (final version)

Deliverable Version:	D 6.1.2, v.3.0
Document Identifier:	local4global wp6_d6.1.2_rwth_trv_tum_v3.0
Preparation Date:	August 13, 2015
Document Status:	Final
Author(s):	Thomas Schild (RWTH), Roozbeh Sangi (RWTH), Roland Ettinger (TRV), Walid Fourati (TRV), Julia Obermaier (TRV), Michael Krause (TUM)
Dissemination Level:	PU - Public

Project funded by the European Community in the 7th Framework Programme

SEVENTH FRAMEWORK PROGRAMME

EU FP7 - SMALL/MEDIUM-SCALE FOCUSED RESEARCH PROJECT (STREP), FP7-ICT-2013.3.4: ADVANCED COMPUTING, EMBEDDED AND CONTROL SYSTEMS

D) FROM ANALYZING TO CONTROLLING BEHAVIOUR OF SYSTEM OF SYSTEMS (SOS)

European Commission
Information Society and Media

Deliverable SUMMARY SHEET

Deliverable Details	
Type of Document:	Deliverable
Document Reference #:	D 6.1.2
Title:	Evaluation Plan (final version)
Version Number:	3.0
Preparation Date:	August 13, 2015
Delivery Date:	May 11, 2015
Author(s):	Thomas Schild (RWTH), Roozbeh Sangi (RWTH), Roland Ettinger (TRV), Walid Fourati (TRV), Julia Obermaier (TRV), Michael Krause (TUM)
Document Identifier:	local4global wp6_d6.1.2_rwth_trv_tum_v3.0
Document Status:	Final
Dissemination Level:	PU - Public

Project Details	
Project Acronym:	Local4Global
Project Title:	SYSTEM-OF-SYSTEMS THAT ACT LOCALLY FOR OPTIMIZING GLOBALLY
Project Number:	611538
Call Identifier:	FP7-ICT-2013.3.4
Call Theme:	ADVANCED COMPUTING, EMBEDDED AND CONTROL SYSTEMS
Project Coordinator:	CERTH – Centre for Research and Technology – Hellas
Participating Partners:	CERTH – Centre for Research and Technology – Hellas; ETHZ – Eidgenössische Technische Hochschule Zürich; RWTH – RWTH Aachen University; IK4 – IK4 TEKNIKER; TRV – TRANSVER GmbH; TUC – Technical University of Crete; TUM – Technische Universtität Muenchen;
Instrument:	STREP
Contract Start Date:	October 1, 2013
Duration:	36 Months

Deliverable D 6.1.2: Short Description

The Evaluation Plan will prescribe the general evaluation approach, the assessment objectives, the indicators to be used, the methods and timing of measurement and the measurement conditions, and statistical issues such as sample sizes. Establishing this Evaluation Plan is an iterative process based on a first draft version as basis for further discussion within the project and with the project officer. After that the final version will be consolidated, complimented by questionnaire details for the evaluation of the user acceptance.

Keywords: evaluation, assessment, system of systems, use case, traffic, building, control, automation, monitoring, key performance indicator, junction control, building management system, optimization, data base, transfer, questionnaires, user acceptance, methodology, methods

Deliverable D 6.1.2: Revision History

Version:	Date:	Status:	Comments
1.0	05/05/2015	Draft	For partner`s revision
2.0	11/05/2015	Final	Final version
3.0	13/08/2015	Final	Revision with respect to review recommendations

Table of Contents

List of Figures

List of Tables

Abbreviations and Acronyms

ANMS	Average Network Mean Speed
BMS	Building Management System
BUC	Building Use Case
CBA	Cost-Benefit-Analysis
CBR	Cost-Benefit-Ratio
CCA	Concrete Core Activation
CSS	Central Supply Systems
FCD	Floating Car Data
FOT	Field Operational Test
FVU	Façade Ventilation Unit
GKPI	General Key Performance Indicators
HCM	Highway Capacity Manual
HGV	Heavy Goods Vehicle
HMI	Human Machine Interface
IS	Interconnecting Wall Sensor
KPI	Key Performance Indicators
L4G	Local4Global project
LOS	Level of Service
MCIS	Monitoring Control and Information System
NDS	Naturalistic Driving Study
NEC	Net Energy Consumption
NES	Net Energy Savings
NREC	Non-renewable Energy Consumption
NRES	Non-renewable Energy Savings
PC	Personal Computer
PC (BUC)	People Counter
PEC	Primary Energy Consumption
PES	Primary Energy Savings
PI	Performance Index
PS	Presence Sensor
RMSPE	Root Mean Square Percent Error
SoS	System of Systems
SUS	System Usability Scale
TD	Temperature Demand
TrUC	Traffic Use Case

TSoS	Technological System of Systems
VFC	Variable Flow Control
WS	Window Sensor
WTP	Willingness to Pay

Executive Summary

The Local[4]Global project is meant to develop an ambitious and new methodology that facilitates to control a generic TSoS. Therefore, a plug-and-play and ready-to-use system will be the final product of the project. These goals require an exhaustive evaluation within the development process to advance a focused and goal-oriented progress.

The evaluation of the methodology`s impact bases on two use case TSoS in different technical areas, a traffic TSoS and a building TSoS. The methodology will be tested in real TSoS that facilitate to implement and to monitor the system´s operations in the required quality to be able to analyze and evaluate the achievement of the project`s goals. A successful testing of the Local[4]Global methodology would open a valuable and advantageous perspective for further research activities in the appropriate research areas, thus an extensive and detail evaluation renders a wide-ranging exploitation possible.

Hence, the evaluation plan on hand faces the challenge to prepare a detailed and expressive evaluation of an inchoate methodology that is still in development, solely based on the test bed environment that already exists, but is not yet considered as a TSoS, and a generally capable software architecture. The mean of choice is an extensive data gathering and monitoring that provides the foundation of the evaluation on the level of the particular TSoS as well as on the level of generic TSoS.

For both use case TSoS, detached evaluation plans are developed, which meet the use case specific basic conditions and issues, facilitate to determine meaningful KPIs that are suitable for comparison. The general evaluation plan consists of the transfer principles, how the use case specific KPIs will be transferred and conclusions will be made for generic TSoS. In addition to the analysis based on measurements and derived figures, a pre-analysis will take place to investigate and gather indications if and how the methodology will be applicable to the real TSoS. Therewith, an instance of quality management is implemented as well as a first base for evaluation.

The document on hand embodies the final version of the evaluation plan. It is meant to facilitate a synchronous development of all related disciplines that take part in the project. This is an essential requirement for the capability to achieve the project`s goals. To this end, it contains further perceptions and details and presents the methods for the evaluation of the final Local[4]Global methodology`s impact on TSoS.

1 Introduction

The Local4Global project (L4G) aims to develop a new groundbreaking, generic and fully-functional methodology for controlling and optimizing TSoS[1] at the global level, based on local information. Therefore, a control algorithm will be used that is suitable to control generic TSoS without the need for the use of elaborate and effort consuming modelling, analysis and control design tools.

This algorithm/system will be evaluated based on use cases of different technological disciplines to be able to evaluate the potential impact for generic TSoS. Within the traffic use case, the L4G System will be introduced to a rural traffic network in combination with the control of cooperative vehicles. Within the building use case, the L4G algorithm will be implemented to control heating and cooling devices in a building. To do these implementations, the first step is the preparation of plans for the implementation and evaluation phases. The former has been accomplished in deliverables 5.1.2 respectively 5.2.2, and the latter is made in this documentation.

The evaluation plan prescribes the general evaluation approach, the assessment objectives, the indicators to be used, the methods and timing of measurement and the measurement conditions, and statistical issues such as sample sizes. The core of the evaluation plan is an 'Indicator Table' for each use case, which contains for each indicator

- ✓ The name and number of the indicator (e.g. Average Daily Travel Times, Average Daily Fuel or Energy Consumption, Average Daily Fanger Factor[2], Optimality Accuracy, Time Required to Complete the Computations, Time Required to Re-configure, etc);
- ✓ The maximum and minimum allowable values of the indicator;
- ✓ The European standards related to the particular indicator (if any);
- ✓ The data required to measure the indicator;
- ✓ The data sources (e.g. which sensors and devices to be used, the particular time-intervals to be used within the evaluation, etc) and
- ✓ The methods of data analysis (formulas, statistics tests).

Establishing this evaluation plan is an iterative process, which is obtained through communication, discussion and agreement among the L4G project partners point by point about what is feasible, what is not possible and what alternative solutions are available.

This document aims to define the framework for evaluation of the Local4Global system within both, the building use case and traffic use case. The overall structure of the report contains five chapters.

The first chapter presents the general evaluation approach and addresses the objectives of work and also the required instruments like the above-mentioned "Indicator Tables". The evaluation methodology is described in chapter two, where general key performance factors for evaluation are also defined. Chapters three and four deal with the evaluation procedure of the traffic and the building use cases, respectively. A brief description of the systems, how to obtain the required parameters for calculation of key performance factors and how to evaluate the gained results are discussed in these chapters. Additionally, preparation of the use case independent

[1] A Technological System of Systems (TSoS) is an orchestration of human designed systems that built a system of systems (SoS) regarding the way they work and the behavior, characteristics and interaction

[2] Fanger Factor is an indicator for thermal comfort

evaluation is presented in chapters three and four. Finally, comparison of general key performance factors, assessment of the L4G impact and conclusions with a view to generic TSoS are made in chapter five. In addition to evaluation purposes, the results of the developed evaluation plan are applied in the next step of the project, in order to determine additional sensors and monitoring devices that have to be installed at each of the two use cases.

2 General Evaluation Approach

The project aims to develop an algorithm/methodology that will be applicable to generic TSoS as a fully-functional and ready-to-use system, delivered in the form of an embedded, web-based plug-and-play system. Both use cases are meant to test this methodology and to evaluate its impact by implementing it in real-life TSoS. Based on the results and the experience of these tests, conclusions will be made for the deployment on generic TSoS.

Therefore, the evaluation task faces the following challenges:

- Incorporation the project`s assessment objectives and general KPIs (GKPIs)
- Finding comparable KPIs for each use case that relate to the GKPIs
- Preparation of the evaluation procedures to determine all KPIs
- Preparation of the questionnaires that provide the evaluation of users' acceptance
- Preparation the overall evaluation
- Preparation of the assessment of the L4G´s impact on generic TSoS by transferring the evaluation results from both use cases

The general evaluation plan addresses all these challenges and describes the approaches that will be used for evaluation. Chapter 3 gives an overview on the general evaluation methodology, which contains the assessment approach regarding the transferability to generic TSoS and the evaluation methodology that will be used in both use cases structurally. The use case specific evaluation plans are described in chapters 4 and 5, where the specific assessment objectives and methodology, evaluation methodology and the preconditions and circumstances are discussed. Additionally, the instruments will be presented in these two chapters as well as a brief description of the particular use case. In chapter 0, conclusions will be made regarding the future evaluation purpose.

2.1 Comparability

The evaluation of two use cases in different research areas requires a well-aligned evaluation of both use cases, if the outcome of both evaluations has to be comparable regarding the overall assessment objectives. For sure, each subject area contains its own perspective on the control purpose that underlies L4G and, therefore, each particular TSoS requires specific evaluation methods and target figures, while a joint subset of these figures has to be evaluable.

Despite the use case specific peculiarities, the required comparability will be established by focusing on the project`s assessment objectives and the related KPIs. The introduced common KPIs will be used as far as possible to enlarge the evaluation`s findings.

2.2 Transferability to generic TSoS

The comparable KPIs will be transferred to the GKPIs as a foundation to assess the L4G´s impact on generic TSoS. Finding conclusions for a totality based on a sample size of two use cases is hard to achieve, but the use cases gives a first orientation and indication for the impact of L4G. Therefore, the KPIs' average values will be used with a spread in size of the deviations as well as the questionnaires` outcome, discussed concerning the tendency of the users' acceptance.

3 General Evaluation Methodology

3.1 Assessment Objectives

In DoW (1.1.5) the measurable objectives in form of Key Performance Indicators (KPIs) are defined that will lead through the evaluation task. They are related to one or more of the main project objectives that are given in Table 1 in short terms.

Table 1: Main project objectives (DoW, Annex I, Table 1)

Index	Description
O1	The TSoS constituent systems are operating as fully autonomous units that react and interact depending only on their local environment in order to optimize the TSoS emerging performance at the global level.
O2	There will be no need for an elaborate and tedious effort to deploy the L4G system or to redesign/re-configure in cases of changes. The L4G system provides a plug-and-play control mechanism that embed learning, evolving and self-organizing capabilities
O3	There will be no need for an elaborate, “expensive” infrastructure.
O4	The L4G methodology will be applicable to generic TSoS.

These objectives focus on the System-of-Systems (SoS) character that is the foundation for the L4G development and substantiates the ambitious challenge to aim on generic TSoS control.

Both use case TSoS are of SoS character as discussed in D2.3 (Ettinger, Krause and Schild, et al. 2014) and suitable to test the L4G methodology. For both use cases, measurable KPIs are defined that embody the core elements of the evaluation beside the questionnaire`s outcome.

Table 2 gives an overview and a brief description of the main assessment objectives, the target improvements, and the relation to the above-mentioned main project objectives. Both use cases are suitable to evaluate the achievement of the project target as far as a real TSoS is involved. The assessment of O4 is part of the verification process and described in the implementation plans of both use cases (Ettinger und Krause 2014) (Schild and Sangi 2014).

Table 2: Main assessment objectives (DoW, Annex I, Table 3)

Index	Description	Target	Main Objective
GKPI 1	*Traffic Use Case:* Difference in terms of daily average network speed	Greater than 30 % improvements as compared to the base case	O1, O2, O3
	Building Use Case: Difference in terms of daily average energy consumption from non-renewable sources while maintaining users' comfort (Fanger Factor) at an acceptable level		
GKPI 2	Equal to GKPI 1, but when major (artificial) incidents are present	Greater than 20 % improvements as compared to the base case	O1, O2, O3
GKPI 3	Users' satisfaction metric to be calculated through questionnaires	Less than 10 % of users prefer the base case system	O1, O2, O3

3.2 Assessment Methodology

In this section, the methodologies for assessing each particular use case and the transfer to GKPIs are described. Herein, KPIs represent the evaluation`s outcome of measurable figures, delimited to the results of the questionnaires that deliver statistical figures.

3.2.1 GKPIs

Regarding the duality of the evaluation, the GKPIs have to be determined by the equivalent KPIs within each use case. As mentioned in section 2.2, the average values of these figures will be determined with a spread in size of the deviations as far as deeper statistical analysis is not reasonable.

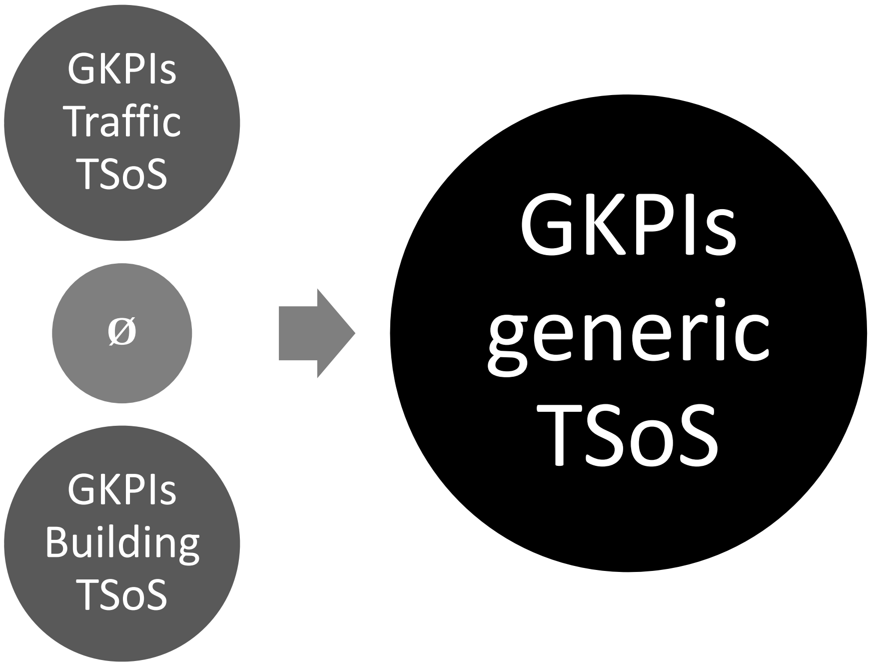

Figure 1: Assessment of general KPI for generic TSoS

The definition of the evaluation procedure is given with the following formulas:

$$GKPI_n = \%GKPI_n \pm \%gkpi_n \tag{3.1}$$

where:

%GKPI = percentage of improvement of GKPI due to the L4G methodology

%gkpi = spread of percentage of improvement above both use cases

$$\%GKPI_n = \frac{\%KPI^{eq}_{n,TrUC} + \%KPI^{eq}_{n,BUC}}{2}$$

where:

%(G)KPI = improvement of KPI due to the L4G methodology (3.2)

TrUC = traffic use case

BUC = building use case

eq = use-case-equivalent to GKPI

n = identifier of figure

The spread is equal to the difference between the GKPI and the KPI of one use case.

$$\%gkpi_n = \%GKPI_n - \min\left(\%KPI^{eq}_{n,TrUC};\%KPI^{eq}_{n,BUC}\right) \tag{3.3}$$

where:

%gkpi = spread of percentage of improvement

%GKPI = percentage of improvement for the general KPI due to the L4G methodology

$\%KPI^{eq}$ = percentage of KPI, equivalent to GKPI

n = identifier of figure

TrUC = Traffic Use Case

BUC = Building Use Case

Therefore, each KPI will be calculated twice, for the system with the current control algorithm (base case) and for the L4G controlled system (test case), to calculate the percentage of improvement.

$$\%KPI_n = \frac{KPI_{n,base} - KPI_{n,test}}{KPI_{n,base}} \cdot 100 \quad (3.4)$$

where:

%KPI = percentage of improvement for the KPI due to the L4G methodology

base = determined for base case

test = determined for test case

n = identifier of figure

3.2.2 Determination of GKPI

GKPI 1 and 2 are based on measurable values, aggregated to the use case specific KPIs and transferred to the GKPIs per use case. Measurements for the values required for KPIs determination will be implemented in the use case TSoS. While the tests are running, these measurements are logged in a data storage and available for evaluation. Using evaluation procedures (e.g. programmed as script or visualized by some exploitation options), the use case specific KPIs will be determined and as far as possible aggregated to the GKPIs equivalent KPIs, that will be transferred to GKPIs and allow assessment and making conclusions for generic TSoS.

Figure 2: Evaluation methodology for use case TSoS

3.3 Evaluation Methodology

In this section, the general evaluation process is described from the point of data gathering to the final use case specific KPIs. Also interrelations to the implementation strategy are presented, which increase understanding, experience and knowledge about the L4G methodology and its operational behavior.

3.3.1 Determination of KPI

A brief description of the meaning and the formula for KPI determination is given in the use case specific sections. Based on the formula, the needed values for KPI determination is identifiable. These obtained values have to be stored in the data storage to be available for evaluation procedures. The evaluation procedures contain the definitions, methods and mathematical relations as described in this deliverable (see sections 4.3 and 5.4.1).

3.3.2 Measured data

The data for KPI determination will be gathered automatically in a data base that facilitates a detailed evaluation as far as required. The before-mentioned simulation data and the measurements that will be done whilst the algorithm is in operation, embody the base of all determinations.

The data acquisition and storage is system specific and will be described in the future chapters in the particular use case context.

3.3.3 Simulations

A pre-analysis will be done before the L4G algorithm is implemented within the real TSoS to prevent incidents because of failures or malfunctions of the algorithm in connection to the TSoS. For both use cases, a base simulation already exists and will be modified as far as needed to get a notion of the SoS` behavior under the L4G control.

The simulation will be simplified and does not raise the claim to be complete in relation to the real TSoS, but it will produce an indication whether the algorithm will operate successfully and what kind of problems can be expected. The results of the simulations will be compared to the later operations, thus the evaluation will be based on all available information about the methodology.

3.3.4 Questionnaires

The evaluation of the users' acceptance is part of the main assessment objectives. Therefore, questionnaires are required that facilitate statistical analysis to achieve representative and comparable figures based on subjective perceptions. Beside the base issues of the questionnaire`s construction, sample sizing and evaluation, the choice of a suitable medium is a challenge regarding the reachability and the willingness to participate of the users' structure.

The questionnaires will be constructed as short as possible to obtain the highest ratio of participants and as focused as possible to achieve a maximum of significance. Both use cases will take previous disseminations of similar questionnaires as well as latest perceptions and latest lessons learned into account.

The questionnaires will be evaluated statistically focusing on the users' acceptance. Therefore, methods of descriptive statistics are suitable. Further details will be given in the use case specific sections.

4 Traffic Use Case

4.1 The Use case

4.1.1 System Description

Aiming to introduce the Local4Global System into practice in the frame of the Traffic TSoS Use Case, a test bed in the north of Munich, Germany was defined to test and evaluate the outcomes of the project. The test bed is a road section of the federal road B 13 between the highway junction Unterschleißheim in the north and the junction of the federal roads B 13 and B 471 in the south (cf. Figure 3). This road section with a length of more than five kilometers allows testing new control approaches and technologies in a rural environment with comparably long distances between intersections and high-speed sections. The speed limit varies along one driving direction between 70 km/h and 100 km/h. Between the traffic signals at Landshuter Strasse (Junction #2) and Inhauser Moos (Junction #1) the maximum allowed speed is 60 km/h. The whole section has two lanes per driving direction and seven signalized intersections.

#1: Inhauser Moors

#2: Landshuter Strasse

#3: Weihenstephaner Strasse

#4: Münchner Ring

#5: Kreuzhof

#6: Franz-Lehner-Straße

#7: B 471

Figure 3: Test Bed Munich for Traffic TSoS Use Case

Due to relative high traffic loads on the one hand and comparably highly developed hardware infrastructure on the other hand, the section has a strong potential for improvement in terms of traffic control. The seven signalized junctions of the road stretch are equipped with induction loops to detect the traffic flow in junction areas. At all junctions, traffic signal controllers of the traffic signal manufacturer Siemens of the latest generation are in operation. The controllers are able to receive and store data from roadside detectors. Furthermore, traffic controllers are able to archive signalization data. Detection and signalization data can be transmitted via mobile network to a central traffic computer for further processing.

4.1.2 L4G Traffic Use Case

The Local4Global system aims to be a generic plug-and-play evolutionary optimization software package to improve the performance of generic TSoS globally on the basis of local information. The product's performance is tested and evaluated by means of two specific use cases. With respect to the project's Traffic Use Case, the product aims to autonomously optimize the current signal traffic control concept as described in the following section.

4.1.3 Test Case Control

After implementation in the Traffic Use Case, the Local4Global system aims to optimize the current traffic signal control in the test bed as described in the previous sections. The implementation of the Local4Global product in the Traffic Use Case incorporates two major applications. First, cooperative vehicles in the test bed shall receive signal timing information while approaching signalized intersections. A smartphone-based application will derive and display appropriate speed recommendations, aiming to avoid stops at signalized intersections. Secondly, the Local4Global system will autonomously optimize signal control on the basis of test bed data. With respect to the technical conditions of the test bed, the Local4Global system can be provided with signalization and detection data as well as GPS trajectories from cooperative vehicles for optimization purposes. Summing up, the Local4Global systems' functionalities aim to improve performance, safety and environmental capability while being highly effective concerning costs and operation defined as capability objectives in Deliverable D2.3 (Ettinger, Krause and Schild, et al. 2014).

4.1.4 Assessment Objectives

Capability objectives describe benefits that should result from the functionalities provided by the system. From the perspective of evaluation, capability objectives lead to assessment objectives defining the qualitative reference goals if aspired targets are met. Assessment objectives used for evaluation can be structured and grouped with respect to different features. Within the TSoS Traffic Use Case in Local4Global, four main assessment objectives are chosen to structure the criteria: efficiency, profitability, human factors and functionality. Each group contains one or several criteria (cf. Figure 4).

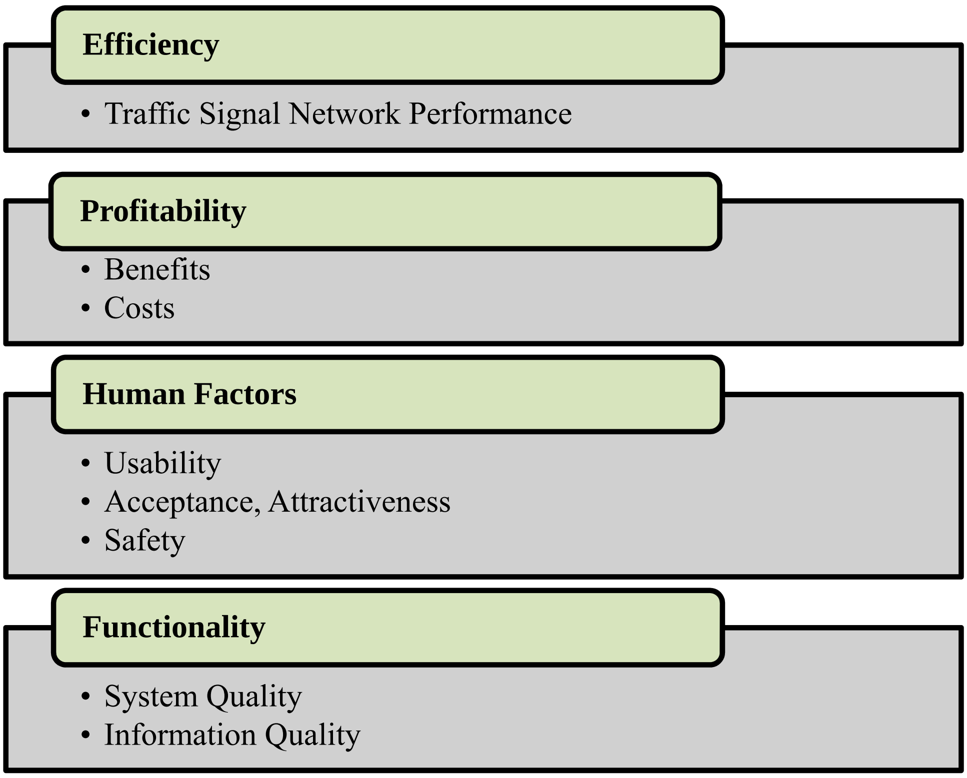

Figure 4: Assessment objectives and criteria

4.2 Baseline

Currently, traffic signal systems in the test bed operate coordinated. The well-timed coordinated system is under operation during off-peak period on weekdays between 9.30am and 2.30pm. The concept of coordination permits continuous movement along an arterial and supports the formation of platoons leading significantly to an improvement of traffic flow along with the reduction of travel times, stops and delay. Figure 5 illustrates the result of coordination showing trajectories of moving vehicles in two directions through a system of coordinated traffic signals using a graphical representation known as a time-space diagram (FHWA 2008).

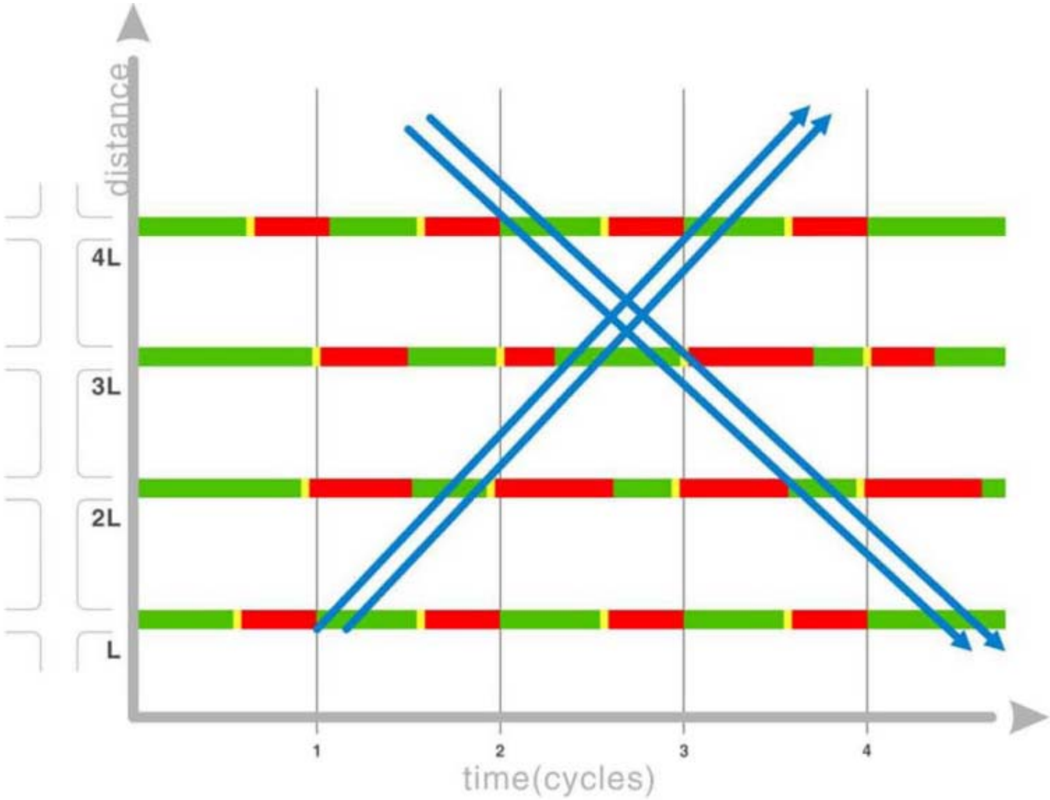

Figure 5: Time-Space Diagram of a Coordinated Timing Plan (FHWA 2008)

Coordination can either be of fixed time or traffic actuated nature. In both cases the cycle lengths of all signal plans at all signalized intersection as being part of the coordinated road stretch have to be equal. Besides requiring a background cycle length for synchronization purposes, coordination of traffic signals leads to further constraints: one signal phase has to be defined as a coordination phase. This phase always receives a minimum amount of assigned green time. Although it is possible to have a portion of the coordinated phase be actuated, i.e. traffic-responsive, the designated coordination phase requires a non-actuated interval that is guaranteed every cycle for the purpose of coordination (FHWA 2008). In other words, traffic dependency needs to be realized within a framework. Concerning base case control in the test bed, a traffic-actuated coordination being bound to a rather tight framework is under operation, i.e. the framework allows only small traffic-actuated deviations at some intersections.

4.3 Evaluation Methodology

4.3.1 Methodological Approach

The method of evaluation chosen for the assessment of the TSoS Traffic Use Case is basically combining aspects of three widely applied approaches. The evaluation methodology follows the

Guidebook for Assessment of Transport Telematics Applications: Updated Version (Zhang, et al. 1998), the V-Model from software engineering projects and integrates ideas from the Information Systems Success Model by DeLone and McLean (2003).

The definition of evaluation steps, grouping of assessment objectives, impact assessment by means of evaluation measures as well as the choice of assessment methods is based on the quite generic guideline from Zhang et al. (1998) developed within the European project CONVERGE. The approach distinguishes between the following major steps concerning the evaluation procedure:

1. **Determination of user needs**: the determination of user needs, capability objectives and system requirements is subject of Deliverable 2. (Ettinger, Krause and Schild, et al. 2014).
2. **Description of applications**: the Local4Global product aims to be a web-based plug-and-play system, evolutionary and self-organizing, optimizing the global performance of the Traffic TSoS based on local information. The system is the basis for the traffic-related applications being developed within the project i.e. advanced traffic control optimization on the one hand and smartphone application displaying speed recommendations on the other hand.
3. **Definition of assessment objectives**: the capability objectives as listed in Deliverable D2.3 (Ettinger, Krause and Schild, et al. 2014)lead to assessment objectives named and grouped in section 4.1.4.
4. **Pre-assessment of expected impacts**: Local4Global intends to globally improve the performance of TSoS. With respect to the Traffic Use Case, the main objective of Local4Global is to improve efficiency in traffic control described by a variety of Key Performance Indicators (KPI) like Level of Service (LOS), network mean speed or Traffic Performance Index (TPI). The assessment objectives according to the DoW are listed in section 3.1 of this document. According to this, Local4Global aims to e.g. improve the daily average network speed by 30 %.
5. **Assessment methodology**: The assessment methodology developed for Local4Global as documented in this section is grounded on assessment objectives and the available data to quantify the system’s benefits or detriments in the line with the objectives by means of KPIs. In this context, a KPI is a quantitative or qualitative indicator, derived from one or several measures, agreed on beforehand, expressed as a percentage, index, rate or other value, which is monitored at regular or irregular intervals and can be compared to one or more criteria (FOT-NET 2011).

Besides incorporating the major steps towards evaluation from the CONVERGE methodology, the approach chosen in Local4Global is based on procedures of the V-Model – widely used in software development. As illustrated in Figure 6, the V-Model describes the road map from the idea of the Local4Global system via requirements analysis, design specification and software development towards testing, implementation, verification, validation and operation. These steps are described by the design cycle. The evaluation steps are represented by the assessment cycle. Starting from the idea, especially identified requirements and derived specifications define the goals for evaluation.

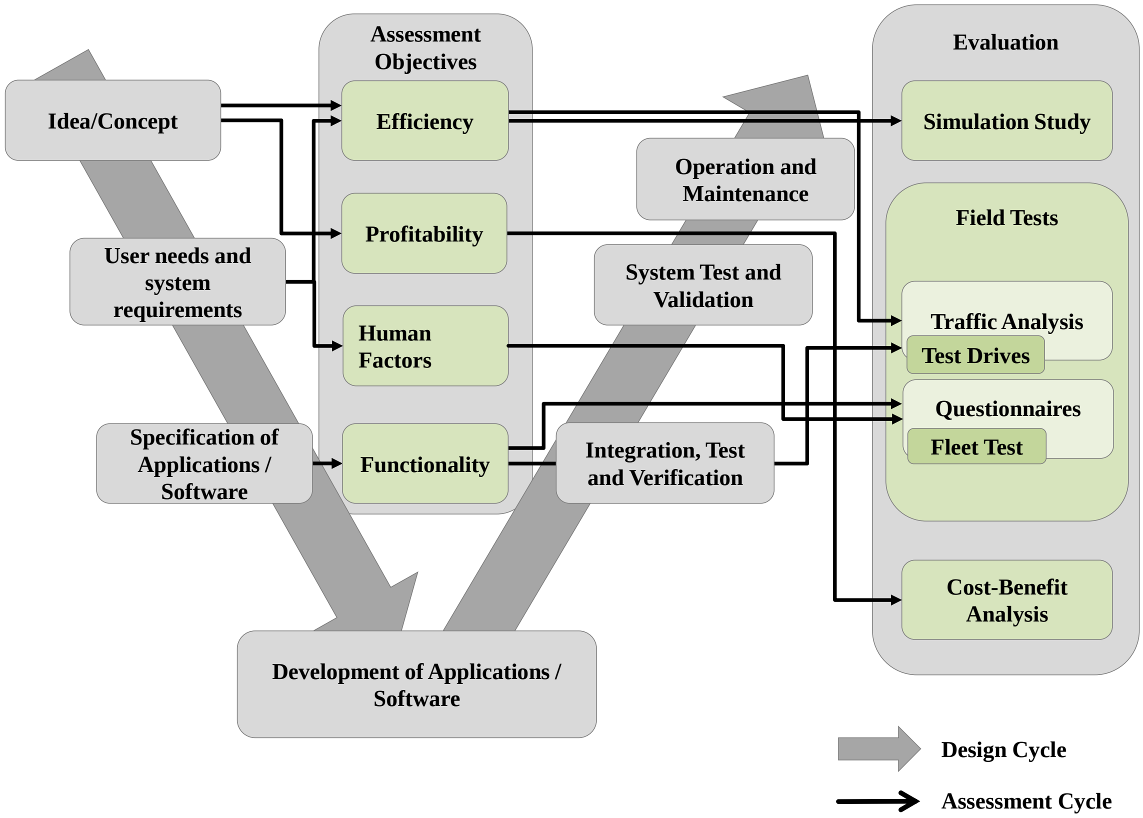

Figure 6: V-shape cycle for TSoS Traffic Use Case Evaluation

As illustrated above, assessment objectives are of technical nature (functionality), cover impact assessment (system performance, efficiency), target human factors (safety, user acceptance) as well as monetary aspects (profitability). Assessment objectives are described by KPIs determined by means of various evaluation methods: Within Local4Global, a simulation study, field tests as well as a Cost-Benefit Analysis (CBA) are considered to be appropriate evaluation methods. Field test data as well as simulation output provide the basis for a traffic analysis. Benefits derived from the traffic analysis as well as identified costs provide input for the CBA. Human factors are determined via experiments and questionnaires to cover both aspects like usability, acceptance and safety as well as availability and quality of functionalities.

The approach to determine the latter aspects is borrowed from DeLone & McLean (2003) but is significantly adapted and shortened. The approach grounds individual impact of a system on the two aspects of system quality and information quality (cf. Figure 7). Although the organizational impact will be excluded in the questionnaire due to the properties of Local4Global being a research project, the individual impact for users of a technical system like the Local4Global product somehow define the organizational impact as users and their usage of the system influence the system’s global performance. In this context, the approach matches the Local4Global idea of optimizing constituent systems on the basis of local information in order to achieve a global optimum. Addressees of questionnaires for determining the individual impacts regarding system and information quality are users of the Local4Global application, i.e. primarily test subjects using and assessing the smartphone application during the test period.

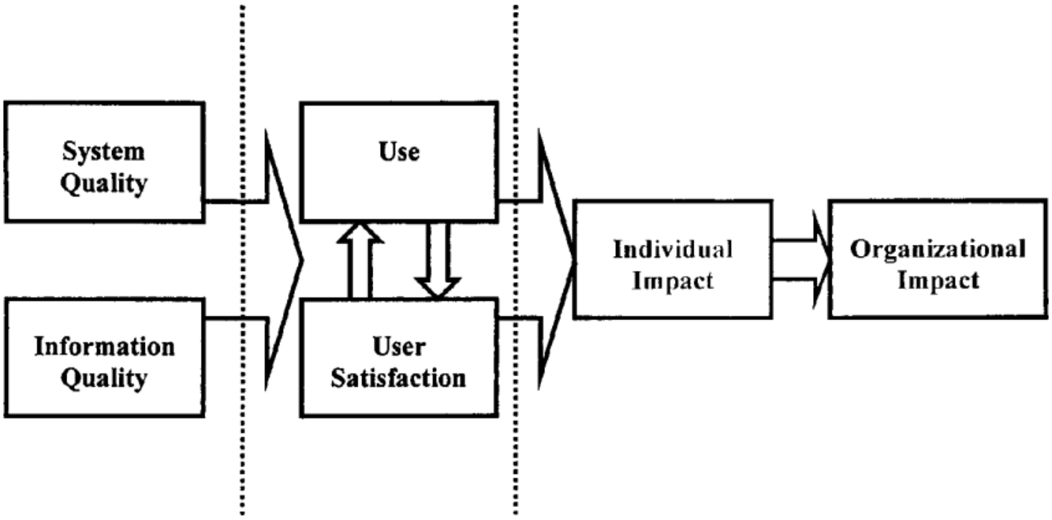

Figure 7: DeLone and McLean Information Systems Success Model (DeLone and McLean 2003)

4.3.2 Measures and KPIs

As described above, KPIs are used to describe assessment objectives quantitatively or at least qualitatively. The indicators are derived from measures basically grounded on available information. From this point of view, available information defines measures whereas measures define evaluation methods enabling impact assessment resulting in KPIs for evaluation purposes. The relationship of chosen methods, measures and KPIs is illustrated in Figure 8. Evaluation methods are in detail described in the subsequent section and KPIs are defined starting in section 4.3.6.

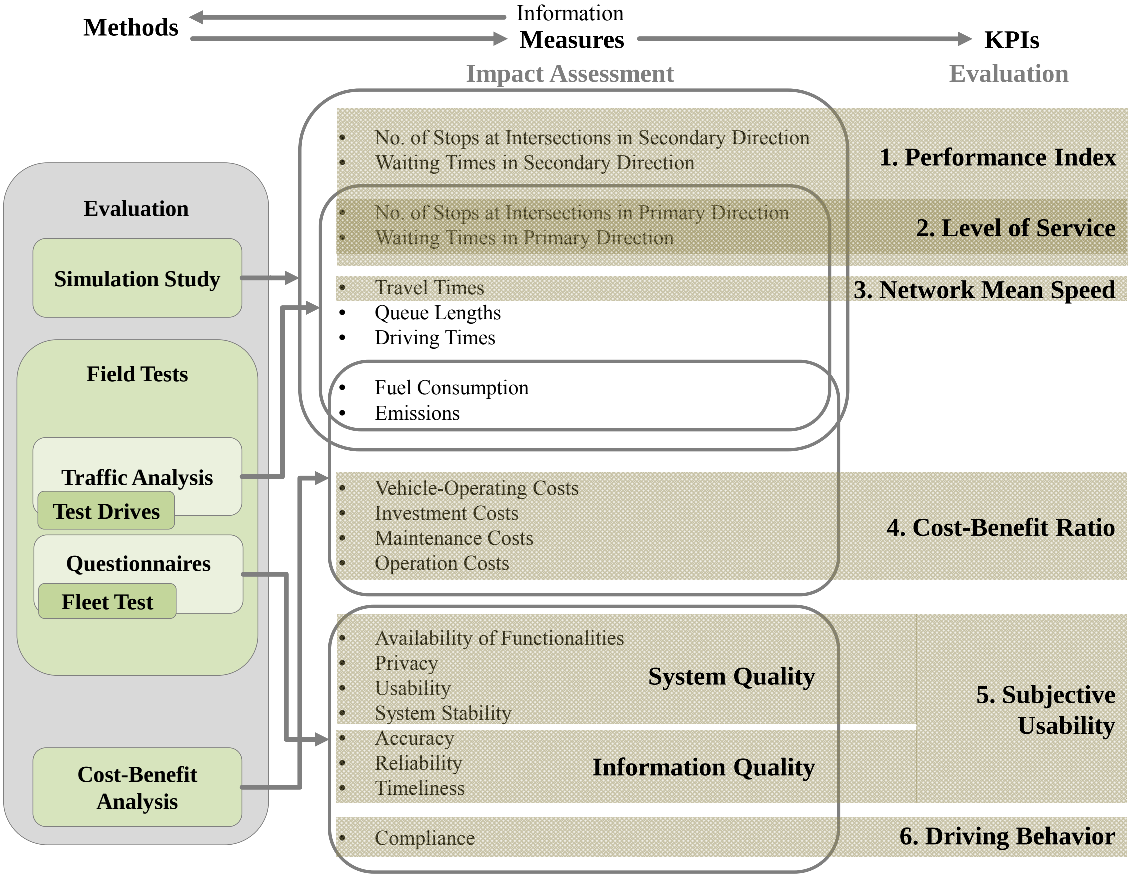

Figure 8: Evaluation Methods, Measures and KPIs

4.3.3 Methods

4.3.3.1 Simulation Study

Before starting the testing period of the Local4Global control system in the test field, the system performance will be evaluated by means of a microscopic traffic simulation. The simulation environment offers the possibility to assess the functionality, operational safety and performance of the Local4Global control approach without being bound to any technical, administrative or legal restraints. A positive assessment of the control approach in the simulation environment may also serve as system verification and provide a basis for argumentation towards the introduction of the system into practice.

4.3.3.1.1 Traffic Simulation Model

The microsimulation study will be carried out in the simulation environment VISSIM. VISSIM is a microscopic, time step and behavior-based traffic flow simulation model for visualization, analysis and optimization of traffic-related scenarios developed by PTV Group, Germany. VISSIM uses the psycho-physical driver behavior model developed by WIEDEMANN (PTV AG 2011). WIEDEMANN basically distinguishes between four different situations to simulate a vehicle's driving behavior as illustrated in Figure 9: free driving (i.e. no reaction), approaching (i.e. reaction), following (i.e. unconscious reaction) and deceleration.

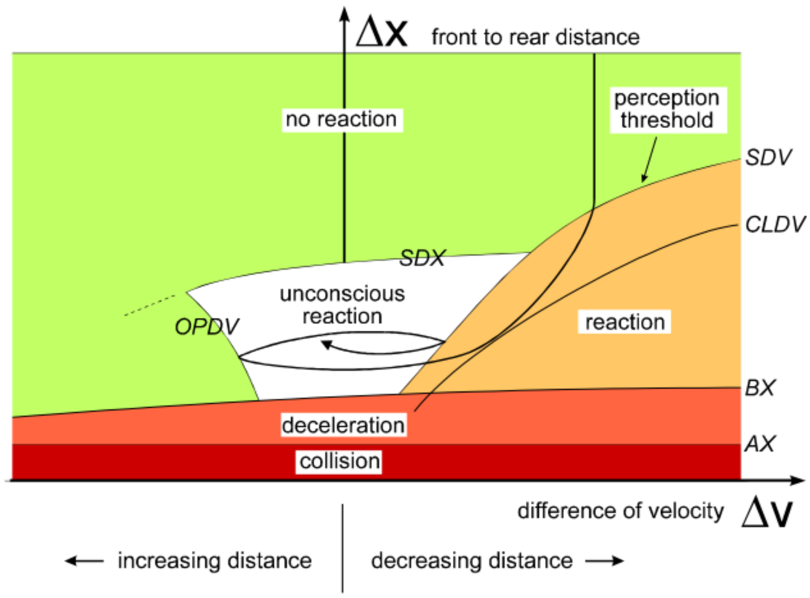

Figure 9: Car Following Logic by WIEDEMANN (PTV AG 2011)

"The basic concept of this model is that the driver of a faster moving vehicle starts to decelerate as he reaches his individual perception threshold to a slower moving vehicle. Since he cannot exactly determine the speed of that vehicle, his speed will fall below that vehicle's speed until he starts to slightly accelerate again after reaching another perception threshold. This results in an iterative process of acceleration and deceleration." (PTV AG 2011) Consequently, in particular parameters like desired speeds and deceleration as well as acceleration parameters in VISSIM all described by distributions are the main variables for modelling the movements of vehicles in the simulation environment. These parameters affect significantly capacities, level of service and travel speeds. The desired speed represents the speed a driver is willing to choose for traveling if not hindered by other vehicles or network elements like traffic signal systems, stop signs or speed limits. In general, the desired speed is geared to the speed limit. In VISSIM, each vehicle gets initially assigned a quantile for speed, acceleration and deceleration distributions when entering the network.

4.3.3.1.2 Microsimulation Study Organization

The process of carrying out a microsimulation study in Local4Global follows the *Guidelines for Applying Traffic Simulation Modeling Software* (FHWA 2004) and accordingly the German equivalent (FGSV 2006). Under these guidelines, the overall process can be subdivided in the steps preparation, base model development, calibration, simulation experiment and documentation. The slightly adapted approach for Local4Global is illustrated in Figure 10.

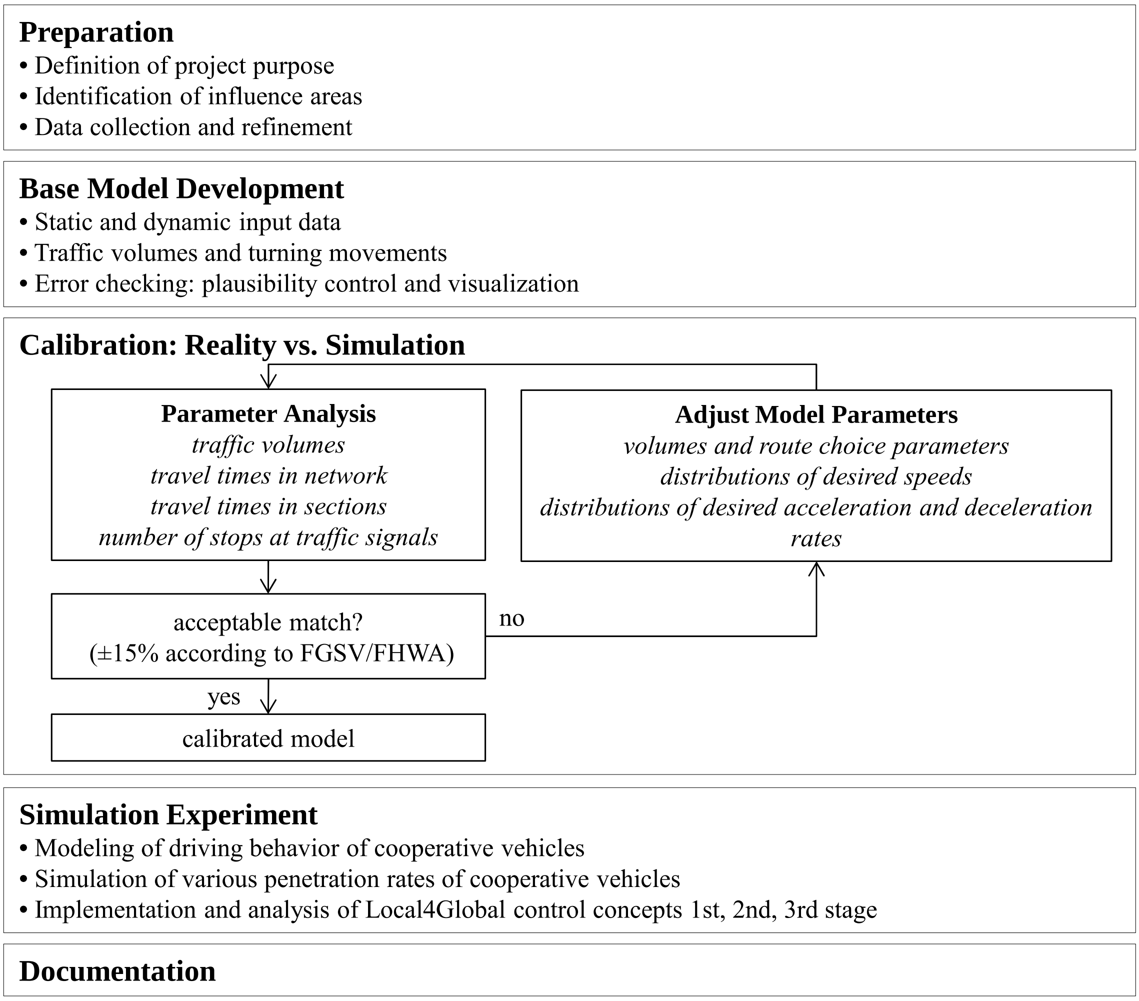

Figure 10: Microsimulation Model Development and Application Process in Local4Global

4.3.3.1.3 Starting Position in Local4Global

The microsimulation study in Local4Global is grounded on an available simulation model set up in VISSIM. The test bed is modelled on the basis of official maps from the *Landesamt für Vermessung und Geoinformation* (Bayern Bavarian Agency for Surveying and Geoinformation) (cf. (PTV AG 2011)). Each intersection is modelled in detail on the basis of signal layout plans as illustrated in Figure 12 for intersection #1 (Inhauser Moos).

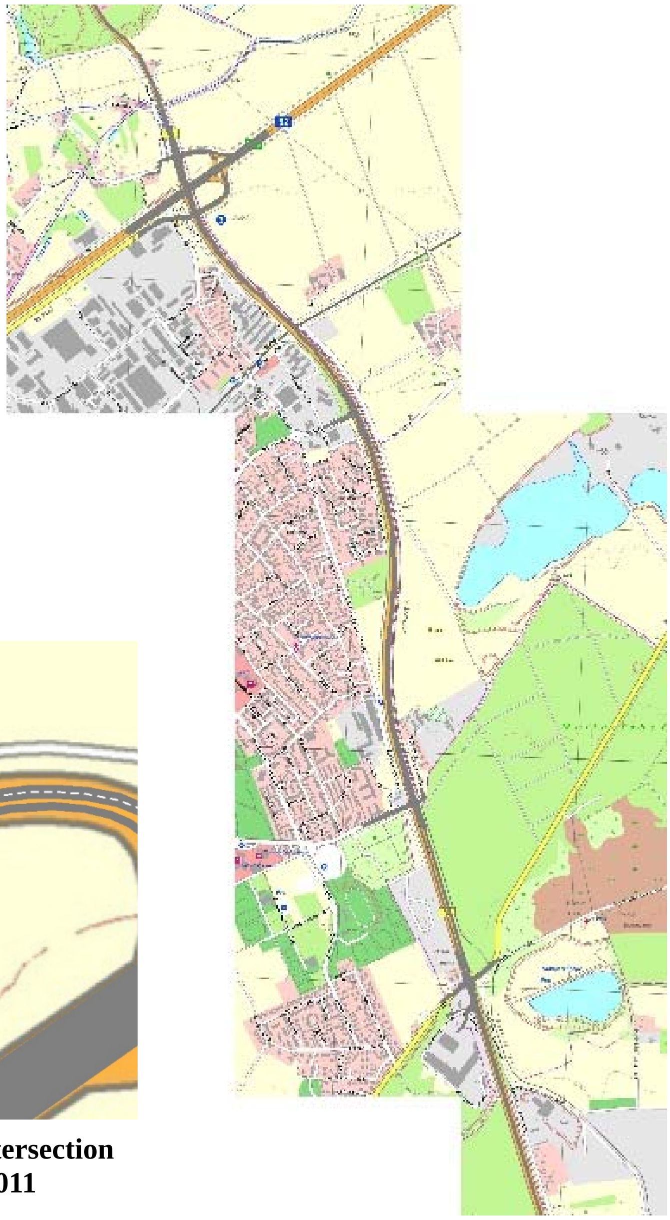

Figure 11: VISSIM network model. Map: Landesamt für Vermessung und Geoinformation Bayern, 1053/08

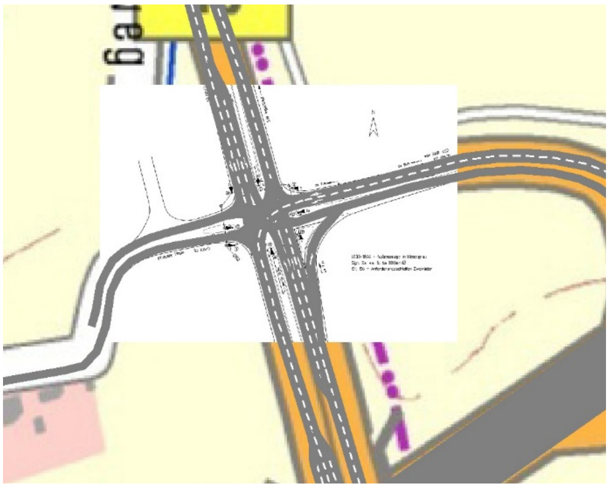

Figure 12: Detailed Modeling of Intersection Layouts from Siemens AG, 2011

The available model is calibrated on the basis of traffic volumes and turning movements from traffic counts carried out during off-peak periods (weekdays, 9.30 am to 2.30 pm). Distributions for desired speeds, accelerations and deceleration are calibrated on the basis of travel times and detected speeds from test drive data (GPS trajectories). Furthermore, the behavior of influenced drivers or rather cooperative vehicles is calibrated and implemented. Considering this background, the microsimulation study in Local4Global needs only few adaptations before integrating and testing the control logic.

4.3.3.1.4 Further steps in Local4Global

Further traffic counts and calibration activities will need to be carried out if the Local4Global system shall operate on a 24/7 basis, especially to adequately model the peak hour traffic conditions. An available macroscopic traffic model of the region could also deliver further information on demand and loads. The behavior of influenced drivers might have to be recalibrated depending on the exact functionalities provided by the Local4Global system. Starting from the previous simulation, the following steps will need to be carried out paving the way towards simulation experiment, determination of KPIs and evaluation:

1. **Re-Calibration** of the existing simulation has to be done since the traffic demand shall be variable following a time variation curve. The demand would be constant per interval of time (e.g. 15 min) to get the maximum closer to reality and most importantly to test the control logic with changing input.

   The calibration process uses the real observations to compare them to the simulation results (e.g. travel time) through the root mean square method.

   $$RMSPE(X) = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(\frac{X_{sim,i} - X_{real,i}}{X_{real,i}}\right)^2}$$

   Where:

   RMSPE = Root Mean Square Percent Error [-]

   n = number of measurements [-]

   $X_{real,i}$ *= average value (traffic volume, travel time) of inverstigation period at measurement i [Kfz/h]*

   $X_{sim,i}$ *= average value (traffic volume, travel time) of some simulation runsat measurement i [Kfz/h]*

   As soon as the Root Mean Square Percent Error is lower than 15% the calibration can be finished. (FGSV 2006)

2. **Implementation** of Local4Global traffic signal system control methods in VISSIM. The software's COM Application Programming Interface (API) allows the implementation of advanced traffic control methods like the Local4Global control algorithms. Figure 13 below shows an overview of external control implementation.

   The external control is made by the library TUC.dll (Diakaki C, et al.). During the initialization of the simulation, TUC is initialized as well through a TUC_INIT function. Signal plan libraries installed in each junction controller are a static input to the TUC (to obey to the plug & play feature) and are preprogrammed in the simulation.

   In the same way, description information for junctions, links (directions) and detectors are stored in an external location (e.g. Excel file). During each time step, the first operation is to collect traffic data (flows and occupancies) from the simulation (equivalent to a data collection by the junction controller and then by the traffic computer), process it and average it. Data from some imaginary detectors is calculated (according to formulas retrieved from the external junctions description file) to compensate missing detectors in real life. Traffic variables are then rearranged per links.

The following operation in each time step is to check for each junction if we attained the desired control interval (e.g. a cycle). When reached, collected data are written to an exchange file that will be read by the L4G control unit to generate a signal plan selection for the concerned junction. A second exchange location is reserved for new signal recommendation by L4G, with a file for each junction. Via the COM interface the new plan is sent to the corresponding Vissim junction controller to be applied in the next control interval.

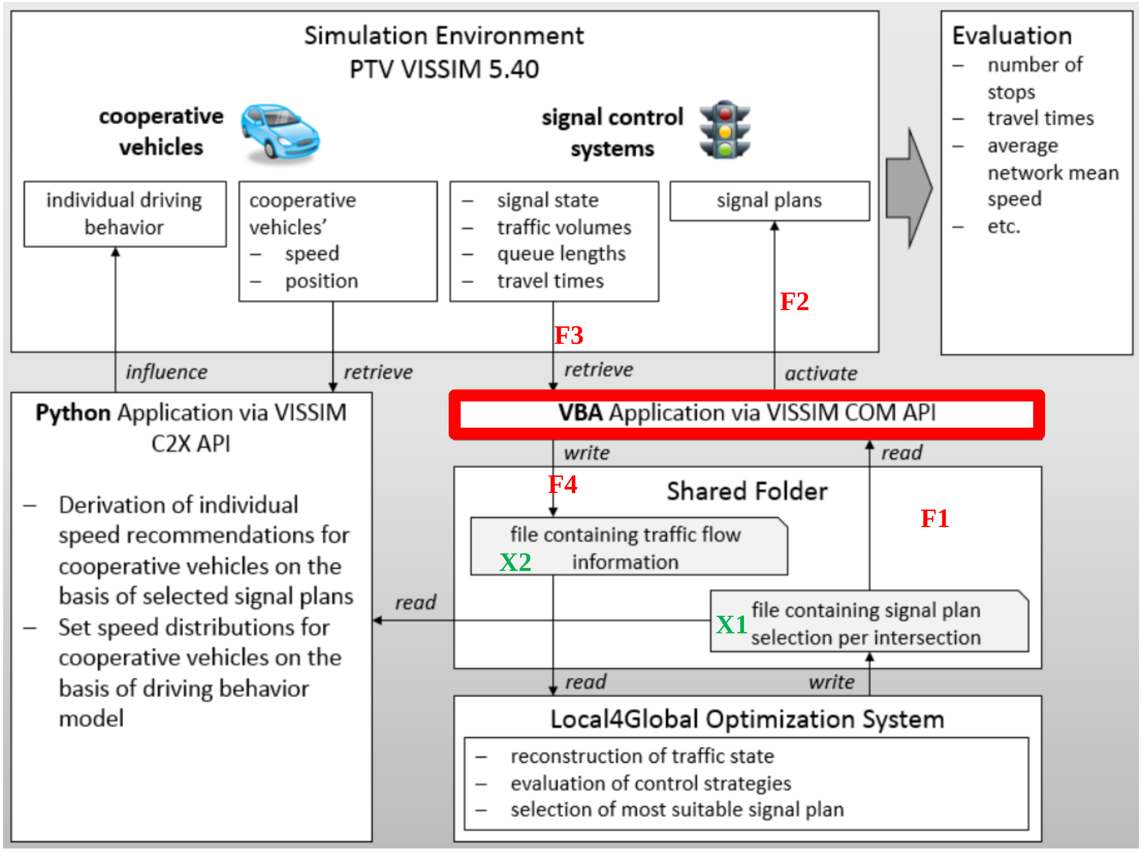

Figure 13: Implementation of external signal control in the simulation

F1: Read an exchange file (X1) containing signal plan selection per intersection

F2: Activate a selected signal plan for each intersection in VISSIM

F3: Retrieve information from the simulation (signal state and detector data)

F4: Write in an exchange file (X2) information retrieved from the simulation

The existing C2X interface is upgraded to be able to consider dynamic signaling after new signal plans generated by L4G. First this needs to describe the libraries in a static file (e.g. XML file containing all junctions' possible signal plans, with the relevant information for the cooperative vehicles as cycle time, start and end of green time). The C2X component reads at regularly the control output from the same exchange location of the COM interface. With this modification, cooperative vehicles inside the simulation would be recommended to run in speeds corresponding to the new signaling configuration.

All of these operations happen each time step and should happen in a very particular order to ensure the validity of the exchanged data between all of the control modules, as shown in Figure 14.

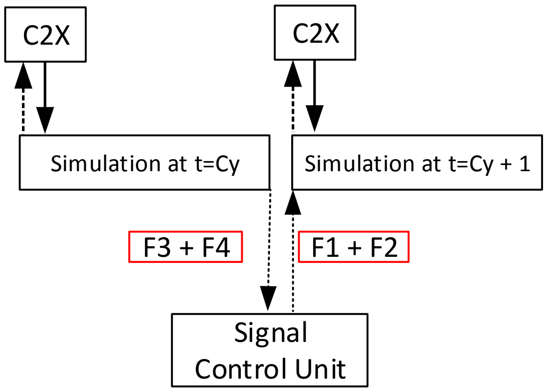

Figure 14: Execution order at each time step

At the end of the simulation, the learning mechanism of L4G is called to assess the performance of the past control from data written by the TUC.dll and fine-tune the optimization parameters.

3. **Writing of the new signal plans** to start with in the simulation. The simulation is the perfect occasion to write and fine-tune the final signal plans that will be installed in the junctions' controllers.
4. **Required number of simulation runs**: In order to gain results with statistic confidence, a single simulation run is not sufficient. The standard deviation of major parameters is required to estimate the number of repetitions. Based on a pre-test, the standard deviation is initially estimated and calculated as below (FGSV 2006).

$$s^2 = \frac{\sum(x - \bar{x})^2}{n-1} \tag{4.1}$$

where:

s = *standard deviation*

x = *variable (such as delay, travel time) for which the sample variance is desired*

$\bar{x}$ = *average value of the variable produced by the model runs*

n = *number of model runs*

Upon determination of the variable's standard deviation, a desired confidence level and confidence interval are selected. The confidence level is the probability that the true mean lies within the target confidence. The confidence interval is the range of values within which the true mean value may lie (FHWA 2004). The usual approach is to pick a 95-percent or 97.5-percent level of confidence requiring a higher number of simulation runs. The selection of the confidence interval depends on the variable to be tested and on the purpose for which the results will be used. The required number of repetitions is finally computed using the following formula (FGSV 2006).

$$n_{req} \geq \frac{t(\propto, n-1)^2 \cdot s^2}{e^2} \tag{4.2}$$

where:

n_{req} = *required number of simulation runs*

t(α*,n-1)* = *Student's t-statistic for the probability of a one-sided error summing to confidence level* α *where n equals the number of repetitions*

s = *standard deviation*

e = *confidence interval*

5. **Simulation runs for L4G calibration** is necessary since it relies on a learning mechanism. Before starting evaluation of the whole system, the control logic should be prepared and auto-adapted to the concerned junctions and signal configuration. This is done by running a certain number of simulations by varying stochastically the demand within a defined range, and without reinitializing the control parameters so that the fine-tuning process can build on previous runs.

After completing the aforementioned steps, the re-calibrated and adopted model can be used to simulate the TSoS Traffic Use Case.

Purpose of the study is to demonstrate functionality, operational safety and performance of the Local4Global system. Assessment objectives as defined in previous sections are described by KPIs. Figure 8 shows that Traffic Performance Index, Level of Service and Network Mean Speed are outcomes of the microscopic simulation study. All KPIs are described in detail starting from section 4.3.6 of this document.

4.3.3.2 Field Tests

Field tests are carried out to collect data providing the input for the evaluation of the performance of the Local4Global system in the test bed. Two types of field tests are planned to be carried out. One-vehicle test drives are performed in order to determine objectively the quality of traffic control on the basis of acknowledged key figures. A fleet test aims to introduce the Local4Global system to a wide range of users using intensively Local4Global applications on a voluntary basis willing to contribute to the evaluation of the system.

4.3.3.2.1 One-Vehicle Test Drives

The objective of one-vehicle test drives is to collect data for the quality assessment of traffic control at intersections and sections. GPS data from test drives is used e.g. to determine the quality level (LOS) of the Local4Global system compared to the traditional coordinated traffic control being currently under operation. Average waiting times are the measure to determine LOS for uncoordinated intersection branches.

The execution of one vehicle test drives for data collection follows the principles of Module N2 of the German Guideline for Quality Management for Traffic Signal Control (Friedrich, et al. 2008). Test vehicles cruising in the test bed are equipped with GPS data loggers and collect information on current position, heading and time stamp once a second. The travel speed is derived from distances between positioning points and time intervals between logs. In order to be able to obtain an objective view on traffic quality, drivers of test vehicles carrying out test drives are asked to adapt their speed to the traffic flow and “swim” in platoons.

Similar to the required number of simulation runs as described in context of the simulation study, the variance of a sample (e.g. travel time or travel speed per direction in the test bed) defines the required number of test drives in terms of grounding the outcomes on a sufficient level of statistical confidence. The required sample size n is calculated according to following equation:

$$n \geq \frac{t^2 \cdot s^2}{{e_r}^2 \cdot x^2} \qquad (4.3)$$

where:

n	*= required sample size for test run*
t	*= Student's t-distribution, two-sided test at required statistical accuracy (e.g. 0.95) and degrees of freedom (= number of test runs carried out in pre-test)*
s^2	*= variance derived from pre-test*
e_r	*= permitted relative deviation (e.g. 0.10)*
x	*= mean from pre-test*

The pre-test generally requires 5-10 test values in order to be able to estimate the variable's deviation.

4.3.3.2.2 Fleet test

The intended fleet test is classified as an experimental field test. Experimental field tests are characterized by an enhanced external validity (Bortz und Döring 2006). Thus, the results can be generalized to other conditions. Field test have naturally a high ecological validity, because they are carried out in the "real-world" and not in a simulation or laboratory. On the other hand, the influence to control interfering variables (e.g. weather conditions or interaction between cars) is often reduced or even not possible. Sometimes, test subjects are preselected or filtered in pre-tests, according to the subject of the experiment. (e.g., the test population should resemble a worst case scenario), but most often the test subjects are chosen to compose a typical user group of a product (e.g. conform to a given distribution of age).

Two terms that are closely connected to field tests in the automobile sector are: Field Operational Tests (FOT) and Naturalistic Driving Studies (NDS).

Field Operational Test: "A study undertaken to evaluate a function or functions under normal operating conditions in environments typically encountered by the host vehicle(s) using quasi-experimental methods." (FOT-NET 2011)

Naturalistic Driving Study: "Naturalistic Driving observation refers to studies undertaken using unobtrusive observation when driving in a natural setting." (http://wiki.fot-net.eu)

Thus, NDS does not fit to characterize the fleet test because the information by the traffic light information system on the smartphone are intended to change the "natural setting".

The fleet test is best characterized by the term FOT. Nevertheless, FOTs are often connected to products or technologies that are already available or close to market entrance. The purpose of the project is to test a System-of-System approach in two use cases (building and traffic). Thus, even the term FOT does not fully fit.

The test subjects will not be intentionally preselected. There is a kind of filtering due to the facts that the application will be implemented for Android (biggest market share) and the focus is on light vehicle drivers. Another big constraint is, that a valuable test subject should use the road (best on a regular basis), when the system is operating.

The upper limit for test subjects is estimated during the project by a load test on the server architecture, which provides the traffic light information to the smartphones and gather the data from the floating vehicles. Thus, it can be judged how many users can be managed at the same time by the servers.

In four former experiments, e.g. the analysis of reduced speeding behavior with and without the system (Krause, Yilmaz und Bengler, Comparison of Real and Simulated Driving for a Static Driving Simulator 2014) with the traffic lights assistant on this road section, the sample size was typically around 20 resulting in significant outcomes. Thus, N=20 will be assumed as lower limit for human factors in this project.

There is also a differentiation of the sample size and user group fluctuation for different purposes. For FCD collection and as input for the optimization algorithms in general, it would be desirable to have a constant group as large as possible. To judge about human factors like usability and acceptance, even one-time-users and drop-outs (and their reasons) can be of more interest than continuous users.

The Human Machine Interface (HMI) was extensively tested in the KOLIBRI project[3] in driving simulator and field trials, e.g. regarding visual (Krause, Knott und Bengler 2013) and cognitive demands and showed suitability while driving on this track. The HMI is used as before. If (small) changes are necessary due to project requirements, an expert judgment will be made with respect to driver distraction issues. However, this project includes no further objective driver distraction assessment e.g. with eye-tracking.

The test subjects need to follow a registration procedure subscribing with their email address to use the system. This assures that even one-time users can be asked for their opinion. The used questionnaires preferably incorporate the System Usability Scale (SUS), (Brooke 1996) and the AttrakDiff2 (Hassenzahl 2003). These were used several times in the KOLIBRI project. Thus, comparison would be enabled.

The SUS is a wide-spread, short and easy to administer questionnaire. The ten items are statements that have to be judged on a five point Likert scale. The overall score ranges between 0 and 100 and represent a value for the overall usability. Bangor et al. (2009) connected the outcomes to an adjective rating scale. Their results also give indication how to interpret the values: In general an acceptable system should get a SUS score above 70. An overview of the SUS values of KOLIBRI project and a short SUS literature review can be found in Krause und Bengler (2013). An example of a KOLIBRI SUS score on the real road is 86 (excellent) (Krause, Knott und Bengler 2013)

The AttrakDiff2 has 28 items, each rated on a seven point Likert scale. It consists of four sub dimensions: pragmatic quality (PQ), hedonic quality stimulation (HQ-S) and identification (HQ-I) as well as attractiveness (ATT). The results are typically communicated in a portfolio diagram (cf. Figure 11). For the portfolio HQ-S and HQ-I are combined into one axis. The results are interpreted regarding their position, together with the confidence rectangles (95%). The ATT is typically communicated separately as number. The outcomes of a test with KOLIBRI on the real road, revealed results that could be connected to "desired" (Krause, Weichelt und Bengler 2015).

For methodological reasons it would be an advantage to get the baseline driving behavior of a test subject on the track, without the traffic light information on the smartphone. Therefore, the app will include a function to track the driving, without showing any traffic light information. Solid continuous users, will be contacted and asked, if they could enable the option for their next ride.

Inspired by the of the approach from DeLone & McLean (2003) aiming to describe the success of a system by analyzing system quality and information quality, a simplified determination of the fulfillment of assessment objectives in terms of functionalities from a user's point of view is planned to be carried out. Participants in the fleet tests will be addressees of questionnaires being asked to assess aspects of system quality like availability of functionalities, usability, stability, Willingness To Pay (WTP) for the app etc. as well as aspects of information quality like accuracy, reliability and timeliness by means of checklists and ordinal scale. The latter aspects can also be quantified and compared to objective metrics by calculating e.g. the difference between predicted and actual signal timing described e.g. via Root Mean Square Percent Error (RMSPE).

---

[3] KOLIBRI: Cooperative Traffic Control Outside Urban Areas, Research project funded by the Bavarian research foundation, 2011-2013

Figure 15 illustrates relevant aspects covered by questionnaires in terms of system quality, information quality and attractiveness in the view of assessment objectives linked to suitable assessment methods.

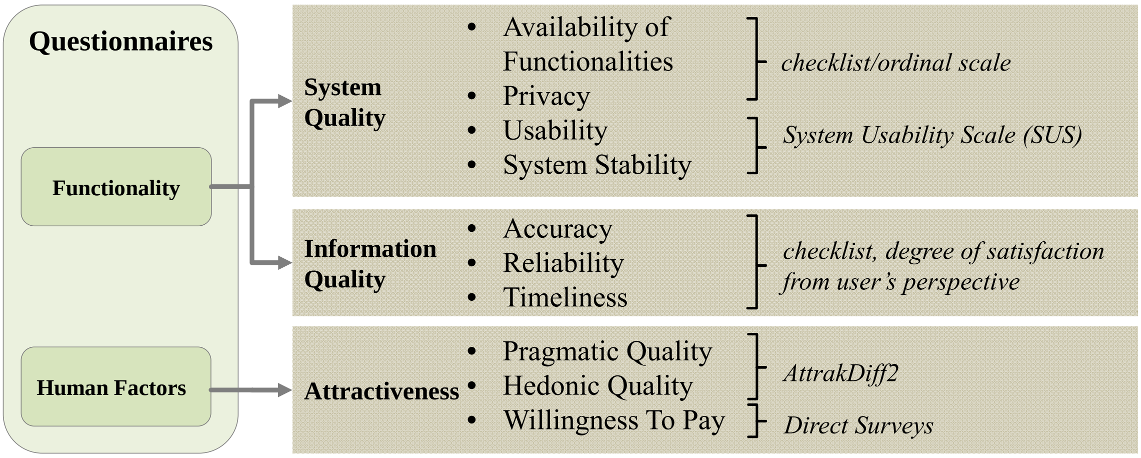

Figure 15: Assessment Objectives and Contents of Questionnaires

4.3.3.3 Profitability analysis

The profitability of the Local4Global system is going to be assessed by means of a Cost-Benefit-Analysis (CBA). The CBA represents an economical assessment method intending to maximize social welfare. In this context, the methodology is often used to evaluate an entire strategy with respect to the impacts of investments on the benefits of the group on whose behalf the analysis is carried out. It is the main objective of a CBA to identify configurations which can improve the level of overall welfare (Commission of the European Communities 1991).

The analysis covers different kinds of impacts and different groups affected, but tries to combine all of these impacts into a single measure. The interest groups generally affected by projects are producers, consumers and third parties. The costs of producers (i.e. companies or public administrations) may be directly measured in monetary units. For consumers and third party groups the project's outcomes may either result in a financial advantage or disadvantage depending on whether travel costs increase or decrease. It is common however to value the project impact on the quality of life that is mainly affected by it.

A cost-benefit calculation technique often used is the "Benefit-Cost Ratio". The comparison of use case related costs and benefits as far as they can be determined lead to a KPI describing the system's economic performance.

4.3.4 Uncertainties and artificial circumstances

Uncertainties and artificial circumstances are mainly related to field tests. They are subject of the risk assessment and contingency plan documented in Deliverable D5.1.1 (Ettinger und Krause 2014).

4.3.5 Data Base

The traffic control methods implemented in the test bed during the KOLIBRI project define the reference cases for the evaluation of the Local4Global system. The availability of wide data material from the KOLIBRI project enables the comparison of impacts of the Local4Global System's control strategies to the results of the uncoordinated, fixed time coordinated and traffic actuated control evaluated within KOLIBRI. Available material consists of

- findings of a traffic analysis carried out in the frame of the project KOLIBRI (2011-2013) based on one-vehicle drives quantifying measures like travel times, waiting times, number of stops at intersections, quality of traffic flow in different control setups (uncoordinated, fixed time coordination, traffic actuated coordination).
- user questionnaires from KOLIBRI project (2011-2013)

- SUS scores from five experiments
  - AttrakDiff2 values from four experiments
  - individual willingness to pay for the app, from an online survey and from two field trials
- driving data from usability tests within KOLIBRI project (2011-2013)
  - from three driving simulator experiments
  - from four real road experiments

  (Typically around N=20 participants)

The KOLIBRI experiments with test subjects on the real road were mainly carried out during fixed time coordination periods. The last subject test of KOLIBRI was managed during a traffic actuated coordinated period and also included an experimental aspect with (intended) malfunction of the traffic lights information system on the smartphone.

4.3.6 KPI 1: Performance Index

The Performance Index expresses the overall traffic quality in a road network based on measures like stops at traffic signals or waiting times. For evaluation purposes the criteria are not regarded as isolated parameters. They are weighted and included in a so-called Performance Index (PI) e.g. according to Brilon et al. (2007) based on the findings of Robertson (1969) in the context of the traffic control tool TRANSYT. The PI represents the weighted average number of stops and waiting time and is calculated according to the equations below.

$$PI = \left(G_W \cdot \sum_i \sum_z \left(W_{i,z} \cdot g_i \cdot g_z \right) + G_H \cdot \sum_i \sum_z \left(H_{i,z} \cdot g_i \cdot g_z \right) \right) \quad (4.4)$$

$$PI_R = \frac{PI}{\sum_i \sum_z \left(Q_{i,z} \cdot g_z \right)} \quad (4.5)$$

where:

PI = *performance index*

PI_R = *relative performance index*

G_W = *weight of waiting times*

$W_{i,z}$ = *sum of waiting times per hour for vehicles of type z on intersection branch i*

g_i = *weight of route i*

g_z = *weight for vehicles of type z*

G_H = *weight for number of stops*

$H_{i,z}$ = *sum of number of stops per hour for vehicles of type z on intersection branch i*

$Q_{i,z}$ = *traffic volumes of vehicles of type z on intersection branch i*

i = *route index*

z = *vehicle index*

The parameters "number of stops" and "waiting time" can be determined by means of test drives or microscopic traffic simulation. The determination of weights follows published examples or predefined numbers based on local conditions (Brilon, Wietholt and Wu 2007). The factor g_i can be used to put more weight on the main direction if needed. It can also be assumed that relative to the energy consume, a stop represents a waiting time of 60 seconds. A lower PI stands for higher traffic quality. The value of the PI has no significance by its own and is only relevant for comparison reasons. For evaluation purposes, the last calculation step is normalizing the performance index due to the traffic volumes.

4.3.7 KPI 2: Level of Service

The KPI of Level of service (LOS) is used to analyze roads by categorizing traffic quality. It is a qualitative measure describing the operational conditions for a traffic stream. The assessment is based on service measures such as travel time, waiting time, travel speed or coordination measure. The differentiation into six different service levels is both used in the American standard *Highway Capacity Manual HCM* (FHWA 2000) and in the German equivalent *Handbuch für die Bemessung von Straßenverkehrsanlagen HBS* (FGSV 2001). The categories are defined by letters from A to F. Each letter stands for a state of flow. The definitions are based on the *Highway Capacity Manual*.

LOS A describes free flow operations. The vehicles are unimpeded in their ability to maneuver within the traffic stream.

LOS B describes a reasonably free flow. The ability to maneuver within traffic stream is slightly restricted.

LOS C describes a stable flow of traffic and stable operations. The ability to maneuver and change lanes is more restricted than at LOS B, which may contribute to lower average travel speeds.

LOS D is approaching unstable flow. The freedom to maneuver within the traffic stream is noticeably limited and even minor incidents can create queuing.

LOS E is the highest operational density level. It describes operation at capacity. The ability to maneuver at this Level is minimal due to closely spaced vehicles and lack of space for the traffic stream to absorb disruptions. Any incident can create a serious breakdown.

LOS F describes a breakdown in vehicular flow. Queues are forming and the maneuverability within the traffic stream is down to zero.

The German standard HBS distinguishes between two cases for the quality assessment of signal traffic control: coordinated traffic signal control and uncoordinated traffic signal control.

In order to evaluate a road section with coordinated traffic, it is necessary to establish the coordination measure k. The value of k is expressed in percent and represents the rate of signal pass through, i.e. passing a traffic light without stopping. A higher value stands for a better coordination and higher traffic quality.

$$k_i = \frac{D_i}{(N_k - 1) \cdot n} \cdot 100 \tag{4.6}$$

$$k = \frac{k_1 + k_2}{2} \tag{4.7}$$

mit:

k_i = *coordination measure traffic direction i [%]*

k = *coordination measure for the road section [%]*

D_i = *number of signal pass through at signalized intersection for each traffic direction i [-] (except starting intersection)*

N_K = *number of signalized intersections for the road section [-]*

n = *number of test drives [-]*

For uncoordinated traffic control the relevant characteristic is the average waiting time. It represents the average time in seconds to wait for a signal to turn green on a road section. Table 3 summarizes the LOS definitions in the context of coordinated and uncoordinated traffic signal control.

Table 3: Level of Service Definitions (FGSV 2001)

LOS	coordination measure [%]	allowed average waiting time [s]	description according to HCM (FHWA 2008)
A	≥ 95	≤ 20	free flow
B	≥ 85	≤ 35	reasonably free flow
C	≥ 75	≤ 50	stable flow
D	≥ 65	≤ 70	approaching unstable flow
E	≥ 50	≤ 100	unstable flow
F	< 50	> 100	breakdown flow

4.3.8 KPI 3: Average Network Mean Speed

The Average Network Mean Speed (ANMS) is a suitable indicator for determining the performance of a traffic control system. A quality assessment on the basis of the ANMS can be carried e.g. by comparing two systems, provided that sufficient data is available for both setups when comparing e.g. the Local4Global system to the baseline control from the KOLIBRI project. Furthermore, the KPI is suitable for measuring the system's performance over time. This approach is in particular suitable if the evolutionary capabilities of a system shall be assessed as in the case of the Local4Global product. For a net-wide analysis, the microscopic traffic simulation is considered to be an appropriate tool as it allows the integration of all vehicles in the modelled network.

4.3.9 KPI 4: Cost-Benefit Ratio

The CBA's KPI considered beneficial for the TSoS Traffic Use Case is the widely applied *Benefit-Cost Ratio*. For public investments, the ratio between costs and benefits typically needs to be > 1. In this particular case, it shows the value of traffic-related benefits over costs. Mainly the following indicators for costs and benefits are considered:

Indicators for costs:

- **Investment costs** for Local4Global system (infrastructure and equipment) e.g. installation costs for server infrastructure, detection equipment, devices, traffic control facilities (traffic controllers, communication facilities)
- **Maintenance costs** for Local4Global system (running costs) for data transmission and communication, data processing and storing, operation of traffic computer and/or computing centers

Indicators for benefits are e.g. reductions in:

- **Travel times** incl. monetary assessment e.g. on the basis of the German guideline "*Empfehlungen für Wirtschaftlichkeitsuntersuchungen an Straßen*" (FGSV 1997) indicating time cost rates for passenger cars and HGV
- **Fuel consumption** including monetary valuation on the basis of fuel prices
- **Externalities** indicating benefits resulting from a reduction of emissions from road traffic due to e.g. avoiding stops at signalized intersections.

4.3.10 KPI 5: Subjective Usability / User Experience

For the subjective ratings, the two standardized questionnaires

- **System Usability Scale** (SUS), overall usability rating and
- **AttrakDiff2** user-experience

along with project specific questions (e.g. willingness to pay) are used.

The results are related with literature values and scores from KOLIBRI project. Also, changes of the SUS and AttrakDiff2 over time (repeated use) are of special interest. As well as, reasons why users stopped, to use the app. Additionally, the usage frequency of the app is graphically assessed (e.g., histogram).

4.3.11 KPI 6: Driving Behavior / Compliance

To evaluate and characterize the driving behavior the following metrics will be used (whenever possible related to baseline driving data from Local4Global and/or data from KOLIBRI):

Compliance/behavior to speed recommendation:

- Percentage of time within the traffic light assistant speed recommendation.
- Traffic light approach (speed, acceleration/deceleration), if the system shows "You will arrive at red"

Safe Driving:

- Percentage of distance travelled above the speed limit (with tolerance level; speed limit +5km/h)
- Count of severe speed violations (e.g., > 15km/h)
- Characterization of speeding excess (e.g., Root Mean Square speed above speed limit)

4.4 Comparison baseline

Evaluation of the new Local4Global control strategy would rely on a comparison with a baseline. The baseline describes how good the performance of the traffic control was before the L4G was active.

Outcomes of the KOLIBRI project will be used as a baseline, as data and analysis are available from this period regarding travel times, waiting times at intersection, number of stops and LOS at intersection.

In this paragraph, the conditions of the KOLIBRI experiments are first described, in order to show the main differences compared to Local4Global. After that general outcomes are presented. These results might be detailed in comparison to other equivalent outcomes of Local4Global in the next evaluation deliverables.

4.4.1 Experimental conditions of the baseline

4.4.2 General description

The KOLIBRI project aimed to improve traffic performance of one particular direction, which was the main stream along B13, a suburban 2-lane road near Munich. This is remarkable difference to Local4Global, where the improvement effort is addressed to local optimization independently of the direction.

The project had two main contributions to the existing situations: first provide an optimal "human made" green wave coordination to the road section using advanced calculation and calibration methods. Second provide cooperative drivers with speed recommendations so that the overall traffic performance is improved with action on both actors.

4.4.3 Scenarios comparison

To evaluate the proposed contributions of KOLIBRI, experiments have been conducted on 3 configurations:

- Uncoordinated traffic-actuated (UTA) signaling, no cooperative vehicles
- Coordinated fixed-time (CFT) signaling with cooperative vehicles
- Coordinated traffic-actuated (CTA) signaling with cooperative vehicles

These scenarios have been compared according to the following indicators:

- Quality of coordination (% of no-stop drives through the road section)
- Number of stops per drive
- Wait time / congestion length
- Travel time

4.4.2 General results of the baseline

All results have been reported for both directions of the test bed main stream (north>south and south>north) and are listed in the Annex.

The KPI 1 (Performance index) has been used in KOLIBRI to compare variations in simulation environment. The outcomes are listed in the table below.	**CTA**	**UTA**
PI main stream	19249	16515 (-14,2 %)
PI transversal streams	25510	29 134 (+14,2 %)
PI total	20969	19 984 (-4,7 %)

4.5 Transfer to General KPI

The Local4Global System intends to provide generic methods applicable to a variety of technical applications. Therefore, the transferability of KPIs to other use cases is interesting in terms of describing the generic outcomes of the Local4Global system. KPIs like LOS or ANMS are solely traffic related, i.e. they are established measures for assessing the quality of the Local4Global system in the context of traffic. The PI as introduced in section 4.3.6 is based on traffic related parameters like travel times and waiting times. The impacts of a traffic related application of the Local4Global System can be expressed in e.g. percentage changes of the PI. Although not mentioned in the DoW, a variation of the PI, the Cost-Benefit Ratio provides a suitable measure for comparing different use cases of the Local4Global system. When it comes to the assessment of L4G generic methods, for each GKPI (cf. section 3.2.1) equivalent figures are chosen as KPI that facilitate a direct transfer to GKPI and comparison above use case boundaries. See Table 7 for GKPI-KPI-correlation in traffic use case.

Table 4: GKPI-KPI transfer relation

GKPI	KPI	Description	Target Improvement
GKPI 1	ANMS	Difference in terms of daily, average network speed	30 % (**according to DoW**)
GKPI 2	ANMS under artificial circumstances	In both use cases equal to GKPI 1, but when major incidents are present	30 % (**according to DoW**)
GKPI 3	SuS / AttrakDiff2	System usability rating an and acceptance to be calculated through questionnaires	equivalent to findings from KOLIBRI (no deterioration)