SUNSET

Sustainable Social Network Services for Transport
www.sunset-project.eu

Grant agreement n°: 270228
Start date: Feb 1, 2011
Duration: 36 months
Area: ICT for Transport
Project Officer: Mr. Stefanos Gouvras

Deliverable D6.1
Evaluation approach for operational success and effectiveness of incentives

Version: revised final
Due date of deliverable: Nov 30, 2012
Actual submission date: July 17, 2013
Dissemination level: PU
Responsible partner: UTWENTE

The research leading to these results has received funding from the European Union Seventh Framework Programme (FP7/2007-2013) under grant agreement n° 270228. The project's website is at www.sunset-project.eu.

Summary

This deliverable aims to present a methodological approach for evaluating the operational success and effectiveness of incentives in SUNSET. It takes as input the SUNSET deliverables D3.1 and D3.2 on the individuals' travel goals, D3.3 and D3.4 on feasible and potentially successful incentives. The output of this deliverable, together with D6.2, will be used in the evaluation of the SUNSET Living Labs (LLs), leading to D7.2, D7.3, D7.4, and D7.5.

The **operational success** of the SUNSET system, as agreed within the consortium, relates to the ability of the whole system to deliver against four criteria:

- Enabling individuals to meet their travel objectives (e.g. to reduce travel time)
- Creating synergy through the use of social media and social networks
- The system functioning as intended (e.g. detection accuracy and battery life)
- User friendliness of the system (e.g. the system's learnability, user satisfaction)

To understand *how well the tripzoom app can help individual travellers achieve their desired travel goals*, personal objectives are firstly identified. Tripzoom users will be asked for their perception on the extent to which the tripzoom app helps in obtaining those goals, using various qualitative techniques. A set of **subjective indicators** proposed as Key Performance Indicators for **operational success** is:

- travel time, distance and cost by trip purpose and distance
- scheduling effort by trip purpose and distance
- experience with social networks (e.g. motivational support, feedback, and satisfaction)
- output (the added value of tripzoom app in relation to each of the desired mobility)

In order to evaluate the operational success based on the above mentioned KPIs a number of prerequisites, i.e. additional indicators are required to scrutinize and explain the results based on groups of individuals. These are:

- Indicators on socio-demographics and household characteristics
- Personality, lifestyle, desired mobility
- Preferred mode and type, accessibility to modes, mobility constraints
- Number of trips by trip purposes
- Total travel time by trip purpose and distance
- Objective indicators on social networks
- Subjective indicator of travel time by trip purpose and distance
- Subjective indicators of social networks
- Attributes related to travel dislike, freedom, and stress

The second criterion of operational success focuses on indicators suitable to evaluate the use of *social media related aspects* in the SUNSET system. Measurements are taken by the app itself and using Google Analytics, and/or Radian6, in accessing success based on the following indicators:

- Number of unique visitors to the web portal
- Number of registrants, LL participants
- Number/percentage of participants recruited using LL-Facebook, via friends on external social networks, via friends invited by email
- Size and activeness of the tripzoom social network

The third and fourth criteria of success are related to the evaluation of the tripzoom app regarding *functionality* and *usability*. It is recommended that a design approach is applied where the design process proceeds in iterations. End-users are involved during every stage of development, providing feedback on concepts and prototypes regarding functionality and usability. In order to reveal user needs and evaluate concepts at an early stage of development, feedback is gathered in two ways: quantitative user data from large user groups using surveys; qualitative user data from smaller groups applying focus-groups and interviews. During development, evaluation should be conducted with several iterations in the following ways:

- Ad-hoc functionality testing by developers
- Functionality testing by non-developers (in the role of potential users of the system) whilst executing a structured task list, whereby developers are notified of bugs and functions that are not working properly
- Usability testing by non-developers in the role of usability experts and executing the same structured task list, using evaluation techniques such as heuristic evaluation and cognitive walkthrough
- Functionality and usability evaluation with real end-users, by applying evaluation techniques such as think-aloud protocol, interviews, focus groups, field testing, experience sampling, and surveys

To evaluate the **effectiveness of the SUNSET incentives** (as designed in D3.3), both changes in behaviour and changes in attitude should be taken into account. *Behavioural indicators* are designed based on the mobility pattern of an individual as detected by tripzoom. They are quantified by the distribution of trip measurements over trip characteristics. Key performance indicators are

- number of trips
- distance travelled
- travel time
- travel cost
- CO2 emission

Where all trips are characterized by mode, purpose, timing, route and location.

Electronic questionnaires and experience sampling will be used to collect data on user attitude. Key performance Indicators for measuring the travellers' **attitude change** due to incentives are:

- Awareness of (the impact of) the personal mobility pattern;
- Awareness of the existence and/or performance of alternatives (modes, routes, etc.);
- Awareness of the societal impact of traffic (externalities);
- Self-categorisation, preferences, satisfaction, social attitude.

Hypothesis testing and longitudinal data analysis techniques are employed to evaluate the effectiveness of incentives. For this purpose an experimental design is proposed where the LL participants are divided into seven groups, with each group receiving a specific combination of incentives over the LL periods. An incentive is expected to affect only certain types of trips (e.g. commuting trips using the car mode) and certain aspects of user attitude. Therefore it is recommended that the LL coordinator should only ask the most relevant experience sampling questions, and the data analysis should focus on the categories of trips that are most likely to be affected by the incentives offered.

Several other issues also need to be dealt with when evaluating the effectiveness of incentives through the LL experiment:

- Sample size: it is recommended that each participant group should have at least 40 users. Whilst all seven groups will be used in the main LL (Enschede), the number of groups may be fewer in the reference LLs (and the total number of participants recruited may also therefore be lower). During the group formation stage of the experiment, it is proposed that the LL coordinator should follow a prioritization plan, in order to ensure that at least the most important groups contain a sufficient sample size. The priority of the groups is determined according to the need to answer particular research questions in each LL reflecting, for example, the local transport priorities.
- Missing data: it is recommended that a reward or compensation should be provided for participation in the LL, so that users are more motivated to participate, to manually correct their travel data, and to answer questionnaires.

Document Information

Authors

Name	Partner	Email
Diana Kusumastuti	UTWENTE	d.kusumastuti@utwnte.nl
Jing Bie	UTWENTE	j.bie@utwente.nl
Sander Veenstra	UTWENTE	s.a.veenstra@utwente.nl
Susan Grant-Muller	UNIVLEEDS	s.m.grant-muller@its.leeds.ac.uk
Nikolaos Thomopoulos	UNIVLEEDS	tranth@leeds.ac.uk
Erik Klok	ENSCHEDE	e.klok@enschede.nl
Ynze van Houten	NOVAY	ynze.vanhouten@novay.nl
Eric van Berkum	UTWENTE	e.c.vanberkum@utwente.nl

Editor

Name	Eric van Berkum
Partner	UTWENTE
Address	University of Twente, Postbus 217, NL-7500AE Enschede
Phone	+31 (0)53 489 3821
Fax	+31 (0)53 489 4040
Email	e.c.vanberkum@utwente.nl

History

Version	Date	Changes
V0.1	05-09-2012	First outline
V0.2	17-09-2012	Final outline
V0.5	10-10-2012	First preliminary draft
V0.6	29-10-2012	Combining first draft, updated
V0.8	11-11-2012	Combining second draft
V0.81	14-11-2012	Further update by Nikos, Erik, Ynze & Jing
V0.82	15-11-2012	Updated second draft
V0.85	16-11-2012	Revised final draft, missing summary & conclusions
V0.90	18-11-2012	Full version for internal review
V0.99	28-11-2012	Final draft, revised after internal review
V1.0	30-11-2012	**Final version, approved by PCPs and PMT, sent to EC**
V1.1	14-07-2013	**Revised version after review**

Distribution

Date	Recipients	Email
17-07-2013	SUNSET partners	sunset@lists.novay.nl
17-07-2013	Project Officer	Stefanos.GOUVRAS@ec.europa.eu
17-07-2013	Project Archive	INFSO-ICT-270228@ec.europa.eu

Table of Content

List of Figures

List of Tables

1. Introduction

In the broader sense, the SUNSET project aims to encourage sustainability in the transport system and to support people's mobility. These goals are articulated as more specific objectives – specifically to reduce traffic congestion and CO2 emissions, and to improve personal safety and well-being. To support delivery of these project goals, the "tripzoom" application (app) has been developed for use with smartphone (on the iOS and Android platforms). The app has a range of functionality, but can be used as a channel through which various types of incentives can be offered. The incentives may encourage people to make smarter travel choices and to travel in more sustainable and health-promoting ways (e.g. by cycling or walking more often). The tripzoom app will be fully operational and tested in three living labs (LLs), namely: Enschede (NL), Leeds (UK), and Gothenburg (SE), with Enschede being the main LL and Leeds and Gothenburg being the reference LLs. During these LL trials various incentives will be tested, which may be tailored according to the design of the individual LL and local transport priorities.

The tripzoom app is not the only output from SUNSET, as the concept has also required the development of a 'city dashboard' for the administration of incentives and a web portal for the administration of new and existing users. These two components are prerequisite for a fully functional tripzoom service. The development of the various work strands is reported in deliverables; they can be found on the SUNSET general information site: http://sunset-project.eu/. In SUNSET, the city dashboard has been specifically developed for use in the project to enable the offering of incentives in the living lab (LL) operations. In a non-experimental context, the city transport operators will need a similar tool to issue their own incentives. This should allow the selection of particular traveller groups from the set of participating citizens and the ability to time-schedule when incentives are offered. The web portal functions as a landing page, to provide information and support user registration.

Within SUNSET, Work Package 6 (WP6) is responsible for providing a general assessment framework. This includes providing a methodological approach, together with recommendations on indicators (or measurements) for the criteria. Specifically, WP6 contributes to the project through two deliverables:

- D6.1 (T6.1 and T6.4): Evaluation approach for operational success and effectiveness of incentives;
- D6.2 (T6.2 and T6.3): Evaluation methodology and measurement approach.

To summarise the relationship between the two deliverables, D6.1 takes a 'bottom up' approach concerning operational success and individual responses to a range of incentives, whereas D6.2 provides a system level evaluation which interfaces with the business case and the bottom up recommendations from D6.1.

The research reported in this deliverable (D6.1) takes inputs from several other SUNSET deliverables, as indicated below:

- T3.1/D3.1 and T3.2/D3.2 on individuals' travel goals
  D6.1 will use the individuals' travel goals defined in D3.1 and D3.2 to derive indicators needed to assess how well the system meet travellers' personal goals.

- T3.3/D3.3 on potential incentives

D6.1 will use the proposed list of potential incentives suggested by D3.3. This means that D6.1 will give suggestions on methods of analysis which are suitable to assess the effectiveness of all the potential incentives, previously identified in D3.3.

- T3.3/D3.4 on the final design of incentives (forthcoming)
  D6.1 will use the input from T3.3, specifically on D3.4 which is responsible for the final design of incentives. This means that design parameters, such as the frequency of offering incentives, will be taken into account in the indicators and methods of analysis.

The research in this deliverable will also be used by other tasks. Together with D6.2 it will provide instruments that can be used in the evaluation of the living labs (T7.4/D7.5) with the experimental data generated by lab participants in Enschede, Leeds and Gothenburg.

The various SUNSET deliverables that this deliverable is related to are listed in Table 1.1.

Table 1.1 Related SUNSET deliverables

Title of deliverable	Referred to in this deliverable as
D1.1 Preliminary scenarios and user and system requirements	D1.1
D1.2 Revised scenarios and user and system requirements	D1.2
D2.1 Mobility sensing and experience sampling services	D2.1
D2.2 Mobility pattern detection and visualisation service	D2.2
D3.1 Objectives	D3.1
D3.2 Individual objectives versus system objectives	D3.2
D3.3 Impact of Incentives	D3.3
D3.4 Feasible and potentially successful incentives	D3.4
D4.5 Final mobile application design	D4.5
D5.2 System integration	D5.2
D6.2 Evaluation methodology and measurement approach	D6.2
D7.1 Living lab plan	D7.1
D7.2 Living lab report Enschede	D7.2
D7.3 Reference Living lab report Leeds	D7.3
D7.4 Reference Living lab report Gothenburg	D7.4
D7.5 Evaluation report	D7.5

In the remainder of this chapter, the goals of this deliverable are given in §1.1 and the main contributions and innovations of D6.1 will be presented in §1.2. Our general approach to deriving the methods and indicators will be summarised in §1.3. Finally, the structure of the overall document will be explained in §1.4.

1.1 Goals

The overall objectives of WP6 are as follows:

1) To provide a set of key indicators that allow evaluation of the implementation and operational success of the social traffic scheme (success will be measured by a combination of mobility efficiency and sustainability indicators);
2) To outline a general framework to evaluate the SUNSET system in against broad EU objectives for improved mobility in the future, including objectives relating to efficiency, sustainability and society;
3) To provide specific recommendations to the living lab experiments on the indicators and measurement approach for the analysis of case study data in assessing the achievement of objectives;

4) To outline an analysis approach for the effectiveness of the use of incentives in the SUNSET system.

D6.1, contributes to the first and the fourth goals above, which lead to the following objectives:

- **To provide indicators and methods to evaluate operational success**
  - Objective 1: To provide indicators to evaluate how well the system meets individuals' travel goals;
  - Objective 2: To provide indicators to evaluate how well the social media concept is implemented in the project;
  - Objective 3: To provide indicators and to propose methods to evaluate the tripzoom app during the development and design stages;
  - Objective 4: To provide indicators and to propose methods to evaluate the tripzoom app from the point of view of users, based on users' experience. This work will be done as a follow-up study to assess tripzoom app during the course of LLs.

- **To provide indicators and methods to evaluate the effectiveness of incentives**
  - Objective 5: To provide indicators to evaluate travel behavioural changes;
  - Objective 6: To provide methods to evaluate changes in indicators of travel behaviour;
  - Objective 7: To provide indications on data requirements (in relation to the Objectives 5 and 6) and data gathering techniques.

The research reports in D6.1 and D6.2 are interrelated, with D6.2 providing an overall framework for the assessment of operational success. Therefore, resulted indicators that address Objectives 1 and 2 will feed into an overall evaluation of the project.

1.2 Main results and innovations

The main results of D6.1 are:

- An assessment approach to reflect how the system meets individuals' travel goals;
- An assessment approach for the use of social media concept in SUNSET;
- An assessment approach for the functionality of the tripzoom app during the development and design stages;
- An assessment approach for the evaluation of the tripzoom app with respect to user experience;
- An assessment approach for travel behavioural changes and attitude changes in response to incentives.

It is expected that these assessment approaches can be used in both the main LL and the reference LLs in SUNSET, as well as adapted for external use by other projects. Within SUNSET, the proposed methods and tools will be used by D7.2, D7.3, D7.4, and D7.5 to evaluate the success of the project during/after the course of the LL operations.

The innovations of this deliverable are summarised in Table 1.2.

Table 1.2 Contributions of this deliverable to SUNSET innovations

SUNSET innovations	Contribution of this deliverable
Social mobility services that motivate people to travel more sustainably in urban areas	N/A
Intelligent distribution of incentives	N/A

(rewards) to balance system and personal goals	
Algorithms for calculating personal mobility patterns using info from mobile and infrastructure sensors	N/A
Evaluation methodologies and impact analysis based on Living Lab evaluations	(a) recommended methods and indicators for the assessment of the extent to which the system meets individuals' goals (b) recommended methods and indicators for the assessment of social media and social network aspects (c) recommended methods and indicators for the assessment of application development (d) recommended methods and indicators for the assessment of incentives

1.3 Approach

The main focus of the research in WP6 is methodological, resulting in a series of recommended assessment approaches and specific indicators that may be used in practice in the subsequent real life trials. As such the research considers both theoretical ideals alongside pragmatic constraints, for example around data availability. The primary approach therefore involves review, synthesis and design around of the state of the art. This includes journal publications, reports of similar projects, and existing best practices in appraisal and assessment. The design element has evolved by adapting, merging and proposing new schema within the novel SUNSET context. The outputs of the research are such that they will provide guidance which may be adapted for use in other related research, which is a source of added value for the work.

1.4 Document structure

This document is organised into four chapters. They are recapitulated below and their relations to the objectives defined in §1.1 are shown in Table 1.3.

- Chapter 1
  This chapter provides the introduction to the document, including summarising the goals of WP6 and D6.1, the main results and innovations, the approach, and the document structure.

- Chapter 2
  This chapter focuses on indicators and methods to evaluate the operational success of the project. The criteria of operational success, as derived by the SUNSET consortium, will be addressed. In brief, the operational success should be viewed in terms of how well the system: (1) can monitor travel related characteristics to support insight and feedback for different types of goals (D6.1 will place emphasis on individuals' travel goals), (2) fulfils the social media concept, and (3) be adequate from the experts' and (4) users' points of views. Therefore, §2.2 to §2.4 of this chapter are structured to address those four objectives. Finally, our recommendations for SUNSET in terms of indicators and methods to evaluate the operational success are summarised in §2.5.

- Chapter 3
  This chapter places emphasis on indicators (§3.2) and methods (§3.3) used to evaluate changes in individuals' travel behaviour due to the offering of incentives. The indicators will take into account stages of behavioural changes as indicated in existing behavioural

theories. The indicators will be derived not only from the observable characteristics of behavioural change (e.g. reduction of the number of trips by car) but also from the non-observable characteristics (e.g. changes in attitudes and intention towards car use). In addition, data requirements and methods for data gathering are discussed in §3.4. At last, our recommendations for SUNSET concerning indicators, methods, and data requirements are summarised in §3.5.

- Chapter 4
  This chapter contains the final conclusions of D6.1 and summaries of the recommendations for SUNSET.

Table 1.3 Document structure in relation with objectives

Content	Chapter 1	Chapter 2	Chapter 3	Chapter 4
Introduction	✓			
Objective 1 Individual goals		✓		
Objective 2 Social networks		✓		
Objective 3 Functionality		✓		
Objective 4 Usability		✓		
Objective 5 Behaviour indicators			✓	
Objective 6 Methodology			✓	
Objective 7 Data requirement			✓	
Conclusion				✓

It is worth noting that, although the objectives above are well defined, it is inevitable that some interrelation exists among them. For example, the amount of help that tripzoom can provide for achieving individual travel goals (Objective 1) is correlated to the effectiveness of incentives (Objectives 5~7). It is therefore likely that the indicators proposed for evaluating these objectives would overlap with each other to a certain extent. In this document, Chapter 2 and Chapter 3 will be written independently, whilst Chapter 4 will provide an overview on the indicators derived in the preceding chapters.

2. Evaluation approach for the operational success of SUNSET

2.1 Introduction

During the LL periods in Enschede, Leeds, and Gothenburg, a number of predefined incentives (presented in D3.3 and discussed in Chapter 3) will be offered to users. The objectives of those incentives are to persuade people to voluntarily change their travel behaviours towards sustainable forms and to support sustainable daily mobility. In Chapter 3, indicators and methods that can be used to assess how well those incentives perform will be addressed. However, besides the success of incentives, there are other criteria of success related to the operationalization of the system. These criteria have been defined in the context of the SUNSET system are listed in Table 2.1.

Table 2.1 Criteria for the operational success of SUNSET

Criteria	Description	Related system components
The system meeting particular goals	The system should be able to reach EU objectives on improved mobility and the system objectives. In addition, the system should be able to help individuals to meet their travel objectives.	app
Success of the social media concept	This definition of success is assessed based on how well the use of social media and social networks works towards achieving the goals of the project.	app web portal
The system functioning as intended	This definition of success is assessed based on the technical and system requirements set by the technical work packages, such as the acceptable level of the accuracy of the detected mobility pattern data and the acceptable level of battery life.	app web portal city dashboard
The "usability" of the system	This definition of success is defined based on the usability point of view. Therefore, it focuses on users' perceptions regarding the system's learnability, efficient of use, effectiveness, memorability, and users' satisfaction with the system.	

This chapter aims to address the above criteria of success by providing sets of methods and indicators that can be used to evaluate the success of the system from these four criteria. The first criterion of success is interpreted here as being related to the effectiveness of the system to meet individuals' travel goals and to contribute to the city and EU goals defined in D3.1. To avoid overlap with D6.2 (which takes a system wide perspective), this deliverable will only address the indicators related to individuals' travel goals. Indicators to evaluate how well the system contributes towards the achievement of the city and EU goals will be addressed in D6.2.

The second criterion of success deals with the innovation of the SUNSET project in the use of social media and social networks. Specifically, it focuses on indicators suitable to evaluate the use of social media related aspects of the SUNSET system.

Finally, the third and fourth criteria of success are related to the evaluation of the tripzoom app in SUNSET. The criterion on functionality addresses the methods and indicators used to evaluate the tripzoom app during the development phase. It is therefore strongly related to the app release cycle, during which the app is assessed periodically by experts who are familiar with the technology and the SUNSET concept. The results of the release periods are reported within other deliverables (arising from WP7). However, the selection of method and indicators for assessment

are provided in this chapter. The criterion on usability places emphasis on methods and indicators that can be used to evaluate users' experiences with the tripzoom app during the LL periods.

In line with the definitions of success, the remainder of this chapter is structured as follows. The indicators related to individuals' travel goals will be addressed in §2.2. Subsequently, the success of the social media concept will be addressed in §2.3. The methods and indicators to evaluate the tripzoom app during the development stage and the LL periods are provided in §2.4. The last section, §2.5, will summarise our recommendations for the practical LL trials to be undertaken in SUNSET, arising from the issues addressed in §2.2 to §2.4.

2.2 The system meeting individual travel goals

This section places emphasis on indicators that can be used to assess how well the SUNSET system enables people individuals' travel objectives. Individuals' travel objectives have been previously defined in D3.1 and D3.2, and therefore form the basis for the derivation of indicators suitable for evaluation. In D3.1, it is concluded that an individual traveller aims to make travel decisions that make the best use out of time, scheduling, household resources and costs, social networks, normative beliefs and expectations, identities and culture, and pleasure. These objectives are further specified in D3.2 for the modelling purposes. For instance, an individual traveller tends to select a transport option that can minimise his/her time, scheduling effort, and travel costs. In addition, an individual inclines to choose a transport mode that can maximise safety/security, pleasure, and so on. Furthermore, Individuals' travel objectives are reflected in people's travel behaviours. For example, an individual who is trying to minimise his or her travel time to go to work tend to select the quickest transport mode or the quickest route. It may also be reflected in the adjustment of departure time to avoid morning traffic congestion. The possible impacts of individuals' travel objectives on their travel behaviour have also been identified in D3.2. Travel objectives defined in D3.1 and D3.2 are presented in Table 2.2. The table also summarises the possible impacts of travel objectives on travel behaviour (in the column to the right).

Table 2.2 Individual's travel objectives identified in D3.1 and D3.2

Individual objectives (from D3.1)	Individual objectives in and their potential impacts on travel behaviour (from D3.2)
Time	Minimising travel time Major behavioural impacts: • Choose the (combination of) mode(s) that is the quickest • Always choose the quickest route Minor behavioural impacts: • Depart earlier or later to avoid congestion, subject to constraints (or delay and early arrival penalties) • Tendency to use a high speeds in the vehicle and keep a short headway • Tendency to change lanes and/or overtake more frequently
Scheduling	Minimising scheduling effort Major behavioural impacts: • Choose the travel mode that is available/known/offered as options (by the scheduling tool) Minor behavioural impacts: • Choose the "default" route (normally the fastest route) • Tendency to make fewer trips; tendency to visit the destination

	available/known/offered as options
Household resources and costs	Minimising costs Major behavioural impacts: • Choose the cheapest travel mode (based on out-of-pocket cost) • Choose the cheapest departure time (e.g. off-peak discount fare on public transport) • Choose the cheapest route (in terms of fuel cost and any applicable road toll) Minor behavioural impacts: • Avoid unnecessary/long trips • Avoid unnecessary ac-/decelerations (in order to save fuel)
Safety/security	Maximising safety/security Major behavioural impacts: • Prefer safer routes, including: (i) preference of the highway over local roads; (ii) avoidance of congested route; (iii) preference of heavily instrumented route (e.g. lighting), especially at night • Prefer a safe (lower) speed and (longer) headway with the front vehicle • Prefer safer/securer travel mode, e.g.: (i) car may be considered more secure than transit; (ii) bicycle might be considered as (un)safe with(out) exclusive bicycle lanes Minor behavioural impacts: • Higher concentration/attention level • Less risky lane change/overtaking (i.e. longer critical gaps) • Always comply with traffic rules
Social networks	Maximising overlap/synergy with normative Major behavioural impacts: • Do not try to avoid peak hour or congestion • Follow a similar choice of that of peers in mode selection • Tendency to follow the "default" route Minor behavioural impacts: • Comply with speed limit • Tendency to comply with traffic rules
Normative beliefs and expectations	Maximising identity recognition Major behavioural impacts: • More consistent in mode choice over time, e.g.: (i) travellers with a 'green' identity prefer the transit mode; (ii) travellers with a 'car user' identity always choose car • Drivers with a 'green' identity tend to have a steady speed profile; drivers with a 'motor biker' identity tend to have high speed and make more ac-/decelerations • Do not try to avoid peak hour or congestion Minor behavioural impacts: • Make more planned trips (esp. to locations where fellows frequent) and less spontaneous trips
Identities and culture	Maximising capital Major behavioural impacts: • Prefer to use personally owned facilities (e.g. cars, bicycle) Minor behavioural impacts:

	• Tend to make more trips when facilities are available
Pleasure	Maximising pleasure Major behavioural impacts: • Choose the most pleasurable/comfortable route, such as a route with a good view or alongside sites of attraction • Choose departure time based on congestion avoidance and pleasure-related preferences, such as weather Minor behavioural impacts: • Choose the most pleasurable/comfortable travel mode • May lose concentration (due to e.g. scenery, music) • Tend to speed and tailgate when deemed pleasurable

The SUNSET system tries to address individuals' travel goals through the offering of incentives in tripzoom. Those potential incentives have initially been defined in D3.3 and are summarised in Table 2.3 below. In practice the incentives will be selected from this list and given a local 'flavour' according to the objectives and context of the particular LL. In brief, the incentives cover real-time traffic information on the road network, travel feedback on the individual's mobility pattern, setting targets and feedback on those targets, challenges (using points with and without exchange value), and social network incentives for peer-to-peer messages, sharing location, find a travel buddy, and treasure hunt.

Table 2.3 Incentives identified in D3.3

Type of incentive (from D3.3)	Description of incentive
Real-time travel information provided by the system	• The system gives information about the most recent conditions on the road networks. • The system gives alerts to users whenever there is a relevant event (either expected or unexpected) that may influence their travel behaviours. In addition: a) Users should be able to enable/disable the incentive; b) When enabled, alerts can only be given based on regular activity-travel patterns (i.e. related to the spatial parameter). Therefore, when a traveller uses a new route for the first time, alerts related to that route will not be available.
Social networks for peer-to-peer travel information/messages	The system provides an infrastructure for users to exchange messages among each other. In general, there are two types of messages: • Information alerts related to the conditions on the road or infrastructure. • Tips/advice on travel.
Feedback based on self-monitoring of own travel behaviour	The system records users' daily activity-travel patterns and presents the recorded information back to the users. This is the basic incentive and a feature of tripzoom app.
Feedback based on setting targets	The system allows the users to set their own travel targets.
Challenges (using points without an exchange value)	Every user who exhibits certain travel behaviours (e.g. cycling or walking) will be awarded points. This can be related to a competition with other users based on points (akin to on-line games).
Challenges (using points with an exchange value)	This category is related to: • Challenges set by the system or by the 3rd parties. • Periodic offers akin to a loyalty card. For instance, once a user reaches 100 points, he or she can redeem the points for a tangible reward.
Social networks for sharing location	Every user can share their current location to selected users.
Social networks for finding a buddy	Every user can use find a buddy to find a travel companion.

Social networks for treasure hunt	This is another type of challenges. As an example, treasure (in the form of points) can be hidden at a specific coordinate and can only be unlocked whenever users cycle pass the coordinate.

Based on individuals' preferred travel goals (Table 2.2) and incentives (Table 2.3), a qualitative assessment is made to examine the possible impact of each incentive on individuals' travel goals. This is done by mapping the likely influences of an incentive on individual behaviour (Table 2.3) with the main behavioural impacts of a travel objective (Table 2.2). For instance, real-time traffic information is likely to help individuals to minimise their travel time by enabling them to avoid traffic congestion. It may also contribute to reducing the scheduling effort (by reducing the time needed to search for information related to the road conditions) and reducing travel costs (through the reduction of fuel spent in congestion). The assessment of the possible impacts of each incentive on travel behaviour is listed in Table 2.4.

Table 2.4 Incentives and individuals' travel goals

Type of incentive	Influence of incentives on travel goals
Real-time travel information provided by the system	This incentive can help an individual to: • Minimise total travel time by reducing time spent in traffic congestion or to provide detour due to road work or traffic accident; • Minimise scheduling effort by reducing time needed to look for traffic information from different sources; • Minimise travel costs (by car) by omitting the extra fuel spent in traffic jam, thus reducing total costs.
Social networks for peer-to-peer travel information/messages	Since it is a user generated content incentive (through social networks), the influence of this incentive on individuals' travel goals depends on the content of the messages. Several examples can be found below: • Messages containing information about traffic congestion/accident/road work may help a traveller minimise travel time and furthermore, reduce travel costs; • Messages containing tips/advice related to unsafe bus stop/road or bad road surface may impact on individuals' safety and security; • Messages containing information on CO2 emission due to car use or similar types of information may increase behavioural awareness and influence normative belief ; • Messages containing community activities, such as cycling or walking, may make a traveller feels accepted and part of a group with similar interests/views/hobbies and furthermore may give identity.
Feedback based on self-monitoring of own travel behaviour	Feedback on travel patterns can help an individual to achieve his or her travel goals only if it is combined with self-reflection. This is due to the nature of the incentive, which does not give direct feedback, such as: *"If you leave home at your usual time at 8:30 with your bicycle instead of car, you will arrive at your work at your office at your usual time at 9:00. You will improve your health and reduce your travel cost"*. When combined with self-reflection, self-monitoring can help traveller to: • Minimise travel costs; • Optimise total travel time (for instance by trip-chaining) and scheduling.
Feedback based on setting targets	Similar to self-monitoring, feedback on target can help an individual to: • Minimise travel costs (e.g. setting cost target); • Minimise travel time; • Improvement of health (e.g. setting calorie target); • Give identity (e.g. setting CO2 target).
Challenges (using points without an exchange value)	Challenges can be used as a mechanism to acknowledge the "desired" behaviours, to relate to social networks, and to add the fun factor (or pleasure).
Challenges (using points with an exchange value)	Similar to the above, challenges can be used as a mechanism to acknowledge the "desired" behaviours, to relate to social networks, and to

	add the fun factor (or pleasure). Additionally, the exchange value may also work as an additional motivational factor.
Social networks for sharing location	The basic idea of this social network incentive is to have someone trusted to watch over a traveller. Thus, it aims to address individuals' safety and security.
Social networks for finding a buddy	This incentive is based on the provision of a medium through which individuals may find a travel companion (in social networks), or to have someone who can motivate a traveller to cycle or walk together (influencing normative belief), and to make daily trips more enjoyable (to give pleasure).
Social networks for treasure hunt	The main purpose is to make daily trips more fun (to give pleasure).

From Table 2.4 above, it can be seen that the incentives can hypothetically help individuals achieve their travel objectives. Whether or not the tripzoom app can help travellers to attain their travel goals in practice will be investigated during the LL period. For this purpose, indicators/techniques for the assessment have to initially be selected. Selecting the indicators is not a straightforward process because individuals' travel goals vary considerably in different contexts, such as: trip purposes (e.g. commuting vs. leisure trips), attitudes (e.g. having travel dislike), and personality (e.g. adventure seeking vs. organiser) (Ory & Mokhtarian 2005). Moreover, a change in travel behaviour does not necessarily improve all individual travel goals. For example, an individual may choose to travel by bike instead of by car to avoid congestion. This may decrease the costs of travelling and the scheduling effort, but may increase the travel time. Therefore, assessing the incentives and behavioural change requires a multi-objective approach instead of single-objective indicators. In the remainder of this section, possible factors influencing individuals' travel goals are discussed, an experimental hypothesis is proposed for assessing the achievement of the goal, whilst alongside this an indicator and data collection process is described. Within the following sections some notation is used to refer to particular data collection approaches that are proposed and this notation is briefly outlined here.

The data needed for evaluating these factors can be collected using three methods:

- Automatic data detection by tripzoom
- Manual data collection through travel diaries
- Questionnaires

The questionnaire method solicits answers from the tripzoom users by explicitly asking them questions, which can be divided into three categories (Table 2.5):

- Questions that the researcher would want to ask once only (**QR1**), in order, for example, to categorise participant types or to establish background factors (independent variables) that may influence travel choices. These questions can be delivered by an electronic questionnaire (or potentially through focus groups too);
- Questions that the researcher may wish to ask twice or more (**QRM**), for example to monitor changes in beliefs. They are suitable for delivery by electronic questionnaire as they do not require the participant to recollect 'current/recent' tripzoom related experiences;
- Questions where the best quality responses will come from a reflection of very recent transport experience or of the tripzoom experience. These questions may be asked once or more than once through the experience sampling (**XP**) mechanism in tripzoom.

Table 2.5 Questionnaires: question categories

Category	Repetition	Data source	Main usage
QR1	Only once	Electronic questionnaire	To-ask-once-only type of information, e.g. to

			categorise participants
QRM	Twice or more	Electronic questionnaire	Hypothesis testing
XP	Once or more than once	Experience sampling	Hypothesis testing

2.2.1 Travel time

In many travel surveys that use a travel diary (e.g. Axhausen 1997; Axhausen *et al.* 2002), travel time is often measured by the difference between the starting and ending time of the trip. For tripzoom, this type of information can automatically be obtained from the recorded travel times. However, the objective measurement of the actual travel time should not be the only indicator to assess an individuals' goal of travel time. Subjective measurements should also be included. For instance, in Ory & Mokhtarian (2005), subjective indicators are also used to evaluate the extent to which individuals like to travel. Subjective indicators can be used to gain insight into how an individual values their travel time. This can be done, for instance, by asking travellers whether or not they feel they spend too much time commuting. In addition, since the main purpose is to measure how useful the tripzoom app is in helping the travellers obtain their desired travel time, measuring participants' relative desired travel time becomes important. This can be done by asking participants (during the base case period) how much travel time they are wishing to spend in the coming months for commuting trips compare to their current level (with the responses anchored by "much less" and "much more"). This way, we can further ask participants (during the incentive period) how the tripzoom app helps them obtain their desired travel time. Possible indicators to measure travel time are listed in Table 2.6 below. The table also shows the possible data sources. A combination of techniques to obtain data is proposed: the recorded daily mobility data for objective variables, questionnaires and experience sampling for qualitative data (e.g. subjective variable).

Table 2.6 Variables for measuring travel time

Variables/indicators	Data collection process
Objective indicator of travel time (H1: there has been a change in observed travel time, following the use of tripzoom and introduction of an incentive or a package of incentives)	
Travel time (minute) by trip purposes: • Work/school commute • Work/school related • Entertainment/recreational/social	tripzoom
Subjective indicator of travel time (H1: there has been a change in perceived travel time, following the use of tripzoom and introduction of an incentive or a package of incentives)	
Travel time by trip purposes (five point semantic-differential scale anchored by "none" and "a lot"): • Work/school commute • Work/school related • Entertainment/recreational/social	QRM
Relative desired travel time (H1: there has been an increase in satisfaction with current travel time, following the use of tripzoom and introduction of an incentive or a package of incentives)	
Relative desired travel time individuals wish to spend compare to current level by trip purposes (anchored by "much less" and "much more"): E.g. How much travel time are you aiming at for work/school commute in the coming months compare to the current level? • Work/school commute • Work/school related • Entertainment/recreational/social	QRM
Output of desired travel time	

(H1: there has been an increase in the contribution that tripzoom makes to achieve the desired mobility, following the use of tripzoom and introduction of an incentive or a package of incentives)	
Tripzoom contribution in helping an individual achieve their desired mobility by trip purposes: • Work/school commute • Work/school related • Entertainment/recreational/social How does tripzoom help you obtain your desired travel time for work/school commute? (five point semantic-differential scale anchored by "none" and "a lot")	QRM

2.2.2 Scheduling effort

Similar types of indicators related to indicators of travel time are proposed for scheduling effort. However, it is infeasible to obtain objective measurements on scheduling effort. Therefore, we propose only indicators based on qualitative measurement, namely: subjective, desired mobility, and output mobility indicators. Users can be asked how much effort they spend to schedule different trips (e.g. work commuting and recreational). In addition, they should be asked about their desired scheduling effort (during the base case period) and whether tripzoom helps them to achieve it (during the incentive period). Table 2.7 shows proposed variables and data sources related to the scheduling effort.

Table 2.7 Variables to measure scheduling effort

Variables/indicators	Data collection process
Subjective indicator of scheduling (H1: there has been a decrease in perceived scheduling effort, following the use of tripzoom and introduction of an incentive or a package of incentives)	
Scheduling by trip purposes (five point semantic-differential scale anchored by "none" and "a lot"): • Work/school commute • Work/school related • Entertainment/recreational/social	QRM
Relative desired scheduling effort (H1: there has been a move towards greater satisfaction with travel scheduling effort needed, following the use of tripzoom and introduction of an incentive or a package of incentives)	
Relative desired scheduling effort individuals wish to take compare to current level (anchored by "much less" and "much more") E.g. How much scheduling effort are you aiming at for your daily trips in the coming months compare to the current level?	QRM
Output of desired scheduling effort (H1: tripzoom provides assistance with trip scheduling, compared with not having tripzoom)	
How tripzoom help you in your trip scheduling? (five point semantic-differential scale anchored by "none" and "a lot")	XP

2.2.3 Household resources, identities, and culture

Indicators related to household resources and identities are strongly related to household characteristics (such as age and gender), preferred mode (based on the most used modes which are recorded), and individuals' perception towards their car. This variable is added based on the extent to which individuals like to travel (Ory & Mokhtarian 2005). Ory & Mokhtarian (2005) stated that an individual may choose to use a car because of the image that it portrays (e.g. for sporty car or luxury car). In addition, constraints in individuals' mobility are also taken into account in one of the variables. These constraints are distinguished as those due to stress when driving in a highway and at night, and also accessibility constraints as a function of resources.

Finally, indicators related to subjective desired mobility and output of desired mobility due to tripzoom are also proposed. These variables are listed in details in Table 2.8.

Table 2.8 Variables to measure household resources, identities, and culture

Variables/indicators	Data collection process
Household characteristics and resources	
Age (unit)	QR1
Income (€, £, kr)	QR1
Household size (unit)	QR1
Number of children under 18 in HH (unit)	QR1
Number of household workers (unit)	QR1
Employment type (full time; part time; not working; others)	QR1
Education level	QR1
Gender (male; female)	QR1
Distance travelled by modes	tripzoom
Accessibility to resources	
Preferred mode and type	
Most used modes (car; public transport; bicycle; on foot; motorbike; moped or others)	tripzoom
If the used mode is car: Vehicle categories (small; compact; mid-sized; large; luxury; sport utility vehicle; minivan/van; pick-up truck; sports)	QR1
Mobility constraints	
Mobility constraints measure inability to travel freely, for instance due to personal difficulty or stress travelling (''No limitation'', ''Limits how often or how long'', ''Absolutely prevents''). • Driving on the highway • Driving at night	QR1
Access to a car (anchored to "none" and "always")	QR1
Access to a bicycle (anchored to "none" and "always")	QR1
Access to public transport (anchored to "easy" and "difficult")	QR1
Subjective indicators on distance and costs (H1: there has been a change in the perceived distance travelled by car / perceived monthly travel cost, following the use of tripzoom and introduction of an incentive or a package of incentives)	
Estimated distance travelled by car in a year	QRM
Estimation of monthly travel costs	QRM
Relative desired mobility (H1: there has been an increase in satisfaction with perceived distance travelled by car / perceived monthly travel cost, following the use of tripzoom and introduction of an incentive or a package of incentives)	
How much money are you aiming at for your daily trips in the coming months compare to the current level? (anchored by "much less" and "much more")	QRM
How many kilometres travelled by car are you aiming at for your daily trips in the coming months compared to the current level? (anchored by "much less" and "much more")	QRM
Output of desired mobility (H1: tripzoom provides assistance with achieving desired travel cost or travel distance, compared with not having tripzoom)	
To what extent does tripzoom help you to reach your desired travel costs? (five point semantic-differential scale anchored by "none" and "a lot")	XP
To what extent does tripzoom help you to reach your desired kilometres travelled by car?	XP

2.2.4 Social networks, normative belief, and expectation

Social networks are commonly used to keep in touch with contacts, share photos, play games, follow celebrities, organise social events, get recommendations, and gain/share information. In the health research field, social networks are often used to provide meaning to life and role of satisfaction, to provide emotional support, to provide practical and logistic support, to provide feedback, and to facilitate maintenance of daily routines and provide normative incentives (Sluzki 2010). In the transport research field, the role of social networks has been identified. For instance, results of the focus groups sessions conducted by Binsted & Hutchins (2012) showed that social networks have the potential to spread travel-related information and to increase people's awareness needed for behavioural change. However, the behavioural awareness should places emphasis on individuals' personal benefits (e.g. cost reduction and health improvement) and not on the society's benefits (e.g. pollution, CO2 emissions, and carbon footprint) (Tertoolen *et al.* 1998).

In the existing research on social networks, several indicators are suggested (Table 2.9). These indicators are not only based on the size of the networks (e.g. number of friends in the list), but should also capture the quality of the relationships among users (e.g. family members and close friends). Based on existing research done by Centola (2010), the quality of relationships among members in a network plays a more important role in spreading behaviour compare to the size of the network. This indicates that when information is repeated many times by close friends/relatives, it becomes more likely that a person will adapt the behaviour compare to when the information is repeated by many far distance friends. This goal will also be addressed further in §2.3, as a part of the discussion on how well the social media concept is used in the project.

Table 2.9 Variables to measure social networks, normative belief, and expectation

Variables/indicators	Data collection process
Objective measures	
Number of friends in the list	tripzoom
Number of messages posted	tripzoom
Number of responses sent to others	tripzoom
Number of messages by others that are liked	tripzoom
Number of buddy groups	tripzoom
Number of buddies in groups that have close relations to the user (e.g. Family members)	tripzoom
Subjective measures (H1: tripzoom buddy messages are useful in supporting travel needs)	
Usefulness of the messages posted by (close) tripzoom buddies in supporting daily travels	XP
Emotional support/motivation from (close) tripzoom buddies to cycle or walk more often	XP
Feedback from (close) tripzoom buddies about daily travels	XP
Satisfaction with social networks offered by tripzoom	XP

2.2.5 Pleasure

In-depth studies have been done to investigate and measure individuals' 'travel liking', such as those by Handy *et al.* (2005); Mokhtarian (2005); Ory & Mokhtarian (2005). Travel liking is related to how much a travel is enjoyed. In a conservative point of view, travel is usually viewed to give negative utility, as it is only a media to link different activities that are located at different places and times. Therefore, it is assumed that individual travellers always want to minimise their travel time, costs, or effort. However, recent studies also indicated that travel may give positive utility on its own. This means that some people may enjoy travelling for its own sake. For instance,

travel can reduce stress and act as buffer between different activities. Many people do not mind taking a longer route with a better scenery and taking a slower mode to be in an open air. Ory & Mokhtarian (2005) further specifies variables that are used to measure travel liking, which is determined not only by people's socio-demographic characteristics (such as age and gender), but also by other parameters, such as travel distances (short vs. long distance), trip purposes (e.g. commuting vs. leisure), and most importantly, attitudes towards travelling, personality and lifestyle. Attitudes measure people's attitude towards travel dislike, pro-environmental policy, commute benefit, travel freedom, pro-high density, and travel stress. The personality variables are divided based on people's personality traits, namely adventure seeking, organiser, loner, and calm. At last, lifestyle variables are based on: frustrated, family/community oriented, status seeking, and workaholic people. Since socio-demographic variables have been discussed in one of the previous sections, only variables related to attitudes, lifestyle, and personality are emphasised and listed in Table 2.10 below. Additional indicators of excess travel and desired mobility are added. The former measures how often people engage in unnecessary travel and the latter measures the added value of the tripzoom app in addressing the goal.

Table 2.10 Variables to measure household pleasure: attitudes, personality, and lifestyle

Variables/indicators	Data collection process
*Attitudes on travel dislike** (H1: there has been a change in attitudes to travel liking, following the use of tripzoom and introduction of an incentive or a package of incentives)	
Travel is boring.	QRM
I like travelling to new places.	QRM
The only good thing about travelling is arriving at your destination.	QRM
*Attitudes on pro-environmental policy** (H1: there has been a change in attitudes to environmental issues, following the use of tripzoom and introduction of an incentive or a package of incentives)	
To improve air quality, I am willing to pay a little more to use an electric or other fuel clean vehicle.	QRM
We should raise the price of gasoline to reduce congestion and air pollution.	QRM
We need more public transport, even if taxes have to pay for a lot of the costs.	QRM
We should use congestion fees to reduce congestion and to finance more public transport	QRM
*Attitudes on commute benefit** (H1: there has been a change in attitudes to commuting time, following the use of tripzoom and introduction of an incentive or a package of incentives)	
My commute is a real hassle.	QRM
My commute trip is a useful transition between home and work.	QRM
The travelling that I need to do interferes with doing other things I like.	QRM
I use my commute time productively.	QRM
*Attitudes on travel freedom** (H1: there has been a change in attitudes to travel freedom, following the use of tripzoom and introduction of an incentive or a package of incentives)	
In terms of local travel, I have the freedom to go anywhere I want to.	QRM
In terms of long-distance travel, I have the freedom to go anywhere I want to.	QRM
*Attitudes on pro-high density**	
Living in a multiple family unit wouldn't give me enough privacy.	QR1
I like living in a neighbourhood where there is a lot going on.	QR1
*Attitudes on travel stress** (H1: there has been a change in attitudes to travel stress, following the use of tripzoom and introduction of an incentive or a package of incentives)	
I worry about my safety when I travel.	QRM
Travelling makes me nervous.	QRM
Travelling is generally tiring for me.	QRM

I tend to get sick when travelling.	QRM
I am uncomfortable being around people I don't know when I travel.	QRM
Personality: Adventure seeking	
Adventurous	QR1
Variety seeking	QR1
Spontaneous	QR1
Risk taking	QR1
Personality: Organiser	
Efficient	QR1
On time	QR1
Personality: Loner	
Like being alone	QR1
Like being independent	QR1
Personality: Calm	
Aggressive	QR1
Patient	QR1
Lifestyle: Frustrated	
I often feel like I don't have much control over my life.	QR1
I am generally satisfied with my life.	QR1
Lifestyle: Family/community oriented	
I'd like to spend more time with my family and friends.	QR1
My family and friends are more important to me than my work.	QR1
Lifestyle: Status seeking (H1: there has been a change in attitudes to car and status, following the use of tripzoom and introduction of an incentive or a package of incentives)	
To me, the car is a status symbol.	QRM
A lot of the fun of having something nice is showing it off.	QRM
Lifestyle: Workaholic	
I'm pretty much a workaholic.	QR1
I'd like to spend more time on work.	QR1
Excess travel (H1: there has been a change in attitudes to discretionary travel, following the use of tripzoom and introduction of an incentive or a package of incentives)	
Excess travel measures how often people engaged in activities involving unnecessary travel (never/seldom; sometimes, often): • How often do you travel with no destination in mind? • How often do you travel just for the fun of it? • How often do you travel mainly to be alone?	QRM
Output of desired mobility (H1: tripzoom introduces fun into the travel experience, compared with not having tripzoom) (H1: tripzoom raises awareness of travel cost, compared with not having tripzoom) (H1: tripzoom is perceived to help with commuting trips, compared with not having tripzoom) (H1: tripzoom is perceived to help with travel freedom, compared with not having tripzoom) (H1: tripzoom is perceived to help with route finding, compared with not having tripzoom) (H1: tripzoom is perceived to help with environmental awareness, compared with not having tripzoom)	
Does tripzoom help make your trip more fun? (five point semantic-differential scale anchored by "none" and "a lot")	XP
Does tripzoom make you be more aware of your own travels (e.g. travel time, cost, and distance)?	XP
Does tripzoom help you in your commuting trips?	XP
Does tripzoom help you gain more freedom during travelling?	XP
Does tripzoom help you discover new routes?	XP
Does tripzoom help you be more aware of the environmental impact of your daily travels?	XP

* Anchored by ''hardly at all'' to ''almost completely''
Source: adapted by SUNSET authors from Ory & Mokhtarian (2005).

2.3 The success of the SUNSET social network concept

The use of social media and social networks lies at the heart of the SUNSET concept. The concept is based on the proposal that system users (travellers) can benefit from transport related social-network activity that are enabled through dynamic and pervasive social media and technology. From the perspective of the transport system operators, the social network concept is intended to support 'smarter choices' - for example in terms of encouraging mode shift to greater public transport use or other more sustainable transport behaviours. This two way flow of benefits is also attuned with the people centred mobility paradigm (whereby the focus is on transport more tailed to individual needs) and the SUNSET objective of the transport system meeting both strategic level objectives and user centred goals. The intention is to achieve these aims through the use of social networks and as a consequence of engaging the resource capital of influences, shared experience and encouragement that the social network may provide. In summary, it is hypothesised that an individual end-user of tripzoom can:

- benefit from the factual information concerning travel that may be contained in transport related postings that the community may make;
- contribute to the value of the social network by placing their own posts;
- receive very timely transport related alerts from community members who can provide more 'instant' information of the current traffic situation than may be possible for centralised information providers;
- receive updates and alerts as a community from the city operators, for example on known advance changes to the system or to encourage longer term changes to travel habits;
- benefit from the positive encouragement of others with similar transport or lifestyle goals;
- compare their behaviour with friends and colleagues; and
- interface their transport related social network activity with broader social networks activity and communities (i.e. those centred within Facebook, Twitter) to share their (positive) transport related behaviour with a wide set of social contacts.

From the perspective of the city transport authorities, social media enabled social networks can also be used to recruit users into the system.

The design of the SUNSET system has two main foci of social media activity for participants and this reflects a distinction to be drawn generally between web-based social network activity (typically seen in Facebook, twitter and similar) and application (app's) based social network activity. The former is accessible through any web-enabled technology, either static or portable, including home computers, smartphones, laptops, tablets, plus a centrally provided connectivity to the internet. Application based social network activity tends to be based on pervasive technologies and to be developed around more focused functionality. This may be transport related or with another focus that links to transport (such as a health related application). Apps may be designed to have either their bespoke social network functionality that allows information flow between participants or a combination that links bespoke social network functionality with web-based social networks. The way in which the social network activity is enabled is key to the design of an assessment method for the success of a social network concept given it drives the type of data that is generated, the method of data collection and the availability of data. A full description of the functionality of tripzoom and other SUNSET features that support social network activity is provided in D4.5 and D5.2. However, a brief summary in the context of evaluation is included below.

2.3.1 Social network activity in SUNSET living labs

The living lab trials will comprise both a structured recruitment into a maximum of 6 groups plus a free-form group of participants who may join by invitation of current LL members or through finding the tripzoom app through other means. The reference LL may prioritise recruitment to a smaller number of groups in order to explore particular research questions of local relevance. As a result the social networks that develop will be a result of both artificial construction (in the experimental conditions) plus 'naturally' evolving social networks over which the assessment process has no experimental control. This is noteworthy as the extent to which success is related to additional participation and growth of the networks is lower than would be the case if a completely unstructured trial implementation were to take place. This is further noted in §2.3.3. However it may also be the case in a real life implement there is a need for some structured recruitment around a particular geographic location or a particular O-D combination - hence the recruitment and assessment approach described here may ultimately be fairly representative of a real life context.

In terms of functionality, SUNSET has social network enabled activity via both the tripzoom app and web-based social media on dedicated pages, with links to non-SUNSET focused social network activity through users' general social media accounts on Facebook, twitter and the like, more specifically:

1) The tripzoom app for the phone (iPhone and Android) – this includes the ability to check individuals 'travel performance' against a community and share their profile with friends. The app also has the ability to connect to users' general Facebook and Twitter accounts, with a push- message facility.

2) A community which can build (or be encouraged to build) around web-enabled social media, specifically the tripzoom twitter feed on the portal, the tripzoom Facebook (F/B) page, the (closed) F/B pages that will be established for particular experimental groups recruited to the living labs and also posts on LL participants own F/B pages.

This section focusses on a methodology to evaluate the success of the SUNSET concept and particularly the social network element. The evaluation approach has been designed to evaluate the success of the SUNSET (social media) concept that is seen to take place through both fields of social network activity. In this respect the approach has a generic added value beyond the SUNSET project and may be readily adapted for use with other schemes that either concern app based social network activity or web-media enabled social network activity or both.

2.3.2 Evaluation of social media enabled social network schemes: case studies

The evaluation of the success of social-media based social-networks schemes in transport (or more generally) is one which has been barely addressed by the academic literature. This is a consequence of the field as a whole being novel and largely in its infancy in the transport sector. In terms of scope of the literature, the use of social networks more generally to influence participants' choices and behaviour has been discussed in detail in D3.1 in the context of the design of incentives - this literature is therefore not repeated here. The focus here is also not to mine the social networks created themselves in order to understand the structures or relationships between different groups as participants join and membership functions – a body of computer science based literature exists in that domain, see for example Kwak *et al.* (2009).

The focus here broadly concerns evaluation of the success of the social network concept in acting as a dynamic platform for information exchange, social encouragement through buddies and influence towards smarter choices.

The analysis of publically available data from users of social media sites is now a well-established tool in the marketing sector (Melville *et al.* 2009). Two types of data that are commonly obtained are user profile data (to understand the socio-demographics of the customer base) and preference or sentiment analysis of contents (for example to understand customer preferences or reactions to a new product. A location based example of the use of profile data is the study by Pontes *et al.* (2012) based on a large dataset collected from Foursquare using the system API. The researchers crawled user profile data consisting of user type, user home city, list of friends, 'mayorships', tips, 'dones', total number of check-ins, Twitter screen names and Facebook identifiers. The aim of the study was to explore issues around attitudes to privacy and also the accuracy of the some of the personal profile data obtained. It is worthy of note that the accuracy of twitter based location indications was as low as 66% with the conjecture that users did not provide real locations to avoid third party marketing. An assessment of the predictive power of social media for outcomes is given by Asur & Huberman (2010) based on the 'tweet rate per hour' and a simple correlation analysis with illustrations related to new film releases. A similar correlation analysis between social-media and traditional mass surveys results is provided for the political arena by Ceron *et al.* (2012).

A selection of case studies concerning evaluation processes are presented in more detail here. These have been selected on the basis that they illustrate the types of techniques that have previously been used for social media enabled social networks.

- The first example concerns the evaluation of an environmental website, with some similarities to the SUNSET concept as the intention there is to use reflective processes to encourage more sustainable choices and behaviours.
- The second example concerns the evaluation of social media when used for recruitment. This is pertinent to the SUNSET scheme as one of the functions of the social network activity would (in real life implementation) concern recruitment of participants by a city transport operator or transport scheme operator.
- The third case study highlights the assessment criteria that were used for social media as a significant element in influencing activity within a community and particularly in allowing open communication of social values and social expression. This type of role for social media is one which has possibly had most media attention and illustrates the considerable 'power' of social media (considering power = volume * depth). The case is of relevance to SUNSET given the prior lethargy in delivering behaviourally orientated demand management approaches in the transport sector. The conjecture within the SUNSET concept is that the free flow of information and user generated content may serve as a more effective means to encourage smarter choices than previous mechanisms.
- The fourth case study is likely to be of particular interest to transport sector stakeholders and operators as it concerns the evaluation of a real-life incentives scheme delivered using social media.
- The fifth example illustrates how a social media scheme to encourage health improvements and employing the encouraging and supporting role of buddies was evaluated. This has relevance to the SUNSET concept on two accounts – the use of buddies and encouragement through social networks, which is embedded in the functionality of the software and also the need to evaluate a scheme concerned with engendering health related impacts. For each case study the means of data collection

and measurement criteria are recorded to inform the design of the SUNSET social network evaluation.

- The sixth example gives an overview of how web 2.0 applications are used in the field of education and identifies good practices and key impact areas of those applications. The relevance of this case for SUNSET is that it provides support for the idea that the use of social media can contribute to the ability and motivation of people to reach their goals.

These case studies are discussed in details below.

Case (1): Energy Saving Behaviour (Mankoff *et al.* 2010)
StepGreen.org was a site designed to promote energy saving behaviours and this paper reports on which of the design features were successful and how the social network aspects were evaluated. The software scheme involved a combination of interactive progress visualization on a StepGreen.org website, plus a scaled-down down app. Progress for participants was reflected after just one or two desirable actions were taken. In terms of evaluation, a deployment trial was designed to assess the usability of the system, rather than to test behavioural responses. StepGreen.org was deployed for 3 weeks to 32 members of the local community who had an active MySpace account. Participants completed an online pre-test questionnaire about their energy-saving behaviours, environmental attitudes, decision-making styles and daily life habits (e.g., car ownership, home rental vs. ownership). For three weeks they were asked to log into their MySpace accounts and view their profile page at least twice a day. A new action suggestion was presented each time they visited their MySpace profile. Logging tools were used to record when each person viewed an action, committed to an action, and reported fulfilling an action; and when a stranger viewed a participant's MySpace profile page. After the three weeks were over, participants completed an online post- test questionnaire asking their feedback on various features of the site (e.g., the actions, the visualizations). Logging tools were used to record when each person viewed an action, committed to an action, and reported fulfilling an action; and when a stranger viewed a participant's MySpace profile page. To summarise, the evaluation approaches used included qualitative 'before and after' electronic surveys, self-reporting through logging actions, automatic monitoring e.g. of pages visited and interrogation of posts on public MySpace walls or blogs about the site. The research also notes the number of friends added to their network during the trial.

Case (2): Recruitment (Fenner *et al.* 2012)
Because recruiting young people for health research using traditional methods has become more expensive and challenging, an Australian research (Fenner et al, 2012) experimented with recruitment using social media. They placed an advertisement on Facebook to recruit young females from Victoria, Australia. From their research the most important indicator for success is the average cost in advertising fees per compliant participant. For this study the average cost per participant was US $20, which was considered to be highly cost-effective. In addition to this value-for-money indicator a number of general indicators are described, such as: the number of responses to the advertisement, the number and percentage of those females that agreed to participate and the number and percentage of those females that completed the survey in the end. For the advertisement itself, the number of clicks on the advertisement and the number of visitors of the website were also reported.

Case (3): Influencing power of User Generated Content (Bekkers, Moody & Edwards 2011)
In this research different cases of social media usage for mobilization of groups of people and social media influence on general imaging in The Netherlands are described. The most interesting case is about the protests against the "1040 Hour-Norm". The name of the norm refers to the yearly number of hours that students at secondary schools at least had to follow. Because

a lot of schools could not meet this norm students were required to remain at school for "self-study" hours or for doing their homework. Mobilization of protesters and organisation of the actions against the norm were done using social media such as MSN and Hyves (a Dutch equivalent of Facebook). One example is mentioned in which the MSN posting of one guy calling people to strike at 23 November 2007 led to large protests on that date. Because of the size of the protests traditional media picked up the story and started covering it leading to an even larger mobilization of protesters. The Ministry of Education was totally unprepared for the large scale of the protests and the speed of organisation and mobilization using social media. This led to a wrong communication strategy in which only traditional media were used which led to an even more hostile sentiment under the protesters. Altogether this led to an investigation and, because of that, to an adjustment of the norm. Also social media were from this moment in time monitored more often and taken more seriously. When we look at the way social media activity was measured is this case the following indicators are mentioned: the number of posts/blogs/uploads on (social) media and the number of likes or new members per day.

Case (4): Traffic Related Incentive Scheme (Spitsmijden 2007)
Spitsmijden (Dutch for: Peak hour avoidance) is a Dutch initiative that tested, and currently still tests, how people react in response to a financial reward for not travelling on a certain road section within peak hours. The first measure of success is the number of participants in the experiment. The base idea of Spitsmijden is that people get a financial reward when they do not travel on a certain busy motorway within peak hours. In order to receive the reward people can either use other modes or travel via other routes, outside the peak period, together with somebody else or not at all. Spitsmijden (2007) describes the learning experiences from the first experiment. This experiment was conducted with people travelling over the A12 between Zoetermeer and The Hague and led to a substantial reduction (50 to 60%) of the number of trips inside the peak period. In the experiment the recruitment of participants was done some time before the start of the pilot. When the experiment started - which was later than intended due to various delays - a number of people no longer wished to participate in the Spitsmijden scheme. The pre-start drop out of respondents is therefore another indicator for the success of the concept.

Case (5): Buddies in health improvement (Centola *et al.* 2010)
This research studies how network characteristics influence for behavioural change of individuals in the network. Two networks are compared: clustered-lattice networks, in which people have mutual friends, and random-networks, in which people don't have mutual friends. The core of the research was that people were asked to promote healthy behaviour (by encouraging registration for a health forum) within their social network. The outcome of the experiment suggests that behavioural change is more likely to occur when people have mutual friends in the social network than when friends are more randomly distributed. The main reason given for this higher degree of behavioural change is that it is more likely that people receive an invitation more than once in such a network. An indicator for the quality of a network therefore is the density of friends/buddies. This can be measured by the number of an individual's friends that also know each other.

Case (6): Social media use in Learning (Redecker *et al.* 2009)
Another example of the use of social media can be found in the field of education and training. Redecker *et al.* (2009) performed a desktop research on current practice of using web 2.0 applications in this sector and did some case studies to study the impact of those tools. An interesting conclusion for SUNSET is that it turned out that in some good examples the motivation of learners increased because of the use of social media. The main reason for this increase in motivation was that the web 2.0 applications allowed new and diverse learning experiences, but also the social, personalised and collaborative component of the learning are mentioned as

explanatory variables for this increase. Besides that the study concludes that web 2.0 tools support more active learning processes and support the sense of ownership of content which, in turn, encourages learners' motivation. Because of these findings Redecker *et al.* (2009) conclude that one of the key impact areas of the use of web 2.0 applications in learning is that learners can increase their ability to achieve goals. The main assessment method was based on consultation of stakeholders using an on-line survey.

In Table 2.11 an overview of the different data collection methods and indicators of the success of a scheme is given, alongside the projects that used or mentioned these indicators.

Table 2.11 Indicators of the success of a scheme from case studies

Indicators arising from previous studies	Project(s)
Number of (unique) clicks on the advertisement	Fenner *et al.* (2012)
Number of (unique) visitors to the study website	Fenner *et al.* (2012)
Number of participants/users recruited	Spitsmijden (2007), Fenner *et al.* (2012), Mankoff *et al.* (2010)
Drop-out numbers/rate before start of the study	Spitsmijden (2007), Fenner *et al.* (2012)
Completion number/rate of study (i.e. survey)	Fenner *et al.* (2012)
Average advertisement costs per participant	Fenner *et al.* (2012)
Number of posts on social media	Bekkers, Moody & Edwards (2011)
Number of likes/new members per day	Bekkers, Moody & Edwards (2011)
Density of the network (% of mutual friends)	Centola *et al.* (2010)
Visits to particular pages	Mankoff *et al.* (2010)
Qualitative questionnaire 'before and after' trial implementation	Mankoff *et al.* (2010), Redecker *et al.* (2009)
Interrogation of social media post content	Mankoff *et al.* (2010)
Self-logging of actions in response to site (via web interface)	Mankoff *et al.* (2010)

2.3.3 Success of the social network concept in SUNSET

In order to be able to determine whether or not the concept is a success it is first necessary to define what the nature of success is and this can then be translated into particular indicators and finally more detailed measurements that can be recorded. The indicators shown in Table 2.12 are intended to act as at a broad level only - the exact indicators for use in a local scheme (or in a living lab) should be determined at the local level. However these broad indicators can be mapped into more specific measurements obtained using particular tools, as described further in §2.3.5 and §2.5.2.

Table 2.12 Translating success of social network concept into broad indicators

Social media activity and benefit	Social network functionality	Overall Indicators for success
System users benefit from the factual information concerning travel, contained in postings that the community may make	Participants join network. Participants read information posted by the network	• The number of participants joining by different routes and referrals • The number of posts relating to travel that the network participants make • The number of visits to the network sites that participants make (to read posts)
System users contribute to the value of the social network by placing their own posts	Participants post information and sentiments	• The number of posts (relating to travel or none travel issues) that the network participants make • Number and type of sentiments that participants make around travel or tripzoom/Sunset
System users receive very timely	Participants read and	• Number of specific alert type posts that

transport related alerts from community members who can provide more 'instant' information of the current traffic situation than may be possible for centralised information providers	post information within the network	participants make concerning travel • Number of specific alert- response type posts participants make
System users benefit from the positive encouragement of others with similar transport or lifestyle goals	Participants have a buddy group on the app and in the wider social network	• Number of buddies on app or in other social networks • Number of visits to buddy function on app • Amount of sharing information with app buddy group or push messages to social network • Longevity and level of activity of buddies on app or social network
System users can compare their behaviour with friends and colleagues	Participants can view their performance against the community and against their own goals. Participants are able to share their performance with a wider social network group.	• Number of visits to 'killer features' representing performance against own and community goals • Number of posts related to achievement of own goals • Number of posts concerning goals of others in networks
System users can interface their transport related social network activity with broader social networks activity and communities (i.e. those centred within Facebook, Twitter) to share their (positive) transport related behaviour with a wide set of social contacts	Participants can link the app buddy group and their wider social network group	• Number of push messages to social networks • Number of posts to non-sunset focused social networks • Sentiment analysis of posts to wider social networks
From the perspective of the city transport authorities, social media enabled social networks can also be used to recruit users into the system	Authorities are linked into social networks as a recruitment method	• Indicators discussed in more detail in §2.3.4 below.

2.3.4 Indicators for the success in recruitment

SUNSET is a novel concept that will be tested in three specific environments (in Enschede, Leeds and Gothenburg) and recruitment to the LL is therefore an issue for the SUNSET project and similar research or implementations. A short reflection on potential indicators that extend from the current state of the art and concern success around recruitment in particular are provided below.

Kraut *et al.* (2012) argue that to be successful, a social site must deal effectively with newcomers and encourage contribution. Knowing how people are recruited is an important means to get insight into the success of the recruitment with social media. Therefore it is important to measure how people are persuaded to go to the portal and/or download the app. This can for instance be by advertisements, Facebook mailings, Facebook friends and "ordinary" mailings. The number of people recruited by a certain medium/track hence should therefore be monitored and converted into an indicator. Besides that the chance that people are recruited in a "social" way is higher when people are more active on social media. Therefore it is important to keep

track of people's activity on those media (e.g. related to the number of friends and the activity in tripzoom). Also the sentiments expressed about tripzoom/SUNSET will make a difference in recruitment as well as providing useful content in their own right concerning satisfaction or otherwise with the system. The sentiments arising from user generated content should therefore also be monitored. Together this gives rise to the following additional indicators:

- Number of participants recruited via Facebook/Twitter mailing
- Number of participants recruited via friends on Facebook/Twitter
- Number of participants recruited via employer group
- Activity of participants (e.g. number of clicks, active days, posts on other media)
- Number of active/passive friends
- Sentiments expressed on the site concerning SUNSET/tripzoom

It is worth a final note of caution due to the slightly artificial nature through which some of the SUNSET networks will be generated and operate, a feature that arises as a result of the structured experimental recruitment process. In a real life implementation there would be a strong need to understand and evaluate effective processes to create, sustain and grow social network activity to the extent that would be appropriate for the local interpretation of the SUNSET scheme.

2.3.5 Measuring success

In order to be able to determine the success of the concept the indicators for the successful described at broad level in Table 2.12 need to be finally measured. Some of the indicators can be measured by the technical components of the software system that is developed within the SUNSET project. Others however can better be measured using external (social) media measurement and analysis tools. These tools give insight in the way people use social media, internet and applications. In this section we describe two of them: Google Analytics and Radian6.

Google Analytics (GA)

Google Analytics is a tool that gives insight in the internet use of (potential) customers or users. It can, for instance, give information about the number of people visiting the website, searching with certain keywords, 'liking' on Facebook, how people are surfing from one digital medium to the other and much more. This allows measurement of many of the above mentioned indicators and as a result, evaluation of the success of the concept. The indicators that are measurable with GA are:

- Number of clicks/visitors
- Posts/likes on social media
- Interactions across different digital media (e.g. search engines, project website, web-advertisements, e-mail, Facebook, Twitter, applications)

Radian6

Radian6 is a social network monitoring platform that gives insight into the online conversations about a certain topic. With Radian6 it is possible to obtain data on:

- The number of times a certain word (or phrase) is mentioned
- Where certain words are mentioned (which medium)
- The sentiments with which certain words are mentioned (in a positive/negative way)
- By whom certain words are mentioned

Radian 6 allows analysis of the content and quality of the social network text-based activity, in contrast to the usage based approach of Google Analytics. It only mines openly available posts

and does not go behind any privacy or pay walls. It offers considerable added value by vastly broadening the scope of SUNSET social media evaluation, for example by the inclusion of sentiments. However, it is worth noting that this software has a charge for use, so the use for evaluation should be judicious and well designed.

2.4 Functionality and usability evaluation

This section discusses the evaluation approaches during the development of tripzoom and the SUNSET living lab operations. Different evaluation methodologies should be applied at different stages of the process. The objective of the evaluation is to ensure that the functionalities and associated services of tripzoom are delivered in a reliable and user friendly way.

2.4.1 Iterative, user-centred design approach

To increase the probability that users actually will use a product when it is introduced in the competitive and open market, it must offer users an added value of some sort, and it must do so in a user-friendly way. The SUNSET project applies an iterative, user-centred design approach. User-centred design (UCD) is a design philosophy and a process in which the needs, wants, and limitations of end-users of a product are given extensive attention at each stage of the design process (UXPA, 2012). Ideally, this is elaborated by also focussing on end-user activities (Norman, 2005). The approach can be characterised as a multistage problem solving process that not only requires designers to analyse and foresee how users are likely to use a product, but also to test the validity of their assumptions with regard to user behaviour in real world tests with actual users. Applying the UCD approach in the product development process increases the probability that the product offers added value to users and will be more usable, leading to improved business metrics like conversion rate and customer loyalty (see for example Nielsen 2008).

The SUNSET user-centred design process consists of three main, chronological stages:

- A first stage aimed at discovering user needs and defining the user activities that will be supporting by the system
- An iterative, release-based approach in which most of the software development occurs, and the functionality and usability of each release is evaluated
- A living lab stage in which the system is applied and tested in a real-life context with end-users

Needs assessment

To be sure a product is based on the right requirements, user needs have to be discovered, and assumptions about users and what really matters to them have to be evaluated. Techniques for getting this knowledge include surveys, interviews and focus-groups. Users involved typically are recruited from project's target groups. As users can be uncertain of their needs, and can be unable to articulate them clearly (Pitts & Browne 2007), activities in user-centred design are carried out in an iterative manner. Through the iterations, developers and users learn more and more about different problems, hence, different solutions to the problems can be explored.

In SUNSET, user requirements were initially based upon a set of use cases (as described in D1.1). Subsequently, end-users were consulted to evaluate these requirements and state their preferences. An update of the user requirements and the derived system requirements are then presented in D1.2.

Release-based development and evaluation

During the design process, next to functional testing, prototypes have to be tested for being usable and delivering the right user experience. Even successful products may be far from being user-friendly. Surveys show that 75% of all users find their ICT tools more stressing than relaxing (EC-ISM European Commission 2009). Products that are user-unfriendly are more likely to lead to errors and customer stress and dissatisfaction, increasing the chance the product will not be used again or customers will switch to other comparable products (Norman 1990). Techniques for testing the prototypes include heuristic evaluation, cognitive walkthrough, usability tests and field tests (for an overview, see Nielsen & Mack 1994).

SUNSET is a research project and not a software development scheme. However, the scale of the living labs requires a near seamless user experience: actual end-users are going to make use of the system over a long period of time. That means good quality software and user interfaces, without bugs etc. That means a more strict software development process and software quality control mechanism.

SUNSET works with an iterative, release-based approach. The target is to produce a new version of the SUNSET system in a 5-week release schedule until the full deployment of the Enschede living lab. Typically in a 5-week schedule, the development period is then 4 weeks followed by an approx. one week period of two parallel activities: a) technical bug fixing and b) user level evaluation. In this evaluation, a group of test users (approximately 5 at each release test) test the functionality and usability of the current release, and provide feedback, priorities and requirements for the next release(s). In the 4-week development period, a mix of both technical components implementation and improvement and integration activities take place to create a working release for (test) users. To monitor progress use a release-control tool (Redmine) is used.

The focus of the user-level testing and evaluation of the releases is to:

- make sure the system is step-wise improving and extending towards a level which matches stakeholder expectations;
- provide a first check on missing functionality or usability issues prior to going live in a LL setting;
- create confidence in the added value of the system;
- create better understanding for the non-technical project team members in the technical design choices and existing challenges;
- provide input (directions, choices and requirements) for the technical team.

The main activity is to prepare and conduct small studies (using case-based assignments) with a set of test users, who test and evaluate the added features of the release in question. The application of which method/technique is appropriate is dependent on the functionality or experience that needs to be evaluated and can vary over the releases.

A summary of the rationale for using a pool of human evaluators for release evaluation is as follows. Early releases are evaluated internally with the SUNSET project team as a test user group. For this, a pool of evaluators is created with members at the different living lab sites. Depending on the specific evaluation task, their role is usability expert or (potential) end-user. In later releases, the test user team is extended with individuals external to the project who are in some way representative for the eventual targeted LL user groups, and are easily available to evaluators for feedback (in other words, mostly colleagues). In the last releases, the test user team is further extended with a limited number of external individuals, but now explicitly selected from the target groups. Outside the release-based process, and at an earlier stage of the project, potential end-users have also been evaluated using surveys and focus groups to assess their needs and gather opinions on the SUNSET concepts (see D3.3).

The evaluations are prepared and planned by a task team both familiar with the technical development and with evaluation techniques, and led by an evaluation coordinator. Using an evaluation protocol, the team defines a number of evaluation questions (e.g. are users able to manage their privacy settings), selects and executes the evaluation method, defines the test group, creates assignments for the test group, directs and guides the test users and collects the feedback. Finally, the task team creates a concise overall evaluation report.

Living lab evaluation

Evaluating products in real-life contexts is central to the idea of Living Labs (EC-ISM European Commission 2009). Several definitions of what a living lab is exist. Eriksson *et al.* (2005) define living labs as a research and development methodology whereby innovations, such as services, products, and application enhancements, are created and validated in collaborative, multi-contextual empirical real-world settings. All stakeholders in a product, service or application participate directly in the development process. In the SUNSET context, this for example includes end-users and municipalities. Others define a living lab as a real-life test and experimentation environment where users and producers co-create innovations (ENoLL 2012). Living lab experimentation strives for the same level of observation as is common in, for example, a usability lab, but in an organic, multi-contextual space. This means that customers participating in a living lab are observed across many aspects of their lives, such as their roles as citizens, workers, at home, travelling, and so on. As such, the living lab studies provide user feedback of high ecological validity.

One main characteristic of the SUNSET design process approach is that end-users are involved at every stage, thus increasing the chance that the ultimate product "fits" the end-user. Another characteristic is the iterative approach, providing constant feedback from an end-user's point-of-view to the developers, with corrections leading to product improvement. A final important characteristic is that the system is tested in a real-life context, which increases the ecological validity. So, the characteristics of the design process itself increase the chance that that the system will be a success regarding functionality and usability.

2.4.2 Evaluation techniques

For each of the stages described above, different evaluation techniques are appropriate. Regarding functionality, developers themselves have a basic role in testing whether what they developed works or not. When a function works at an acceptable level, people outside the development team have to get involved. These are either usability experts or potential end-users of the system.

The functionality the SUNSET system offers is a mixture of descriptions in the SUNSET DoW (Description of Work), feedback from evaluations, and feasibility (what was possible to develop during project time by the project members). A detailed description of the functionality is offered elsewhere. Related to functionality are design requirements which generally apply to any software, but which have a clear impact on the user experience. These include:

- Non-interference when used on users' own mobile – issues around battery, non-obtrusive with other apps
- Robustness –ability to work or at least not crash with network problems, ability to recover unobtrusively, ability to signal to a user that there is a problem to contact the developer
- Security – ability to maintain confidentiality and integrity of data
- Privacy – have sufficient procedures in the operation to support privacy

The evaluation techniques applied to the several releases are mostly aimed at functionality (does the release offer the functionality to the end-user as agreed within the project?) and usability (does the release offer the functionality to the end-user in an efficient, effective, and satisfying way?).

Regarding *usability*, two important standards regarding usability have been set, namely International Standards Organization (ISO) 9241-11 (1998) and ISO 13407 (1999). The former defined the meaning of usability and the latter provided guidance for usability design focusing on the UCD approach. ISO 9241-11 (1998) defines usability as "*… the extent to which a product can be used by specified users to achieve specified goals with effectiveness, efficiency and satisfaction in a specified context of use*".

As an attempt to measure the usability of a system quantitatively, several usability attributes have been defined based on product characteristics. For instance, ISO 9241-11 (1998) measures usability in terms of efficiency, effectiveness, and satisfaction. ISO 9126-2/3 (2003) added more usability attributes, namely understandability, operability, and attractiveness. Nielsen (1993) specified five usability dimensions, namely learnability, efficiency, memorability, errors, and satisfaction. Shackel (1991) measures the usability of a system with regard to its effectiveness, learnability, flexibility, and attitude. Thus, these publications have defined the usability measures (or indicators) in terms of effectiveness, learnability, flexibility, attitude, memorability, efficiency, satisfaction, errors, understandability, operability, and attractiveness. These indicators are shown in Table 2.13 and are explained below.

- Effectiveness: ISO 9241-11 (1998) defined effectiveness in terms of user's task accomplishment with accuracy and completeness
- Learnability: According to Nielsen (1993), learnability concerned with how easy it is to learn the system so that a user can immediately start carrying out his or her work with the system.
- Flexibility: It measures the easiness to adapt to various tasks and the freedom to use several commands to do a similar task (Shackel 1991).
- Attitude: Shackel (1991) described attitude as "*acceptable levels of human costs in terms of tiredness, discomfort, frustration and personal effort*". This also gives an indication of likeability of the system.
- Memorability: It is an indicator to measure how easy it is for a casual user to use the system again after a period of not using it (Nielsen 1993).
- Efficiency: It is an indicator of how efficient it is for a user to accomplish tasks once he or she has learned to use the system (Nielsen 1993). Thus, it is related to "*the accuracy and completeness with which users achieve goals*" (ISO 9241-11 1998).
- Satisfaction: It measures the pleasantness to use the system, including whether or not a user is subjectively satisfied when using the system (Nielsen 1993). ISO 9241-11 (1998) further included freedom from discomfort and positive attitudes towards the system as parts of the satisfaction indicator.
- Errors: This measures the number of errors that a user makes when using the system and the levels of ease to recover from errors when such errors are made (Nielsen 1993).
- Understandability: This indicator measures how easy/difficult it is for a user to know whether the system is suitable for his or her needs and can be used for particular tasks (ISO 9126-2/3 2003).
- Operability: It measures whether or not a user can operate and control they system (ISO 9126-2/3 2003).
- Attractiveness: It places emphasis on the appearance of the system, for instance the design and colour scheme used (ISO 9126-2/3 2003).

Nielsen (1993) further specified several methods and techniques that can be used to assess usability. These techniques are varied, depending on the design stages. For instance, heuristic evaluation, as an expert evaluation technique, is normally used in the early and during design processes. Other methods that directly involve users, such as observation and questionnaire can be done for the follow up studies.

Table 2.13 Usability methods

Method	Lifecycle stage	Users needed	Advantages	Disadvantages
Heuristic evaluation	• Early design • Inner cycle of iterative design	None	• Finds individual usability problems • Can address expert user issues	Does not involve real users, so does not find "surprises" relating to their needs
Performance measures	• Competitive analysis • Final testing	10 (min)	• Hard numbers • Results easy to compare	Does not find individual usability problems
Thinking aloud	• Iterative design • Formative evaluation	3-5	• Pinpoints user misconceptions • Cheap test	Unnatural for users Hard for expert users to verbalise
Observation	• Task analysis • Follow-up studies	3 (min)	• Ecological validity • Reveals users' real tasks • Suggests functions & features	Appointments hard to set up No experimenter control
Questionnaires	• Task analysis • Follow-up studies	30 (min)	• Finds subjective user preferences • Easy to repeat	Pilot work needed (to prevent misunderstandings)
Interviews	• Task analysis	5	• Flexible • In-depth attitude and experience probing	Time consuming Hard to analyse and compare
Focus groups	• Task analysis • User involvement	6-9/group	• Spontaneous reactions and group dynamics	Hard to analyse Low validity
Logging actual use	• Final testing • Follow-up studies	20 (min)	• Finds highly used (or unused features) • Can run continuously	Analysis programs needed for huge mass data Violation of users' privacy
User feedback	• Follow-up studies	Hundreds	• Tracks changes in user requirements and views	Special organization needed to handle replies

Source: Nielsen (1993).

2.4.3 Needs assessment

In SUNSET, user requirements were initially based upon a set of use cases (D1.1). Subsequently, end-users were consulted to evaluate these requirements and state their preferences. An update is then made to the user requirements and the derived system requirements (D1.2).

The end-user feedback can be gathered in two ways:

- Quantitative user data from large user groups applying surveys
- Qualitative user data from smaller groups applying focus-groups and interviews

A combination of these methods at various stages has been employed in SUNSET. Saunders *et al.* (2009) provide an overview of questionnaire based surveys (Table 2.14), highlighting the fit of

internet mediated surveys when focusing on computer literate individuals who are geographically dispersed. Both of these features are highly relevant to SUNSET, therefore an on-line questionnaire has been designed and used at the initial stage (D1.1) which was further supported by focus groups and interviews in each city prior to the launch of the respective living labs (D1.1 and D3.3).

Table 2.14 Main attributes of questionnaires

Attribute	**Internet- and intranet mediated**	**Postal**	**Delivery and collection**	**Telephone**	**Structured interview**
Population's characteristics for which suitable	Computer-literate individuals who can be contacted by email, internet or intranet	Literate individuals who can be contacted by post; selected by name, household, organisation, etc.		Individuals who can be telephoned, selected by name, household, organisation, etc.	Any, selected by name, household, organisation, in the street, etc.
Confidence that right person has responded	High if using email	Low	Low but can be checked at collection	High	
Likelihood of contamination or distortion of respondent's answer	Low	May be contaminated by consultation with others		Occasionally distorted or invented by interviewer	Occasionally contaminated by consultation, distorted or invented by interviewer
Size of sample	Large or small, can be geographically dispersed		Dependent on number of field workers	Dependent on number of interviewers	
Likely response rate	Variable, 30% reasonable within organisations or via intranet, 11% or less using internet	Variable, 30% reasonable		High, 50~70% reasonable	
Feasible length of questionnaire	Fewer 'screens' probably better	6~8 A4 pages		Up to half an hour	Variable depending on location
Suitable types of question	Closed questions but not too complex, complicated sequencing fine if uses IT, must be of interest to respondent	Closed questions but not too complex, simple sequencing only, must be of interest to respondent		Open and closed questions, including complicated questions, complicated sequencing fine	
Time taken to complete collection	2~6 weeks from distribution	4~8 weeks from posting	Dependent on sample size, number of fieldworkers, etc.	Dependent on sample size, number of fieldworkers, etc., slower than self-administered for same sample size	
Main financial resource	Possible web page design,	Outward and return postage,	Field workers, travel,	Interviewers, telephone,	Interviewers, travel,

implications	low if using email	photocopying, data entry	photocopying, data entry	calls, photocopying	photocopying
Role of the interviewer	None		Delivery and collection of questionnaires	Enhancing respondent participation, guiding the respondent through the questions, answering respondents' queries	
Data input	Manual or automated	Manual of automated			

Source: Saunders *et al.* (2009).

Such a design allows identifying user needs at an early stage based on a large sample including diversity regarding gender, age, experience with ICT use, or geographical location. It then builds up on the input from the on-line survey (D1.1) to explore in-depth the needs of individual users through focus groups and semi-structured interviews with stakeholders (D3.3). The latter turns such an approach into a mixed method approach which allows researchers to draw generalisations from a well identified sample to a population whilst gaining better contextual understanding of the population needs being assessed. Mixed methods have been defined as *"the collection or analysis of both quantitative or qualitative data in a single study in which the data are collected concurrently or sequentially, are given a priority, and involve the integration of data at one or more stages in the process of research"*(Gray 2009) and although there is no consensus regarding their definition among scholars yet, it is widely acknowledged that they can unlock a wealth of information to researchers when conducting research in multidisciplinary contexts such as the SUNSET research project.

Furthermore, interviews form an established method of collecting feedback and informing assessment. Since key issues have been explored through the on-line survey (D1.1) in SUNSET, the other two types of interviews have been employed. Structured interviews with stakeholders have been used in Enschede to describe and assess key points, whereas a semi-structured interview format has been used during the focus groups in Enschede and Leeds. Table 2.15 summarises the particular strengths of each type of interview.

Table 2.15 Strengths of different types of incentive

Type of interview	Exploratory	Descriptive	Explanatory
Structured	□	√√	√
Semi-structured	√	□	√√
Unstructured	√√	□	□

Notes: □ – not applicable; √ – less frequent; √√ – more frequent.
Source: Saunders *et al.* (2009).

Moreover, user input has been assessed through a series of focus groups to provide in-depth explanations about specific issues, due to the fact that focus groups offer the opportunity for participants to interact dynamically with each other, stimulating more vibrant and realistic discussions compared to conventional one to one interviews. *"When done well, focus groups offer powerful insights into people's feelings and thoughts and thus a more detailed, nuanced, and richer understanding of their perspectives on ideas, products, and policies"* (Jarvis and Barbarena 2008). Thus, this mixed method approach has the potential to contribute in contemporary methodological debates regarding focus group design and effectiveness given restricted resources (O'hEocha *et al.* 2011).

2.4.4 Release-based development and evaluation

During development, it is recommended to evaluate during the several iterations in the following ways:

- Functionality tested ad-hoc by developers
- Functionality tested by non-developers (in the role of potential users of the system) while executing a structured task list, notifying developers of bugs and not properly working functions
- Usability tested by non-developers in the role of usability experts while executing the same structured task list, applying evaluation techniques including heuristic evaluation and cognitive walkthrough. Usability issues are reported to developers.
- Functionality and usability evaluation with real end-users, applying evaluation techniques like think-aloud protocol, interviews, focus groups, field testing, experience sampling, surveys.

The method selected for the evaluation of the earlier releases was a hybrid of heuristic evaluation and cognitive walkthrough, both usability inspection methods.

Heuristic evaluation (Nielsen 1994) is a usability engineering method for finding the usability problems in a user interface design so that they can be attended to as part of an iterative design process. Heuristic evaluation involves having a small set of evaluators examine the interface and judge its compliance with recognised usability principles: the "heuristics" (they are called "heuristics" because they are more in the nature of rules of thumb than specific usability guidelines).

Based on a factor analysis of a large number of usability problems to derive a set of heuristics with maximum explanatory power, the following set of heuristics was established:

- *Visibility of system status* - The system should always keep users informed about what is going on, through appropriate feedback within reasonable time.
- *Match between system and the real world* - The system should speak the users' language, with words, phrases and concepts familiar to the user, rather than system-oriented terms. Follow real-world conventions, making information appear in a natural and logical order.
- *User control and freedom* - Users often choose system functions by mistake and will need a clearly marked "emergency exit" to leave the unwanted state without having to go through an extended dialogue. Support undo and redo.
- *Consistency and standards* - Users should not have to wonder whether different words, situations, or actions mean the same thing. Follow platform conventions.
- *Error prevention* - Even better than good error messages is a careful design which prevents a problem from occurring in the first place. Either eliminate error-prone conditions or check for them and present users with a confirmation option before they commit to the action.
- *Recognition rather than recall* - Minimise the user's memory load by making objects, actions, and options visible. The user should not have to remember information from one part of the dialogue to another. Instructions for use of the system should be visible or easily retrievable whenever appropriate.
- *Flexibility and efficiency of use* - Accelerators -- unseen by the novice user -- may often speed up the interaction for the expert user such that the system can cater to both inexperienced and experienced users. Allow users to tailor frequent actions.

- *Aesthetic and minimalist design* - Dialogues should not contain information which is irrelevant or rarely needed. Every extra unit of information in a dialogue competes with the relevant units of information and diminishes their relative visibility.
- *Help users recognise, diagnose, and recover from errors* - Error messages should be expressed in plain language (no codes), precisely indicate the problem, and constructively suggest a solution.
- *Help and documentation* - Even though it is better if the system can be used without documentation, it may be necessary to provide help and documentation. Any such information should be easy to search, focused on the user's task, list concrete steps to be carried out, and not be too large.

It is recommended to use three to five evaluators since one does not gain that much additional information by using larger numbers (Nielsen 1994). Heuristic evaluation is performed by having each individual evaluator inspect the interface alone. During the evaluation session, the evaluator goes through the interface several times and inspects the various dialogue elements and compares them with the heuristics. Only after all evaluations have been completed are the evaluators allowed to communicate and have their findings aggregated.

The *cognitive walkthrough* method (Wharton *et al.* 1994) is a usability inspection method used to identify usability issues in a piece of software or web site, focusing on how easy it is for new users to accomplish tasks with the system. Cognitive walkthrough uses an explicitly detailed procedure to simulate a user's problem-solving process at each step through the dialogue, checking if the simulated user's goals and memory content can be assumed to lead to the next correct action. The method is prised for its ability to generate results quickly with low cost, especially when compared to usability testing, as well as the ability to apply the method early in the design phases, before coding has even begun.

A cognitive walkthrough starts with a task analysis that specifies the sequence of steps or actions required by a user to accomplish a task, and the system responses to those actions. The evaluator then walks through the steps, asking him-/herself a set of questions at each step. Typically four questions are asked (Wharton et al. 1994):

- Will the user try to achieve the effect that the subtask has? Does the user understand that this subtask is needed to reach the user's goal?
- Will the user notice that the correct action is available? E.g. is the button visible?
- Will the user understand that the wanted subtask can be achieved by the action? E.g. the right button is visible but the user does not understand the text and will therefore not click on it.
- Does the user get feedback? Will the user know that they have done the right thing after performing the action?

Spencer (2000) proposed a Streamlined Cognitive Walkthrough technique in which one asks only two questions at the user during each action:

- Will the user know what to do at this step?
- If the user does the right thing, will they know that they did the right thing, and are making progress towards their goal?

By answering the questions for each subtask usability problems will be noticed, and afterwards a report of potential issues is compiled.

Whereas cognitive walkthrough is task-specific, heuristic evaluation takes a holistic view to catch problems not caught by this and other usability inspection methods. In a hybrid approach, task scenarios are combined with heuristics. So, when walking through the tasks step-by-step through the lens of the user (what terms they use, the things they'd look for and likely path's they'd take), the evaluator looks for and identifies problems based on the set of heuristics.

Other approaches have been also considered within this context, namely Gerhardt-Powals cognitive engineering principles which are quite similar to those of Nielsen and linked with cognitive ergonomics, yet take a more holistic approach focusing on the user interface. *"Cognitive engineering of a human-computer interface is the leveraging of empirical findings from the cognitive sciences and application of those findings to the design of the interface" (*Gerhardt-Powals 1996). Through the comparison of three different interfaces in that research, it was concluded that there is a strong correlation between a cognitive friendly interface and the utilisation of a set of critical cognitive design principles. In addition, Weinschenk and Barker's (2000) classification has also been considered, but not used due to the high number of attributes and strong duplication with Nielsen's heuristics. Less popular methods such as formal usability inspections or the pluralistic walkthrough have been also reviewed but not explored further since those have now been incorporated in other approaches or abandoned by practitioners and researchers (Novick & Hollingsed 2007). Hence, due to the similarities between Nielsen's attributes and those of both Gerhardt-Powals and Weinschenk and Barker, as well as due to the fact that they are not mutually exclusive, it has been concluded that Nielsen's usability attributes have the best fit with the SUNSET context, restrictions and resources.

During later releases of the tripzoom app, external users will have an increasing degree of involvement in the evaluation. For specific tasks, such as the installation of the app, the think-aloud technique is applied. Think-aloud is an observation technique, in which users verbalise everything that comes to their minds while carrying out certain tasks. Even though it has its origins in problem solving theory (Ericsson & Simon 1984), the technique has become a common approach for user research. Analysing the transcriptions of verbalisations gives a detailed insight in a user's thoughts while interacting with a product or service.

Users are repeatedly interviewed about specific functions and their usability after they have the application for a while in their real-life context (making it a field test). It will generate further added value to bring a number of end-users together in a focus-group session, where they will have the opportunity to interact dynamically with each other, stimulating a different level of discussion compared to conventional one to one interviews.

2.4.5 Living lab evaluation

The transition from the release-based development to the living lab evaluation is an elongated process. The end stage is the experimental set-up in which end-users are assigned to specific experimental groups. These groups are detailed in Chapter 3, where each user group features are explained, along with the particular incentives to be offered to them and during which time period.

Evaluation techniques applied during this stage include the think-aloud technique, interviews, focus groups, field test, experience sampling, and surveys. To measure users in a real-life context, experience sampling is a fitting technique. Experience sampling (Consolvo & Walker 2003) involves taking multiple measurements over time that require little effort from the user. Data can be either qualitative or quantitative. No researcher involvement is needed, apart from possibly short interviews before and after the experience sampling data collection. An advantage of this technique is the direct coupling with real-life behaviour. Disadvantages include the fact that it normally requires a strong motivation on the part of the participants, and that it can be very

obtrusive. This has also been reported in D3.3, where focus group participants stated that both the volume and timing of experience sampling questions may prove critical to the continuous use of tripzoom.

These issues with the participants have been acknowledged by D7.1 where the objectives of experience sampling questions have been clearly defined (restricted). Therefore, due to the research objectives of SUNSET it is crucial to utilise this method wisely during the living labs, since it will also act as a channel to target specific user groups and evaluate their travel experience at real time.

In the experimental phase, travel diaries are also used, to establish travel patterns during time periods where there is no mobility feedback via the app. The diary study is a technique of understanding participant behaviour and intent *in situ* that minimizes the effects of observers on participants (e.g. Tomitsch *et al.* 2010). Diary studies can be broken down into those that use media captured by participants as prompts for discussion in interviews (elicitation studies) and those that require participants to answer predefined questions about events (feedback studies). In SUNSET, the latter approach is adopted. Pros for this technique are that it gives detailed insight over a period of time, that there is a strong relationship with real-life situations, and that it can be used remotely. Cons for this technique are the reliability of retrospective and subjective observations, and that much attention and motivation is required from the participant.

In general, it is anticipated that through the combination of experience sampling questions and travel diaries for certain user groups, a wealth of data will be generated, including stated preference (a-priori), revealed preference (a-posteriori), as well as real time qualitative evaluation of travel experiences. Experience sampling also has the potential of creating a direct channel of communication between users and SUNSET LL coordinators.

2.5 Summary and recommendations for SUNSET

2.5.1 The system meeting individual travel goals

As described in §2.2, there are several travel goals that individual travellers may wish to gain when making a trip, such as to minimise or maximise travel time, costs, and pleasure. However, those goals are also very much dependent on other trip conditions such as distance to destination (e.g. short or long distance trips) and trip purpose (e.g. work or leisure). Moreover, changing travel behaviour does not necessarily show an improvement on all indicators, however the overall utility may have improved. To understand how well the tripzoom app can help individual travellers achieve their desired travel goals, those personalised objectives should initially be identified. Subsequently, tripzoom users will be asked about their perception on how well tripzoom app helps in obtaining those goals. The operational success is than defined as the extent to which tripzoom helps the individual in reaching the individual's personal travel goals.

To be able to evaluate how well the app facilitates users in obtaining their travel goals, a set of indicators are proposed. These indicators have been detailed in §2.2 and are further listed below.

Proposed indicators

The Key Performance Indicators (KPIs) in this context are those indicators that best represent the definition of operational success of the SUNSET system in terms of helping the individual in reaching its personal travel objectives. This is best represented by the subjective indicators. These indicators are based on the opinions of individuals using the SUNSET system for their daily travels.

The extent to which individuals indicate tripzoom helps them to reach their individual travel goals marks the operational success.

Key Performance Indicators for operational success

- Subjective indicator of travel time by trip purpose and distance
- Subjective indicator of scheduling effort by trip purpose and distance
- Subjective indicators on distance
- Subjective indicators on costs
- Subjective indicators of social networks (e.g. motivational support, feedback, and satisfaction)
- Output (the added value of tripzoom app in relation to each of the desired mobility)

In order to evaluate the operational success based on the above mentioned KPIs a number of additional indicators are required to scrutinize and explain the results based on groups of individuals

Prerequisites

- Indicators on socio-demographics and household characteristics (e.g. age, gender)
- Preferred mode and type
- Accessibility to modes
- Mobility constraints
- Number of trips by trip purposes (work related or leisure)
- Number of trips by distance (short vs. long)
- Total kilometres travelled
- Total travel time by trip purpose and distance
- Objective indicators on social networks (e.g. number of friends)
- Attributes related to travel dislike
- Attitudes on pro-environmental policy
- Attitudes on commute benefit
- Attitudes on travel freedom
- Attitudes on pro-high density
- Attitudes on travel stress
- Personality
- Lifestyle
- Excess travel
- Desired mobility (with regard to travel goals by trip purpose and distance)

2.5.2 The success of the concept

In §2.3 an overview has been given on indicators that are commonly used to define the success of social media and a consideration of SUNSET-specific evaluation requirements is provided. Based on that and the findings from §2.3, it is possible to define more detailed indicator measurements for use within SUNSET to assess the degree of success of the social media concept.

Proposed Indicator measurements

- The number of unique visitors to the web portal
- The number of people that register for tripzoom
- The number of people that agree to participate in a living lab
- The number or percentage of participants that are recruited using a LL/city-FB site
- The number or percentage of participants that are recruited using a FB-add

- The number or percentage of participants that are recruited via a friend on an external social network
- The number or percentage of participants that are recruited via a friend invite by email
- The number or percentage of participants engaged in the tripzoom social network
- The number or percentage of participants linked in as 'friends' in the tripzoom social network either actively or passively
- The number of mutual friends within the local tripzoom social network
- The number of messages posted on external social media that relate to tripzoom
- The percentage of positive messages on social media that relate to tripzoom

Key Performance Indicators in the context of the social media concept are the indicators about the amount of friends in tripzoom. This indicator notifies the essence of the social media concept of SUNSET in terms of connecting people via social networks in the transport domain.

Some of the above indicators (e.g. the total number of people registering for tripzoom), can be measured using SUNSET-components. Others however have to be measured with other tools. A large part of the data collection can be achieved by linking both the app and the web portal to Google Analytics. This will allow monitoring of usage statistics of the app (only where specific anchors are built in) and the portal (all pages tracked automatically). Most importantly, there is a need to monitor where people originated when they visit the website, for example to track whether people register through the portal or on the app. This is the analysis that will also detect basic and objective engagement/use of the application.

Data availability from the portal

- Visits: number of visits, unique visits, visit duration, sources (traffic)
- Demographics: language, countries, cities
- System: browser, OS, service provider
- Mobile: OS, service provider, screen resolution
- Invites: number of new users (friends) invited using mail/Facebook

Data availability from the application

- Friends: number of clicks, friend search, invites
- Landing page: number of registers and logins
- Killer feature: number of clicks, features viewed
- Me: number of clicks, pages loaded, refreshing, swiping/sorting behaviour
- Settings: number of clicks, pages loaded, number of changes made

The sentiments and contents of messages relating to tripzoom will be measured using Radian6.

2.5.3 Functionality and usability evaluation

For a successful product with respect to the functionality and usability from an end-user's point-of-view, it is recommended that a design approach is applied with the following characteristics:

- End-users are involved during every stage of development, providing feedback to concepts and prototypes regarding functionality and usability
- The design process proceeds in iterations
- Evaluation (partly) occurs in real-life contexts, increasing ecological validity

For discovering user needs and evaluating concepts in an early stage of development, it is recommended that feedback is gathered in two ways:

- Quantitative user data from large user groups applying surveys
- Qualitative user data from smaller groups applying focus-groups and interviews

During development, it is recommended that evaluation is undertaken with several iterations in the following ways:

- Ad-hoc functionality testing by developers
- Functionality testing by non-developers (in the role of potential users of the system) whilst executing a structured task list, whereby developers are notified of bugs and not properly working functions
- Usability testing by non-developers in the role of usability experts and executing the same structured task list, by applying evaluation techniques such as heuristic evaluation and cognitive walkthrough, whereby usability issues are reported to developers
- Functionality and usability evaluation with real end-users, by applying evaluation techniques such as think-aloud protocol, interviews, focus groups, field testing, experience sampling, and surveys

The system is successful from a functionality perspective if the requirements set by the technical work packages in the SUNSET project are met. Inevitably these requirements will be adjusted during the development process. The functionality that the SUNSET system offers is then a mixture of those functions described in the DoW, feedback from evaluations, and feasibility (i.e. whether it was possible to develop during the project lifetime). The following recommendations summarise feedback from the real-life evaluation of different parts of the SUNSET system during the design stage, including the app, the city dashboard (used as a research tool to issue incentives) and the web portal. The evaluators who generated the feedback were members of the research team, with the app, city dashboard and web portal all being used in a purely experimental context. During the early stages of testing in the LL (with a small group of 'friendly' testers) further feedback will be taken which may add to these recommendations and which will be reported as part of WP7.

Regarding basic, general functionality that apply to most apps, it is recommended that the system should provide acceptable levels for the end-users regarding:

- Battery usage of the mobile phone
- Interference with other uses of the mobile phone
- Robustness –ability to work or at least not crash with network problems, ability to recover unobtrusively, ability to signal to a user that there is a problem to contact the developer
- Security – ability to maintain confidentiality and integrity of data
- Privacy – have sufficient procedures in the operation to support privacy

The extent to which acceptable levels have been reached should be evaluated by the end-users, who can give scores on acceptability using a psychometric scale. How acceptable levels of security and privacy from an objective point-of-view are supported is described elsewhere.

Specific to the web portal (which has been designed for use by LL participants), it is recommended that:

- The information on the website provides end-users with an understandable view on what SUNSET is about, and what registering and installing the app will mean to them
- The website provides helpdesk/FAQ to support users in the interaction with the system
- The website provides clear and usable instructions for how to join

- The process of joining (registering, installing the app) should be user-friendly
- Interaction login using other social media (Facebook, Twitter) is provided in a user-friendly way
- Users can communicate with other users in a user-friendly way

Specific to the tripzoom app, it is recommended that from and end-user's point-of-view:

- Interaction with the app is user-friendly (easy to learn, efficient, effective, satisfying) regarding all specific functions
- The functionality that the app offers provides added value to the user
- Interaction login and account links when using other social media (Facebook, Twitter) is provided in a user-friendly way
- Trips are detected with acceptable accuracy (locations, routes, modalities)
- Communication regarding inaccuracies is presented in a user-friendly way
- Overviews of travel behaviour are presented in a user-friendly way
- Comparisons of the user's travel behaviour with the travel behaviour of other users is presented in a user-friendly way
- Any communication regarding challenges is handled in a user-friendly way (see also Chapter 3 of this document)
- Any communication regarding experience sampling questions is handled in a user-friendly way
- The user has control over settings regarding privacy
- Interactions with friends is supported in a user-friendly way

Specific to the city dashboard (which was developed as research tool and will not be used by LL participants or LL city stakeholders), it is recommended that:

- Challenges can be issued to specific groups of users in order for the challenges to be specific and relevant.
- Responses to challenges can be collected and analysed in order to better understand how to make challenges relevant and effective in improving local transport strategies. Experience sampling questions can be issued to specific groups of users.
- Responses to experience sampling questions can be collected and analysed in order to improve local transport strategies.

For the portal, app, and city dashboard, it is recommended that the following usability issues are addressed:

- Visibility of system status — the system should always keep users informed about what is going on, through appropriate feedback within reasonable time.
- Match between system and the real world - the system should speak the users' language, with words, phrases and concepts familiar to the user, rather than system-oriented terms. It should follow real-world conventions, making information appear in a natural and logical order.
- User control and freedom — users often choose system functions by mistake and will need a clearly marked "emergency exit" to leave the unwanted state without having to go through an extended dialogue. It should support undo and redo.
- Consistency and standards — users should not be confused as to whether different words, situations, or actions have the same meaning. The system should follow platform conventions.
- Error prevention — even better than good error messages is a careful design which prevents a problem from occurring in the first place. It is desirable to either eliminate

error-prone conditions or check for them and present users with a confirmation option before they commit to the action.

- Recognition rather than recall — the system should minimise the user's memory load by making objects, actions, and options visible. The user should not have to remember information from one part of the dialogue to another. Instructions for use of the system should be visible or easily retrievable whenever appropriate.
- Flexibility and efficiency of use — accelerators (unseen by the novice user) may often speed up the interaction for the expert user such that the system can cater for both inexperienced and experienced users. The system should allow users to tailor frequent actions.
- Aesthetic and minimalist design — dialogues should not contain information which is irrelevant or rarely needed. Every extra unit of information in a dialogue competes with the relevant units of information and diminishes their relative visibility.
- Help users recognise, diagnose, and recover from errors — error messages should be expressed in plain language (not codes), precisely indicate the problem, and constructively suggest a solution.
- Help and documentation — even though it is better if the system can be used without documentation, it may be necessary to provide help and documentation. Any such information should be easy to search, focused on the user's task, list concrete steps to be carried out, and not be too large.

3. Evaluation approach for the effectiveness of incentives

In this chapter we focus on the changes in individual travel behaviour as a result of the use of incentives. This means we will zoom in on evaluating the success of incentives in terms of changes in travel behaviour as a result of offering incentives. A large number of indicators described in chapter 2 are also valid to measure the effectiveness of incentives. However, in this context they are used to indicate to what extent the individual travel behaviour has changed (not according to personal travel goals but according to external factors i.e. incentives such as challenges with or without rewards) instead of the extent to which tripzoom helps to achieve the personal travel goals.

3.1 Introduction

The SUNSET system aims to fulfil the following set of objectives relating to the system, the environment and the individual:

- *Congestion reduction: traffic-jams are an increasing problem to tackle. The average travelling times should be reduced. Our objective is 5% less traffic (measured in car kilometres in a specific area) during the rush hours for users of the SUNSET system.*
- *Safety: people must be able to optimise their route, to avoid roads with many cyclists for car drivers, to report local road and weather conditions within community, to detect unusual conditions, or to avoid waiting times on dark and silent railway stations.*
- *Environment protection: for a liveable climate we need reduced CO2 emissions, improved air quality management and reduced noise pollution.*
- *Personal wellbeing of citizens: the system allows individuals to set and monitor personal objectives, like increase individual safety, reduce travel times, reduce costs, improve comfort, and increase health.*

In order to achieve the above objectives, the tripzoom app has been developed with the ability to offer various kinds of incentives to travellers aimed at encouraging smarter choices around travel. These incentives were defined and specified in D3.3, are outlined in Table 2.3 of this report and are given in more details in Table 3.1 below.

Table 3.1 Possible incentives for tripzoom

Type of incentive (From D3.3)	Description of incentive	Characteristics of incentive
(1) Real-time travel information provided by the system	• The system gives information about the most recent conditions on the road networks. • The system gives alerts to users whenever there is a relevant event (either expected or unexpected) that may influence their travel behaviours. In addition: a) Users should be able to enable/disable the	• Duration of incentive: It should be made available 24/7. • Time and frequency to offer/remind user about the incentive: a) Regarding the real-time information on a map, users should be able to set the reminder manually. b) Regarding alerts on expected sporadic events (e.g. road works), an alert should be given to users 1-2 days before the event. Users should be able set manually how often they want to

	incentive; b) When enabled, alerts can only be given based on regular activity-travel patterns (i.e. related to the spatial parameter). Therefore, when a traveller uses a new route for the first time, alerts related to that route will not be available.	be reminded. c) Regarding alerts on unexpected events (e.g. traffic congestion & accidents), an alert should be sent immediately whenever the event occurs.
(2) Social networks for peer-to-peer travel information/messages	The system provides an infrastructure for users to exchange messages among each other. In general, there are two types of messages: • Alerts related to the conditions on the road or infrastructure • Tips/advice on travel	• Duration of incentive: It should be made available 24/7. This means users can post messages at any time and can read old posted messages. • Time and frequency to offer/remind the user about the incentive: a) Alerts related to the conditions on the road or infrastructure should be given to users who may be affected by a message. Based on the results of the empirical work in D3.3, users prefer only relevant information. Therefore, the user should specify the spatial and temporal parameters and the category of the information to share (i.e. alerts on road condition and tips/advice) in every message. This way, massages containing alerts can be sent to users who are likely to be affected by the information. b) Tips/advice on travel should be offered to all users whenever a message containing new tips/advice appears.
(3) Feedback based on self-monitoring of own travel behaviour	The system records users' daily activity-travel patterns and present the recorded information to users. This is the basic incentive and feature of tripzoom app.	• Duration of incentive: It should be made available 24/7. • Time and frequency to offer/remind the user about the incentive: Users should be able set manually how often they want to be reminded to check their recorded patterns.
(4) Feedback based on setting targets	The system allows users to set their own travel targets.	• Duration of incentive: Users should be able to set their targets for a definite time period (e.g. 1 week). • Time and frequency to offer/remind user about the incentive: When a target is set, the user can get a daily reminder in the morning of their performance in relation to their target.
(5) Challenges (using points without an exchange value)	Every user who exhibit certain travel behaviours (e.g. cycling or walking) will be awarded points. This can be related to a competition with other users based on points (akin to on-line games)	• Duration of incentive: It should be made available 24/7. • Time and frequency to offer/remind user about the incentive: Not applicable. User should be able to find the information related to how points can be collected in the help menu.
(6) Challenges (using	This category is related to:	• Duration of incentive:

points with an exchange value)	• Challenges set by the system or by the 3rd parties. • Periodic offers akin to a loyalty card. For instance, once a user reaches 100 points, he or she can redeem the points to a tangible reward.	a) A challenge should last for a period of time (e.g. 1 week). b) The loyalty scheme should be offered for a longer period (e.g. 6 months or 1 year). • Time and frequency to offer/remind user about the incentive: a) Users should get news about a new challenge whenever it is introduced. Users should be able set manually how often they want to be reminded. b) The loyalty scheme should be offered for a longer period (e.g. during the whole LL period). Users should be able set manually how often they want to be reminded.
(7) Social networks for sharing location	Every user can share their current location to selected users.	• Duration of incentive: Depending on users: users should be able to choose to turn on/off their location sharing. • Time and frequency to offer/remind user about the incentive: Not applicable.
(8) Social networks for finding a buddy	Every user can use find a buddy to find a travel companion.	• Duration of incentive: Depending on users: users should be able to choose to turn on/off find a buddy. • Time and frequency to offer/remind user about the incentive: Not applicable.
(9) Social networks for treasure hunt	This is also another type of challenges. For example, treasure (in the form of points) can be hidden in a specific coordinate and can only be unlocked whenever users cycle pass the coordinate.	• Duration of incentive: Depending on users: users should be able to choose to turn on/off treasure hunts. When it is enabled, users should be able to detect treasures within his or her proximity. • Time and frequency to offer/remind user about the incentive: Not applicable.

This chapter discusses indicators and methods/techniques suitable to evaluate the success of the incentives, as offered within the scope of the SUNSET system. To do so, the notion of a 'successful incentives' should firstly be defined.

A 'successful incentive' is characterised by a positive behavioural change in any individual that coincides with the offering of a related incentive. A positive behavioural change can be interpreted as one which has the potential to make a positive contribution to the achievement of system goals. However, to avoid duplication between the scope of D6.1 and D6.2, this deliverable will focus on methods to evaluate changes in individuals' travel behaviour arising from the use of incentives. D6.2 will report on methods to evaluate the impact of incentives at the higher levels, i.e. to accomplish the SUNSET system related objectives and the contribution to city and EU goals. It is worth noting here that 'travel behaviour' includes a number of travel decisions such as the transport mode, route, departure time, location, and trip decision (i.e. whether or not to make a trip).

Various measurements of the degree of success for the incentives can be defined using several indicators. Those indicators should take into account aspects in behavioural change, as

indicated in the Trans-theoretical Model of Behaviour Change (Prochaska & DiClemente 1983; Prochaska & Velicer 1997). This theory argues that behavioural change involve progress which is made in stages, as listed below. The relations between those stages can be seen in Figure 3.1.

- Pre-contemplation: In this stage, an individual is not ready and not intending to change their behaviour in the near future and may not be aware that their behaviour is problematic;
- Contemplation: In this stage, an individual is ready to change their behaviour as they are starting to realise that their behaviour is problematic and to weigh the pros and cons of their behaviour;
- Preparation: In this stage, an individual is intending to change their behaviour in the near future or has started to take small steps toward the change;
- Action: In this stage, an individual is performing the new behaviour as a modification to their previous (problematic) behaviour;
- Maintenance: In this stage, an individual has performed the new behaviour for a while and is working to sustain their new behaviour to prevent relapse;
- Termination: In this stage, an individual is no longer tempted to return to the problematic behaviour.

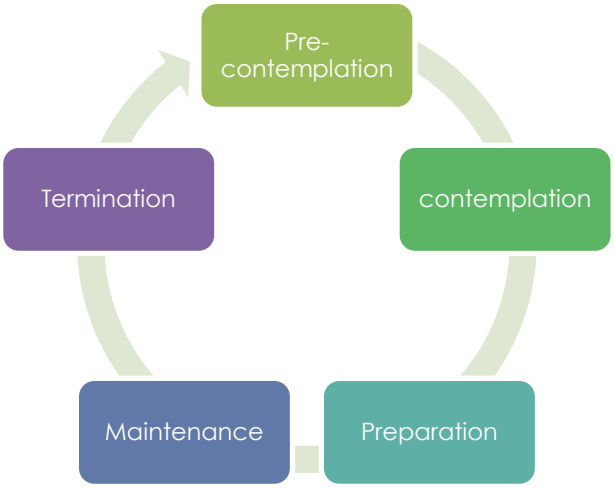

Figure 3.1 Trans-theoretical model of behaviour change
Source: Prochaska & DiClemente (1983)

In the transport research field, several voluntary behavioural change studies (such as IndiMark in Germany and TravelSmart in Australia) have indicated that changes in behaviour are made one step at a time. Therefore, a car user may shift from and to the stages below before arriving at desirable behaviour:

- Pre-contemplation to contemplation: The incentive(s) result in car users starting to realise that frequent car use is problematic;
- Contemplation to preparation: The incentive(s) result in car users intending to change their car use behaviour;
- From preparation to action: The incentive(s) result in car users reducing their car use and increase the use of the transport mode alternatives;
- From action to maintenance/termination: The incentive(s) result in car users becoming regular users of other transport modes in a long term.

Given the stages above, indicators to assess successful incentives should not only be determined by the observable changes in travel behaviours (i.e. changes at the action stage), but also include changes in attitude/intention. People who do not change their travel behaviour during the period of incentives may still be moving from the pre-contemplation stage to preparation stage. This changes can only be measures by changes in people's attitude/intention, such as intention to use transport mode alternatives, working at home, trip chaining, and changing locations (e.g. grocery shopping) to reduce travel distance and time. Those indicators should also take into account the time period to retain the new behaviour (i.e. the maintenance or termination stages).

For the evaluation of incentives, the design aspects of incentives should also (in principle) be taken into account, such as the number of times an incentive(s) is offered before a change in behaviour is observed, the number of times the reminders are given to users before behavioural change, and combinations of incentives to be offered to users. However, taking into account the design aspects of incentives may not be straightforward. In the SUNSET system (and potentially in similar applications), the user has the option to manually adjust these settings based on their preferences. This freedom to manually set preferred incentives has been included as a feature as it may attract new users to use the tripzoom app. For example, a user should be able to set whether they would prefer to get a daily or a weekly reminder on challenges (a form of incentive), or even to turn off the reminder related to challenges. Additional design aspects of the incentives are summarised in Table 3.2 below, covering issues such as which incentives to give, when, and how to give them, to which target groups and why. These aspects may impact on the way the data should be grouped before some analyses are carried out.

Table 3.2 Design aspects of incentives

Issues	Design aspects
What/which incentives	24/7 incentives: Which incentives for which target groups
	Periodic incentives (by LL coordinators): which incentives for which target group
	Individuals' incentive preferences
When to give incentives and how often	When and how often to send a reminder about an incentive to users
	When and how often to offer a new incentive to users
	Individuals' preferences on when and how often to be offered incentives or to be given a reminder
To which target groups	Age
	Location (home area and work area)
	Modes available
	Household
	Children
	Use of specific mode
	Use of specific road segment
	Use of a specific road segment on regular route
	Trips at specific time
	Go to the same work area (including car parks)
Why a certain incentive(s) is offered	To address certain system goal(s)
	To address certain individual goal(s)

Finally, the methodology to evaluate incentives should be sufficiently detailed so that data requirements for evaluation can be understood. It would also be desirable for the evaluation to be undertaken in such a way as to hold some potential for extrapolation to the population under study. In this respect different sources of errors and variability are also addressed in this chapter.

The remainder of this chapter is structured as follows. Identification of the indicators will be presented in §3.2. Subsequently, reviews of several evaluation methods and techniques are provided in §3.3. In §3.4, data requirements and errors will be addressed. Finally, §3.5 will summarise the recommendations for SUNSET in terms of indicators, methods, and data requirements.

3.2 Indicators for mobility behaviour and attitude

The SUNSET incentives are expected to evoke changes in both actual behaviour and attitude of travellers. The incentives include both indirect persuasion (for example by reflecting personal scores against community scores in the 'killer features') and direct persuasion methods (through the challenges issued by city dashboard operators). According to Aleahmad (2008) the former would be expected to result in behavioural change whilst the latter would result in attitude change. Changes in actual behaviour can be measured by changes in travellers' mobility profile. This include changes in amount/frequency/type of trips, shift in timing, mode and route choices, as well as changes in en-route behaviour such as driving behaviour (e.g. speed, car following).

Changes in attitude relate to travellers' perception and opinion of their mobility. Incentives can raise travellers' awareness on the environmental impact of their mobility and the existence of alternative travel options. Incentives may also affect travellers' perception and satisfaction with the transport services. Changes in the decision making process then take place whereby travellers consider a wider set of travel alternatives, take into account more factors in the choice, and/or adjust their weight of each selection criterion within the decision made.

Changes in attitude are believed to be the motivation for changes in actual behaviour. In other words, a change in actual behaviour is the manifestation of a change in attitude. Changes in attitude do not always invoke changes in behaviour – either because the change in attitude is not significant enough or because the travellers are bound by other constraints (for example physical constraints). In either case, if the situation of the traveller changes, behavioural changes may eventually take place. An example is that of parents who currently drive their children to school and then drive to work. Incentives raise their awareness on the environmental impact of car traffic but no change in behaviour takes place - due to the constraint of needing to travel with children and arrive at a particular destination at a particular time. However, if driving the children to school is not an essential part of the journey (any more), the parents may then decide to take a non-car mode for commuting. The overall implication for the SUNSET system is that in assessing the effectiveness of incentives both changes in attitudes and in behaviour should be monitored.

3.2.1 Behavioural indicators: overall measurements

Measuring change in travel behaviour starts with describing how travel decisions are made and what factors are important. The actual travel behaviour of an individual is subject to the interrelation between the needs and wishes of an individual, the activities/destinations available in a spatial context and the level-of-service of the transport system. This interrelation is shown in Figure 3.2. Travel behaviour is a result of an individual's desire to participate in an out-of-home activity, satisfying needs, wishes and obligations (Axhausen and Gärling 1992). For a particular household the latent travel demand thus depends on the environment (spatial distribution of activities and transport system) of the household. The actual traffic conditions and the preferences of the individual result in the actual travel choice, expressed in terms of trips with their accompanying characteristics.

The SUNSET incentives mainly target the trip planning/scheduling behaviour of travellers. They are expected to have negligible impacts on the en-route behaviour of travellers. The exception to this is the possibility of influence by the 'alerts and hazards' incentive where (depending on the length of the trip and timing of the incentive) this may result in adaptation of the route when the trip is already underway. Here we shall restrict our attention to travellers' trip-level behaviour, represented by the travellers' mobility patterns. A mobility pattern can be characterised by the distribution/profile of trip making over time and space. Changes in mobility pattern are then quantified by the variation in mobility pattern from a 'before' period to an 'after' period (or, treatment and no treatment periods), relating to when individuals were offered or exposed to the incentive.

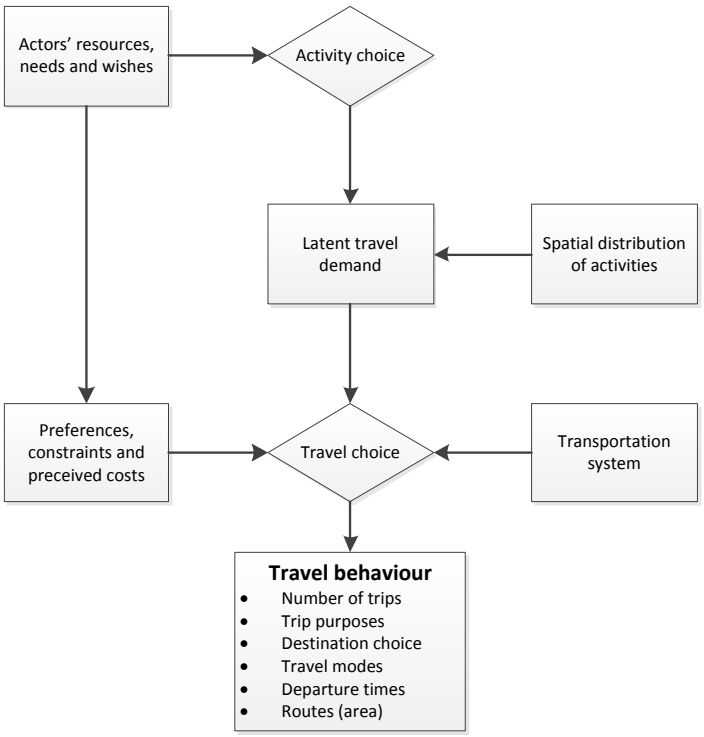

Figure 3.2 Conceptual framework of travel behaviour
Source: Gärling *et al*. (2000), Næss (2003)

The overall measurement of a person's trip making behaviour is to look at the total amount of trips within a time period (e.g. per month, per week). This can be measured in various terms, e.g.:

- Count: the total number of trips
- Distance: the total distance travelled
- Travel time: the total length of time spent *en route*
- Cost: the total amount of money spending on trip making
- Emissions: the total amount of CO2/pollutant emission

Amount of trips
Measuring the amount of trips is straightforward. It can be done by simply counting the number of trips that a traveller has made over the counting period. In practice, however, problems may arise regarding the trip counts. Consider the following three examples:

- Example 1: A traveller first drives the child to school and then continues to the office.

- Example 2: A traveller is driving on the highway and decides to pull over to the service area to take a rest, because there is currently a standstill on the highway. After resting for 15 minutes (or 30 minutes), the traveller continues with the trip.
- Example 3: A traveller cycles from home, first to the post office (to mail a letter), then to the supermarket (to buy some bread), and afterwards back to home.

It is debatable whether the cases in Example 1 and 2 should be considered as one trip or two trips. In both cases there is an intermediate break of journey which also serves a certain purpose (although it can be argued that in the second case this purpose is not so strong or voluntary). For the case in Example 3, it may be categorised as a 1-/2-/3-trip journey, depending on the adopted definitions of what constitutes a trip.

As can be seen, the "trip", as the unit of measurement, is not well defined. For evaluating behavioural changes, unintended biases may be introduced if the number of trips is used as the sole behavioural indicator. Consider the scenario of an incentive which motivates people to better organise their mobility. As a result of this incentive, a traveller now behaves as described by Example 3 above, whilst previously the same traveller would make two independent journeys – one to and from the post office and the other to and from the supermarket. Depending on the definition of a trip, the changes in the number of trips before-after may be calculated as 2-1, 2-2, or 4-3. It is therefore recommended that trip count should be used in combination with e.g. distance/time/cost measurements, in order to capture the user's mobility pattern in a more representative way.

Distance travelled
Measuring distance travelled is also straightforward. Similar to the mileage of a motor vehicle, the mileage of a traveller is derived by summing up the distance of each trip in the counting period. The distance of a trip is measured along the actual route, not the straight-line distance, between the origin and the destination of the trip. Travellers with the same origin-destination pair may therefore have different trip distances, should they take different routes.

Time travelled
Time spent in travelling can be measured similar to the travelled distance, by summing up the time spent in each trip during the counting period.

It should be noted, however, that not all times are equivalent. It may be necessary to distinguish free flow travel time and delay time (or lost time, time in congestion). The former is more acceptable to the traveller while the latter is usually perceived as a waste of their time. Another point of attention is the concept of *productive time*. For example, sitting on the train can be considered as enjoyable and productive where travellers can engage themselves with other activities (such as reading, checking emails, writing, or any other work-related activities).

Travel cost
Monetary spending on travelling can be classified into two broad categories:

- Standing costs: these are the costs incurred on the traveller no matter whether they travel or not.
- Running costs: these are the costs that travellers pay out of their pockets, usually per trip.

Table 3.3 lists the common types of standing costs and running costs per mode. If only short-term behavioural changes are expected, then only out-of-pocket costs need to be considered. If long term behavioural changes are also expected, then both standing costs and running costs should be taken into account.

Table 3.3 Travel cost per mode

Mode	Standing costs (long term)	Running costs (out of pocket)
Driving	Vehicle tax (road tax) Insurance Capital cost Depreciation Servicing (e.g. replacement parts)	Fuel Parking Toll (e.g. congestion fee)
Cycling (private bike)	Capital cost Depreciation Servicing	—
Cycling (rental bike)	Subscription (may or may not be applicable)	Rent
Walking	—	—
Transit (without pass)	—	Fare
Transit (with pass)	Price of transit pass	—

For travelling by the private mode of driving, some of the standing costs may be dependent on actual vehicle usage. For example, annual depreciation is considered to be monotonically increasing with annual mileage. In some cases, the insurance premium is also related to the vehicle mileage. Nonetheless, travellers do not pay for these costs per trip; it is also difficult for them to estimate the contribution of a single trip to these costs. Travellers therefore tend not to relate individual trips to these costs.

Car sharing and carpooling are emerging alternatives of driving the traveller's own car. The cost structure changes slightly or drastically, depending on the exact arrangement (Table 3.4). Rental cars are usually used for short periods (from a few hours to a few days), whereas lease cars are contracted for a fixed period of time (e.g. one year). Carpooling allows drivers to share the available seats in their car with fellow travellers, where the running costs are shared by the two parties.

Table 3.4 Travel cost for car sharing and carpooling

Scheme & role	Standing costs (long term)	Running costs (out of pocket)
Rental car (driver)	—	Rent Parking (may be unnecessary) Toll
Lease car (driver)	Lease payment	Fuel Parking Toll
Carpool (driver)	Vehicle tax (road tax) Insurance Capital cost Depreciation Servicing (e.g. replacement parts)	Fuel (shared) Parking (shared) Toll (shared)
Carpool (passenger)	—	Fuel (shared) Parking (shared) Toll (shared)

Emissions

Emissions are typically measured by the amount of substance emitted. Two types of substances are considered here:

- Carbon dioxide (CO2);
- Air pollutants, including nitrogen oxide (NOx) and particulate matter (PM).

CO2 is a major greenhouse gas (GHG) and contribute greatly to the greenhouse effect. However, it is not widely considered as a pollutant which damages the ecological system and is hazardous to human health.

There are many factors that influence the amount of emission, including: (Wismans *et al.* 2011)

- Vehicle characteristics;
- Driver behaviour;
- Traffic condition;
- Road infrastructure.

Simplified emission models can be used to estimate the amount of emission for a trip, such as

$$E = \alpha M \tag{0.1}$$

Here E is the amount of emission, M is the trip distance, and α is the emission factor which depends on the vehicle class and the average speed.

Besides emissions, other externality-related measurements can also be adopted, such as traffic accident risk and noise pollution. These are discussed in further details in D6.2, alongside a more in-depth consideration of traffic emissions.

3.2.2 Behavioural indicators: distribution over trip characteristics

The behavioural indicators derived in §3.2.1 can be further specified by examining their distribution over the trip characteristics. Characteristics of a trip relate to the context and means of a trip. The trip purpose, its spatial location, and its timing constitute the trip context. The means of a trip refers to the mode(s) and route with which the trip is realised.

Trip purpose

Trips are made in order to access activities and trip purposes can be nominated by these activities. Three categories of trip purposes are identified (as in §2.2), in consistency with the categories commonly adopted in the transport field:

- Work/school commute;
- Work/school related;
- Entertainment/recreational/social.

Location/length of trip

The spatial dispersion of a trip's origin and destination is relevant for analysing the trip patterns. The mobility infrastructure can be divided into the follow types of areas:

- City proper, which can be further divided into central business district (CBD) and residential areas;
- Suburbs;
- Rural areas.

A trip can then be characterised by the locations of its origin and destination, e.g. from suburb to CBD, from a residential area to a rural area. A trip can also be characterised by its length (or in combination with location). Three categories are adopted here:

- Long trips, e.g. inter-city;
- Medium trips, e.g. within city;

- Short trips, e.g. within neighbourhood.

Timing of trip
Two major categories are distinguished:

- Trips that are fully or partially in the rush hour period (peak);
- Trips that are outside the rush hour period (off peak).

The rush hour period is the part of the day when traffic congestion on the road and crowding in public transport are at their peak. The exact definition of the rush hour period varies from location to location. In general, two peak periods are considered during a working day:

- Morning peak: usually between 06:00 and 10:00, when most people commute from home to work or school;
- After peak: usually between 16:00 and 19:00, when most people commute from work or school to their home.

Transport modes
The following modality types are identified in SUNSET:

- Foot
- Bike
- Private motor vehicle (car)
- Train / Light Rail
- Metro
- Bus / Tram
- Ferry
- Other modes

Trajectory of trip
This characteristic is closely related to the location of a trip, but focuses more on the trajectory (route) of the trip instead of its origin and destination and applies mainly to car trips.

The trajectory can be decomposed into segments. In terms of road characteristics, the trip can be divided into segments based on speed limit, or by road categories, e.g.:

- Urban roads;
- Arterial roads (corridors);
- Highway.

In terms of traffic characteristics, the trip can be divided into segments based on the congestion level:

- Free flow condition;
- Slightly congested;
- Heavily congested.

In terms of location specific features of the trip, the trip can be divided into segments based on the following criteria:

- Whether it is within a residential neighbourhood;
- Whether there is a school nearby;

- Whether there are children playing on or near the street;
- Etc.

The overall measurements discussed in §3.2.1 can be made more relevant by examining their distribution over the different trip characteristics. Combining the measurements with these characteristics, Table 3.5 provides an overview on the types of potential behavioural indicators. In Table 3.6, the most useful examples are listed.

For the purposes of evaluating the SUNSET incentives, it could be more relevant to focus on combining several trip characteristics, e.g. trips for work related purposes within the peak period over the car mode. This is discussed in details in §3.2.4.

Table 3.5 Categorisation of behavioural indicators

Characterisation	Measurements				
	Number of trips made	Distance travelled	Travel time	Travel cost	Emissions
Total amount	(#)	(#)	(#)	(#)	(#)
Distribution over purpose	(#/%)	(#/%)	(#/%)	(#/%)	(#/%)
Distribution over space	(#/%)	(#/%)	(#/%)	(#/%)	(#/%)
Distribution over time	(#/%)	(#/%)	(#/%)	(#/%)	(#/%)
Distribution over mode	(#/%)	(#/%)	(#/%)	(#/%)	(#/%)
Distribution over route	(#/%)	(#/%)	(#/%)	(#/%)	(#/%)

Table 3.6 Measurement examples for behavioural indicators

Characterisation	Measurement examples
Total amount (#)	Total travel time within a period (e.g. every month or two months) Total travel cost within a period
Distribution over purpose (#/%)	Number of trips for leisure Travel cost of work related trips Distance travelled of non-work related trips Emissions of work related trips
Distribution over space (#/%)	Number of trips into the city centre Number of trips within own neighbourhood
Distribution over time (#/%)	Distance traveller during weekdays/weekends Travel time (/%) in peak/off-peak hours
Distribution over mode (#/%)	Number of trips on non-/motorised modes Travel cost (/%) by car/PT/bike/foot
Distribution over route (#/%)	Number of trips passing a school zone Distance travelled on urban roads Travel time spent in congestion

3.2.3 Indicators related to mobility attitude

The concept of mobility attitude deals with the travellers' perception and opinion of their own mobility as well as the mobility of wider society. In behavioural science, it is argued (Ajzen 1991) that the intentions to perform certain behaviours can be predicted with higher accuracy from attitudes toward the behaviour, subjective norms, and perceived behavioural control. In personal mobility, researchers have been selecting target groups based on their attitude in order to achieve at certain policy goals. For example, the study in Hunecke *et al.* (2008) analyses the usefulness of an attitude-based approach for predicting the ecological impact of daily mobility behaviour. Different attitude groups are shown to exhibit significantly different behaviour with regard to mode choice, distance travelled, and ecological impact. Attitude is therefore a relevant factor in assessing the effectiveness of incentives. In some cases, incentives may induce

attitude shifts in the traveller but yet no behavioural change is observed. This could be due to several reasons. Firstly, the changes in attitude may not be substantial enough to invoke a behavioural change under the given situation; secondly, travellers may be bound by other constraints. In either case, the change in attitude makes it more likely that the traveller might change his/her behaviour for future situations.

Two broad categories of attitude can be distinguished:

- Awareness
- Opinion

Awareness is used to denote whether the traveller knows about something or not. Unawareness of certain facts means that these facts will not be taken into account in the traveller's decision making process. Changes in awareness will modify the framework of the decision making situation, by including more alternatives, new attributes, new criteria, and/or more (reliable) information.

Opinion, on the other hand, represents how the traveller processes, evaluates, weighs, and compares available information. It affects the prioritisation and optimisation mechanism in the decision making process. A change in opinion will reshape this mechanism and affect the outcome of the process.

Several clusters of mobility attitude are discussed in details below. These include both awareness and opinion related aspects of attitude.

Self-awareness
Self-awareness refers to the traveller's awareness of his/her own mobility pattern and the consequent impact on society regarding efficiency and externalities. This includes:

- Awareness of own mobility pattern: this relates to the subjective/perceived mobility measurements, including:
  - Perceived travel time/distance/cost;
  - Perceived delay time in congestion.

- Awareness of impact of own mobility: this relates to the subjective/perceived mobility impacts, including:
  - Perceived marginal cost on traffic congestion/delays;
  - Perceived impact on traffic safety;
  - Perceived impact on the environment (emissions, pollutants);
  - Perceived impact on citizens' wellbeing (e.g. noise).

Alternative-awareness
Alternative-awareness refers to the traveller's awareness of alternative means to fulfil his/her current mobility pattern. This includes:

- Awareness on feasible alternatives: this relates to the knowledge of potential alternatives, including:
  - E-commuting/-shopping, instead of physical trips;
  - Alternative timing of trips (e.g. peak hour avoidance);
  - Alternative modes;
  - Alternative routes.

- Awareness of the (relative) performance of the alternatives: this relates to the performance of the alternatives, relative to the current choice, in terms of fulfilling the mobility pattern. It includes a comparison of the alternative with the current choice in terms of:
  - Travel time differences;
  - Cost differences.

Societal-awareness

Societal-awareness refers to the traveller's awareness of the society's mobility pattern as a whole and the consequent impact regarding efficiency and externalities. This includes:

- Awareness of traffic congestion: this relates to the subjective estimation of the mobility performance of the society, including:
  - Scope/extent of traffic congestion:
    - Length of traffic jams,
    - Lost time in congestion,
    - Percentage of travellers affected,
    - Average lost time in congestion per person;
  - Economics of traffic congestion:
    - Lost value of time;
    - Secondary loss induced by traffic congestion (e.g. missed appointment).

- Awareness of environmental issues in general: this relates to the general environmental issues the society is facing today, including:
  - Scope/extent of environmental issues:
    - Global warming,
    - Air pollution,
    - Fossil fuels approaching extinction,
    - Hazardous/industrial waste;
  - Economics of environmental issues.

- Awareness of traffic-induced environment issues: this relates to the contribution of traffic to the environmental issues the society is facing today, including:
  - Contribution of traffic to global warming, air pollution, fossil fuel extinction, etc.,
  - Environmental cost of traffic.

Self-identification

The concept of self-identification refers to the traveller's attitude regarding him-/herself. This includes:

- Type of traveller: this relates to the *self-categorisation* of the traveller with respect to his/her personality and/or current mobility behaviour, including:
  - Mode choice, e.g.:
    - I always drive,
    - I take the train as much as possible,
    - I often bike,
    - I seldom take the bus,
    - I never walk;
  - Route choice, e.g.:
    - I take the highway as much as possible,
    - I avoid city centres as much as possible,
    - I do try to avoid passing a school zone,

        - I always follow what my navigation system says;
    - Timing, e.g.:
        - I commute in peak hours;
        - I avoid peak hours whenever I can;
    - Planning, e.g.:
        - I always plan my trip in advance,
        - I'm a spontaneous traveller;
    - Mileage, e.g.:
        - I travel a lot,
        - I'm an inter-city traveller,
        - My trips are mainly long distance trips.

- Preferences: this relates to the way how a traveller prioritises alternatives. It is closely connected to the concept of self-categorisation. The main difference is that self-categorisation focuses on status (or outcome), whereas preference deals with the selection process. Preferences are be characterised by the traveller's decision making process, including:
    - To save time, e.g.:
        - I always choose the mode that is the quickest,
        - I always choose the route that is the quickest,
        - I highly value travel time reliability,
        - I highly value punctuality (or on-time arrival);
    - To save cost, e.g.:
        - I always choose the mode that is the cheapest,
        - I always choose the route that is the cheapest,
        - I avoid any road with a toll.

- Satisfaction: this relates to the extent that a traveller's needs and desire are fulfilled, including:
    - The desired mobility (especially travel time);
    - The planning/scheduling effort;
    - Travel freedom;
    - Travel comfort.

Social attitude/opinion

The concept of social attitude refers to the traveller's opinions concerning the social provision of mobility services. This includes:

- Current infrastructure/services: this relates to the traveller's opinion on the current mobility infrastructure and services, including:
    - Performance and acceptability, e.g.:
        - The current traffic congestion level is okay,
        - The bus frequency is too low,
        - Ticket price for the metro is too high;
    - Impact, e.g.:
        - Traffic congestion is seriously impacting our economy,
        - Traffic-induced air pollution is deteriorating the livelihood of our society;
        - Traffic noise is a big problem.

- Preferred changes: this relates to the traveller's *self-categorisation* of the traveller with respect to his/her personality and/or current mobility behaviour, including:
    - Policy, e.g.:

    - Fuel tax should go up,
    - The city centre should be car free,
    - Congestion pricing should be in place,
    - Public transport should receive more subsidy;
  - Planning, e.g.:
    - More roads should be built,
    - More railways should be built,
    - More bicycle paths should be built;
  - Management, e.g.:
    - Buses should always receive priority at traffic lights,
    - Through traffic should be rerouted via the ring road,
    - Parking should be free at transfer locations (e.g. railway stations).

3.2.4 Indicators for the SUNSET incentives

This subsection proposes a set of representative indicators for each category of the SUNSET incentives. For SUNSET, changes in actual behaviour will be monitored through the mobility pattern based on automatic data collection. Changes in mobility attitude can be assessed using questionnaires and experience sampling, as discussed in §2.2. The exact method of measurement will depend on the design of the individual living lab or other real-life implementation. For example, the implementation may be based on encouraging changes for a particular corridor or for a particular mode or particular destination. The time period over which the amount of change is observed may also vary according to the time over which the incentive is offered (e.g. over a period of a month, three months, etc.).

From D3.2 and D3.3, each incentive is expected to invoke behavioural and attitude changes. Based on these expected areas of change, Table 3.7 lists (for each incentive type) the indicators that are most representative in order to capture the intended impact.

Table 3.7 Indicators for measuring changes due to the SUNSET incentives

Type of incentive	Indicators for behavioural change	Indicators for attitude change
Real-time travel information provided by the system	Number of trips within a given period Travel time per purpose per mode Fuel cost per purpose per mode Emissions per purpose per mode	Awareness of alternative routes Awareness of performance on the alternative routes Societal awareness of traffic congestion on the traffic network Satisfaction Social attitude
Peer-to-peer travel information	Number of trips per mode per timing period Travel time per purpose per mode	Awareness of alternative routes Awareness of alternative modes Awareness of performances of alternative modes Societal awareness Social attitude
Feedback based on self-monitoring on own travel behaviour	Number of trips per mode Travel time Travel costs Emissions	Awareness of own mobility pattern Awareness of the impact of own mobility Self-categorisation Preferences Satisfaction Social attitude
Feedback based on setting targets	Number of trips Travel distance/time/cost Emissions	Awareness of the impact of own mobility Self-categorisation Preferences Satisfaction
Points without an	Number of trips per purpose per	Awareness of performances of alternative

exchange value	mode Travel time per purpose Emissions per purpose	modes
Points with an exchange value	Number of trips per purpose per mode Travel time per purpose Emissions per purpose	Awareness of performances of alternative modes
Social networks for sharing location	Number of trips per purpose per location	Awareness of alternative modes/routes Awareness of performances of alternative modes Societal awareness
Social networks for finding a buddy	Number of trips per mode Emissions per trajectory	Awareness of alternative modes/routes Preferences Satisfaction Social attitude
Social networks for treasure hunt	Number of trips per mode per trajectory Travel distances per purpose Emissions per trajectory	Awareness of alternative modes/routes

3.3 Evaluation methods on effectiveness

The discussion of evaluation methods for the effectiveness of incentives will cover both data collection and experimental methodology. In recent years GPS-enabled devices have been introduced as a data source in travel behavioural research. Besides being a new method for data acquisition, this has offered new possibilities for the analysis of individual travel behaviour. In travel behaviour research and other research fields, several methods of analysis of panel data have been used or could potentially be beneficial to the analysis of travel behaviour change in SUNSET. In this section we will discuss the most appropriate experimental methodology and tools for analysis within the SUNSET context.

3.3.1 Experimental methodology

In transport research, there are two main types of survey that are used to assess the change in behaviour other than automatically recorded trip movements. These are panel surveys and (repeated) cross-sectional surveys (Stopher *et al.* 2009, Brög *et al.* 2009, Ortúzar *et al.* 2010). The nature of the SUNSET project already embodies a panel survey in the sense that individuals are monitored over a longer period of time. This enables research into the dynamics of individual behaviour. However, conducting a panel survey brings some additional points for attention. This section will elaborate on the research issues in using a panel survey approach and will provide an experimental design based on the functionalities of tripzoom, to study the changes in travel behaviour as a result of the incentives and challenges offered.

Panel survey

A panel survey is a longitudinal research approach based on following a particular group of individuals that are subjected to a shared experience within a defined period (Ortúzar *et al.* 2010). In general, two sub-groups are defined: a group that is subjected to an intervention and a group that is *not* subjected to the intervention, but otherwise similar. A panel survey offers the possibility to study the dynamics of behavioural change in relation to the intervention. This research approach has been conducted extensively in economics, medicine, social sciences, as well as in the field of transport research.

When measuring change due to the intervention, other factors that may cause a behaviour change need to be eliminated from the analysis. When using a panel survey a control group is required in combination with the group subjected to the intervention. The control group consists of a comparable group of subjects that are not subjected to the intervention. Changes in external factors affecting the analysis can now be eliminated or at least accounted for. However, introducing a control group brings new challenges to account for (e.g. attitudes of control group may differ from subject group and diffusion effects causing the control group also to change behaviour due to the intervention). Ideally, the subgroup that is not subjected to the intervention is recruited from another population (i.e. comparison group), preventing this subgroup from being indirectly affected from the intervention (i.e. diffusion effect). However, this imposes new challenges on the recruitment process, because the geographical and infrastructural environments of the two subgroups may differ. It may also be problematic as a result of the type of transport measure or intervention being assessed as the practicality of eliminating diffusion could be extremely difficult with measures that are naturally pervasive in location, throughout systems or within the population.

A major concern of the panel surveys is the risk of attrition (i.e. loss of panel members). Panel members may lose interest or move away from the research area and fail to respond on subsequent survey waves or entirely leave the sample. The risk and the extent of attrition rise with increasing survey intervals. On the other hand, more frequent measurements will raise survey fatigue, causing a loss of interest in participation. These drawbacks in panel surveys can be overcome by regularly contacting the panel members and oversampling (i.e. incorporate a significantly larger sample size than strictly required).

Other limitations and concerns are conditioning of participants. When participants are aware of the objectives of the project behind the panel survey, they may show more 'desirable' behaviour and provide more socially desirable answers to experience sampling questions.

(Repeated) cross-sectional survey

The methodological counterpart of the panel survey is the cross-sectional survey. In a cross-sectional survey a representative sample of travel behaviour is taken from the population with the aim of providing data on the entire population. In the case of an investigation into travel behaviour change as a result of an intervention, a repeated cross-sectional survey can be conducted. In this case the representative sample is taken at several moments in time, for example before, during and after the intervention. A major concern in the use of the cross-sectional survey is the large sample size needed to measure a change in behaviour in comparison with panel surveys (Brög *et al.* 2009, Stopher *et al.* 2009, Richardson *et al.* 2004). The covariance between the measurements in a panel survey can substantially reduce the variance of the difference in the before and after study and therefore requires smaller sample sizes in order to detect statistically significant outcomes (Stopher & Greaves, 2007). Moreover, the dynamics of change can be clarified using a panel survey. Every participant in a panel can be compared with itself each time it is surveyed. Simultaneously the changes in personal or household characteristics can be tracked to account for possible changes in behaviour due to factors other than the intervention. In cross-sectional surveys it is only possible to make aggregate comparisons through time. Therefore it is not possible to uncover the underlying processes of the behavioural changes.

3.3.2 Experimental design and data collection

Based on both the intrinsic concept of the SUNSET project of measuring the travel behaviour of the system users and the methodological advantages of a panel survey over a cross-sectional survey, the experimental design of measuring the changes in travel behaviour will be based on

a panel survey design. A number of groups of users will be monitored over a period of time. Each group receives a different sequence of functionality of the SUNSET-system.

It should be noted that the full experimental design is described here, as appropriate to the main LL in Enschede. The reference living labs may cover a reduced number of groups – possibly two to four, with the group selection depending on the local transport objectives and design of the LL. It may also be the case that a reduced number of people are in the group. The reference LL have been designed to show particular variations in the way that the SUNSET system may be used and to demonstrate the link to particular and differing kinds of local transport objectives. Task 6.3 (reported in D6.2) will report on the translation of the design into the exact requirements for the living labs.

Experimental design

To be able to measure the effect of incentives, the LL participants are asked to join an experiment over a certain period of time. Users joining the experiment receive a dedicated version of the tripzoom app and the LL-coordinator should target the participant specifically according to the experimental group they are assigned to. In the experimental design (Table 3.8), a distinction is drawn between participants who are:

- In an experimental group, according to the LL implementation design ('Group parts')
- In a sub-group according to socio-demographic and other characteristics ('subgroup parts'), and
- Have freely joined the Tripzoom community either by invitation or via the webpages ('Free parts')

It is worth noting that some of the incentives are offered by LL co-ordinators, some are offered/generated by the participants themselves and some are system generated.

A distinction is drawn between incentives that need to be targeted towards particular sub-groups (which may be in a particular group, or which may be in the Free part's cohort), and where further design is needed on which subgroups will be offered particular variations of an incentives.

It will be necessary for individuals to be briefed on the issue of incentives (for example that we hope they will engage in posting messages on the group F/B pages, how to use 'find a buddy'). It will also be necessary to consider 'rewarding' participants for taking part with only sub-functions of the SUNSET features.

For the purposes of evaluation, a number of incentives will be assessed in packages (i.e. they are introduced simultaneously). This is due to the constraint the software is unable to 'switch off' particular functions. However, even where some challenges are introduced simultaneously, they may include 'tailored' incentives to the sub-group characteristics.

Table 3.8 Participant groups and the incentives offered during the experiment

Group	Time period					
	1	2	3	4	5	6
1	*Travel diary (manually) to establish non-app travel patterns and mode	Mobility feedback and performance against community	Challenges		Social media	
	Briefing of	*(3) Feedback*	*(1) Real-time travel*		*(2) Social networks for peer-*	

	participants on their role in social media and other participatory challenges	*based on self-monitoring on own travel behaviour (4) Feedback based on setting targets (as a package)*	*information provided by the system (5) Challenges without points reward (6) Challenges with points reward (in sequence)*	*to-peer travel information/messages (alerts and hazards) (7) Social networks for sharing location (8) Social networks for finding a buddy (9) Social networks for treasure hunt (2 and 7 as a package, 8 and 9 in sequence)*
2	*Travel diary (manually) to establish non-app travel patterns and mode	Mobility feedback and performance against community	Social media	Challenges
	Briefing of participants on their role in social media and other participatory challenges	*(3) Feedback based on self-monitoring on own travel behaviour (4) Feedback based on setting targets (as a package)*	*(2) Social networks for peer-to-peer travel information/messages (alerts and hazards) (7) Social networks for sharing location (8) Social networks for finding a buddy (9) Social networks for treasure hunt (2 and 7 as a package, 8 and 9 in sequence)*	*(1) Real-time travel information provided by the system (5) Challenges without points reward (6) Challenges with points reward (in sequence)*
3	*Travel diary (manually) to establish non-app travel patterns and mode	Mobility feedback and performance against community only		
	Briefing of participants that they receive only some functions	*(3) Feedback based on self-monitoring on own travel behaviour (4) Feedback based on setting targets (as a package)*		
4+	Mobility feedback and performance	Challenges		
	(3) Feedback based on self-monitoring on own travel behaviour (4) Feedback based on setting targets (as a package)	*(1) Real-time travel information provided by the system (3) Challenges without points reward (6) Challenges with points reward (in sequence)*		
5#	Mobility	Social media based incentives		

	feedback and performance	
	(3) Feedback based on self-monitoring on own travel behaviour *(4) Feedback based on setting targets* *(as a package)*	*(2) Social networks for peer-to-peer travel information/messages (alerts and hazards)* *(7) Social networks for sharing location* *(8) Social networks for finding a buddy* *(9) Social networks for treasure hunt* *(2 and 7 as a package, 8 and 9 in sequence)*
6#	Mobility feedback and performance	Challenges and social media
	(3) Feedback based on self-monitoring on own travel behaviour *(4) Feedback based on setting targets* *(as a package)*	*(1) Real-time travel information provided by the system* *(3) Challenges without points reward* *(4) Challenges with points reward* *(in sequence)* *(2) Social networks for peer-to-peer travel information/messages (alerts and hazards)* *(7) Social networks for sharing location* *(8) Social networks for finding a buddy* *(9) Social networks for treasure hunt* *(2 and 7 as a package, 8 and 9 in sequence)*
7^	All mob feedback, challenges and incentives	
	(1) Real-time travel information provided by the system *(2) Social networks for peer-to-peer travel information/messages (alerts and hazards)* *(3) Feedback based on self-monitoring on own travel behaviour* *(4) Feedback based on setting targets* *(5)Challenges without points reward* *(7) Social networks for sharing location* *(8) Social networks for finding a buddy* *(all incentives as a package)*	

* For group 1/2/3, travel diary information is collected and for the remaining four groups this is not the case. So for the first three groups, in effect two 'base' cases are established – a true base case and a synthetic base case. The comparison is then made as 1) changes against travel diary data and 2) changes against mobility monitoring (as for the other 4 cases).
+ For group 4, on recruitment, the LL-coordinator should brief the participants on the fact that they receive only some functions of tripzoom.
# For group 5/6, on recruitment, the LL-coordinator should brief the participants on their role in social media and other participatory challenges.
^ Group 7 is the free-form group.

Gathering quantitative data

The experimental design aims to construct a base case for the participants first and then schedules the offering of (sets of) incentives in subsequent time periods. Ideally, all participants in the experiment would complete a travel diary (manually) and have tripzoom running on their smartphone in time period 1. In this ideal case the participants would only use the measuring functionality of tripzoom, and would therefore not be influenced by the functionality of 'mobility feedback and performance'. To get a stable overview of travel behaviour, a measurement period of 2 weeks of trips should provide a sufficient amount of data per individual (Schönfelder

2006). After a base case is established, particular incentives and additional app functionalities are added to be able to evaluate the impact of those specific incentives (with the extra app functions).

However, this approach has a number of practical difficulties which have led to a modified design and which are briefly summarised as follows. Participants will be unable to carry tripzoom during their travel within time period 1 without receiving mobility feedback or an indication of personal performance in return. This is due to the embedded nature of the mobility feature within the software and inability to 'switch off' this functionality. Secondly, the participant burden of requiring a travel diary is considered to be too high to expect this from the large number of participants that may take part. It may be necessary to offer a reward to participants to supply this additional travel diary data. Finally, those participants who join in the free-form group are essentially outside the experimental control of the LL – therefore the expectation of good-will contributions of further data in this way is unrealistic.

As a result, in the experimental design several groups of incentives will be assessed. The base cases are constructed in two ways: manual travel diaries and automatically recorded by tripzoom in combination with the mobility feedback and performance functionality. In this way a distinction can be made between mobility patterns when not using tripzoom and travel patterns collected automatically when using the mobility feedback and performance functionality. The effectiveness of incentives will be investigated by further subdividing the participants of the experiment into 3 groups. The participants in the different groups receive different sequences in the sets of incentives. The effectiveness of using only the mobility feedback and performance can be investigated using groups 1 to 3. The base case is constructed without using tripzoom. The subsequent time period provides the opportunity to study the effectiveness of separate incentives. In this case group 3 is used as a reference for group 1 and 2 to investigate the additional effect of challenges and social media. Group 4, 5 and 6 will provide information about the effectiveness of combining incentives. Moreover these groups serve to study the longer term effects of the particular incentives and incentive groups.

An implication for constructing 6 groups of participants in the experiment is the issue of sample size. To measure changes in travel behaviour groups must be sufficiently large to ensure potentially small changes in behaviour can be detected. The aim is to include 240 participants in the experiment, resulting in 40 participants per group. The numbers of participants joining group 7 is unknown and outside experimental control.

Gathering qualitative data

An important issue with the experimental design is the absence of a control or comparison group. External factors influencing travel behaviour cannot be accounted for. For example (seasonal) weather factors and alterations in the (local) transport system during the experiment may alter travel behaviour. Without a comparison or control group these influences cannot be accounted for in the analysis. Both as a solution for the previous issue and to study the attitudes of participants, qualitative data are also gathered from the participants in the experiment. These data shape the context in which travel choice are made, resulting in actual travel behaviour, from the perspective of both individual (e.g. personal and household characteristics, attitudes and preferences) and the transport system (e.g. influence of weather/seasonal conditions, level-of-service of the various transport modalities).

In similar vein to the data collection method outlined in §2.2, the qualitative data can be classified according to the following question types:

1) Questions to gather background information from the participants (e.g. personal/household characteristics to establish participant types or background factors that may influence travel choices. These questions can be delivered by a single data collection process using an electronic questionnaire **(QR1)**

2) Questions to gather attitudes and circumstances during the timeframe of the experiment, for example to monitor changes in beliefs and changes in availability of transport opportunities (e.g. weather conditions, traffic jams and road works). These questions can be gathered by repeated data collection using electronic questionnaire as these do not require the 'current/recent' tripzoom related experiences **(QRM)**

3) Questions where the best quality responses will come from a reflection of very recent transport experience or of the tripzoom experience, these may be asked once or more than once through experience sampling **(XP)**

The qualitative data is gathered during various time periods prior to or during the use of tripzoom. The timing is shown in Table 3.9. Additionally, focus groups will be used as part of the experiment to study the changes in attitudes more thoroughly.

Table 3.9 Timing of qualitative data gathering

<table>
<tr><th rowspan="2">Group</th><th colspan="7">Time period</th></tr>
<tr><th>1</th><th colspan="2">2</th><th colspan="2">3</th><th colspan="2">4</th><th colspan="2">5</th><th colspan="2">6</th></tr>
<tr><td rowspan="3">1</td><td colspan="2">*Travel diary (manually) to establish non-app travel patterns and mode</td><td colspan="2">Mobility feedback and performance against community</td><td colspan="2">Challenges</td><td colspan="2">Social media</td></tr>
<tr><td colspan="2" rowspan="2">QR1</td><td>QRM(i)</td><td>QRM(ii)</td><td colspan="2"></td><td>QRM(iii)</td><td colspan="2"></td><td>QRM(iv)</td></tr>
<tr><td colspan="2">XP</td><td colspan="2">XP</td><td colspan="2">XP</td></tr>
<tr><td rowspan="3">2</td><td colspan="2">*Travel diary (manually) to establish non-app travel patterns and mode</td><td colspan="2">Mobility feedback and performance against community</td><td colspan="2">Social media</td><td colspan="2">Challenges</td></tr>
<tr><td colspan="2" rowspan="2">QR1</td><td>QRM(i)</td><td>QRM(ii)</td><td colspan="2"></td><td>QRM(iii)</td><td colspan="2"></td><td>QRM(iv)</td></tr>
<tr><td colspan="2">XP</td><td colspan="2">XP</td><td colspan="2">XP</td></tr>
<tr><td rowspan="3">3</td><td colspan="2">*Travel diary (manually) to establish non-app travel patterns and mode</td><td colspan="8">Mobility feedback and performance against community only</td></tr>
<tr><td colspan="2" rowspan="2">QR1</td><td>QRM(i)</td><td>QRM(ii)</td><td colspan="2"></td><td>QRM(iii)</td><td colspan="2"></td><td>QRM(iv)</td></tr>
<tr><td colspan="2">XP</td><td colspan="2">XP</td><td colspan="2">XP</td></tr>
<tr><td rowspan="3">4</td><td colspan="2">Mobility feedback and performance</td><td colspan="8">Challenges</td></tr>
<tr><td>QR1
QRM(i)</td><td>QRM(ii)</td><td></td><td>QRM(iii)</td><td colspan="2"></td><td>QRM(iv)</td><td colspan="2"></td><td>QRM(v)</td></tr>
<tr><td colspan="2">XP</td><td colspan="2">XP</td><td colspan="2">XP</td><td colspan="2">XP</td></tr>
<tr><td rowspan="3">5</td><td colspan="2">Mobility feedback and performance</td><td colspan="8">Social media based incentives</td></tr>
<tr><td>QR1
QRM(i)</td><td>QRM(ii)</td><td></td><td>QRM(iii)</td><td colspan="2"></td><td>QRM(iv)</td><td colspan="2"></td><td>QRM(v)</td></tr>
<tr><td colspan="2">XP</td><td colspan="2">XP</td><td colspan="2">XP</td><td colspan="2">XP</td></tr>
<tr><td>6</td><td colspan="2">Mobility feedback and performance</td><td colspan="8">Challenges and social media</td></tr>
</table>

	QR1 QRM(i)	QRM(ii)		QRM(iii)		QRM(iv)		QRM(v)
	XP		XP		XP		XP	
7	All mob feedback, challenges and incentives							
	QR1 and QRM(i) on joining at time t; QRM(ii), QRM (iii), QRM(iv) successively at time (per month or week) depending on lab design.							

3.3.3 Tools for data analysis

To be able to measure the extent of any changes in travel behaviour, a formal method for data analysis is required. Measuring change in travel behaviour can be conducted by investigating changes in the travel behaviour indicators on an aggregate level and by investigating the characteristics of the separate trips an individual makes (disaggregate level). On the aggregate level the approach is straightforward in the sense that the indicators of travel behaviour for one particular participant or groups of participants are compared for the consecutive time periods. For example, a participant makes 25 trips in the base case and 20 trips when using tripzoom. This could be further specified in terms of specific modalities, travel purposes, etc. (as discussed in §3.2).

Hypothesis testing

For certain indicators on travellers' behaviour and attitude, two subsets of data can be distinguished:

- Data in the 'before' period (the base scenario, without exposure to incentives)
- Data in the 'after' period (the design scenario, after exposure to incentives)

For these indicators, *hypothesis testing* techniques can be applied to assess the statistical significance of changes between the base and design scenarios. Regarding the indicators of change, two types of variables are distinguished:

- Changes in numeric value of the variable, e.g. the user makes fewer car trips now than before;
- Changes in proportions, e.g. the proportion of car trips has decreased (among trips of all modes, whilst the absolute number of car trips may have increased, decreased, or remained the same).

Different hypothesis testing techniques may be employed, and are discussed in Appendix A.

Longitudinal data analysis tools

The approach of hypothesis testing fails to capture the sequential component of travel behaviour. A trip sequence of going from home to work and from work back home via the supermarket, would be the same as going to work via the supermarket and going straight home after work. Accounting for changes as such requires a more trip based approach.

To capture the disaggregate travel behaviour change, several approaches for longitudinal data analysis are described in existing literature, including *similarity indices*, *Herfindahl Hirschmann Index* (HHI), *sequence alignment method*, *tour-based analysis*, *location-based analysis*, *hazard modelling and survival analysis* and *discrete choice modelling*. These tools for analysis are described more extensively in Appendix B. Here an overview of these methods is provided, as well as some elaborations on the appropriateness of use in the analysis of the effectiveness of incentives.

The similarity indices and sequence alignment method seek for the extent of similarity in a sequence of trips (Kitamura & Hoorn 1987, Schönfelder & Axhausen 2002, Schlich & Axhausen 2003). The similarity is then expressed in terms of an index. Although these methods are rather straightforward, they fail in pointing out any direction of behavioural change. Moreover, no 'base case' of travel behaviour can be constructed because the methods rely on actual trip sequences instead of a construct of regular behaviour. The HHI sidesteps this problem by aggregating behaviour from any before and after situation and investigate the stability of travel choices (Schönfelder 2006). However, similar to the similarity indices and sequence alignment method, the HHI does not depict a direction of behavioural change.

A tour-based analysis starts with defining trip sequences and assigning the actual trip sequences to the predefined trip sequences (Golob 1986, Krizek 2003, Stopher *et al.* 2010). In this way the trips of any time periods can be summarized in terms of observed trip sequences. Also, different time periods can be compared. However, this method suffers from a trade-off between accuracy of trip sequence definitions and the number of actual occurrences of the trip sequences. The more trip sequences are defined, the smaller the number of occurrences per sequence and the less statistically sound the outcomes will be.

Location-based analysis focuses on the analysis of the locations individuals visit. Actual visits of locations are represented and the probability of visiting adjacent locations can be determined (Schönfelder 2006). However, this method fails in capturing trips between various locations and therefore seems less appropriate for use in the evaluation of tripzoom.

Hazard modelling, survival analysis and discrete choice modelling focus on capturing behaviour in models. Hazard modelling and survival analysis aim to model the interval between two occurrences (i.e. trips) and can incorporate various personal and household characteristics (Schönfelder 2006). Choice modelling aims at estimating the propensity of making certain (travel) decisions (Frees 2004). The advantage of these models is the ability to find causal relations. However, the modelling process requires assumptions on the model shape. Moreover, the interpretation of the outcomes may be less straightforward.

An overview of the methods discussed above is given in Table 3.10. The table also presents the applicability to the analysis of the effectiveness of incentives. The last column refers to the applicability of the specific method on the indicators listed in §3.2.

Table 3.10 Overview of methods for measuring changes in behaviour

Method	Strengths	Weaknesses	Applicable to analysis of:
Similarity indices	• Picking-up changes in the sequence of trips • Possibility to incorporate trip characteristics	• Difficult to construct 'average/regular' travel behaviour • Index does not depict the direction of behavioural change	Number of trips (comparison of sequences)
Herfindahl Hirschmann Index	• measure of concentration • Enables studying shifting attitudes • Applicable to departure modality choice, time choice and route choice	• does not depict the direction of behavioural change • difficult to interpret with new modalities	Number of trips within a given period Number of trips per mode
Sequence alignment method	• Picking-up changes in the sequence of trips • Possibility to incorporate trip characteristics	• Difficult to construct 'average/regular' travel behaviour • does not depict the direction of behavioural	Number of trips per purpose per mode Number of trips per mode Number of trips

		change	
Tour-based analysis	• Provides a way to construct 'average/regular' behaviour • Includes sequencing of trips	• High number of tour types potentially results in low numbers of occurrences per type	Number of trips Number of trips per mode Number of trips per mode per timing Number of trips per purpose per mode
Location-based analysis	• analysing locations of individuals • graphical representations	• no trip characteristics • more activity behaviour than travel behaviour	Number of trips per purpose per location Number of trips
Hazard modelling and survival analysis	• used to analyse the propensity of change • studying intervals between trips	• complex • more activity behaviour than travel behaviour	Number of trips per purpose Analysing interval between leisure trips
Discrete choice modelling	• Modelling behaviour of individuals or small groups of individuals • Comparing behavioural models for consecutive time periods of experiment	• More complex • Requires model assumptions	Number of trips per purpose and per mode Number of trips

3.4 Data requirements and issues

During the experiment it will be necessary to measure a number of variables such as the number of trips, the amount of kilometres travelled, the amount of time spent on travelling, the amount of travel cost, the total emissions, etc. Every trip also comes with its characteristics, such as purpose, location, timing, modality and trajectory. These data are gathered automatically through tripzoom. To be able to detect change in travel behaviour as a result of using tripzoom, ideally, the accuracy of the tripzoom travel measuring functionality would be very high and the travel behaviour of individuals would be very stable on a daily or weekly basis. In that case, minor changes in travel behaviour can be readily detected. However, measuring real life travel behaviour and estimating trip characteristics will not be flawless – as it the case with most experimental context. Moreover, travel behaviour, in terms of trips with their accompanying characteristics, shows variability over time. Although one can state that after 2 to 4 weeks stability in travel behaviour is reached (Schönfelder 2006), this stability only means that the variability in travel behaviour remains approximately constant from that moment on.

Besides accuracy and completeness of the automated trip monitoring, the information on attitudes, beliefs, preferences and perceptions, level-of-service or the perceptions of level-of-service of the various transport modalities should also be complete for every day of the experiment. However there is a practical limit to the number of questions to be asked to participants without creating an unrealistic burden and workload. Asking too many questions will lead to drop-outs from the experiment or from using tripzoom at all. It is therefore necessary to focus on the most relevant aspects of personal characteristics and subjective opinions.

For the experiment in the SUNSET living labs, data are collected by detecting user activities and the incentives the user is exposed to (as in the experimental design). These data are then used to analyse the effectiveness of the SUNSET incentives. Since different (packages of) incentives are offered to different groups, the experimental design allows the analysis of the effectiveness of various incentives by examining different combinations of user groups, as demonstrated by the evaluation schema in Figure 3.3.

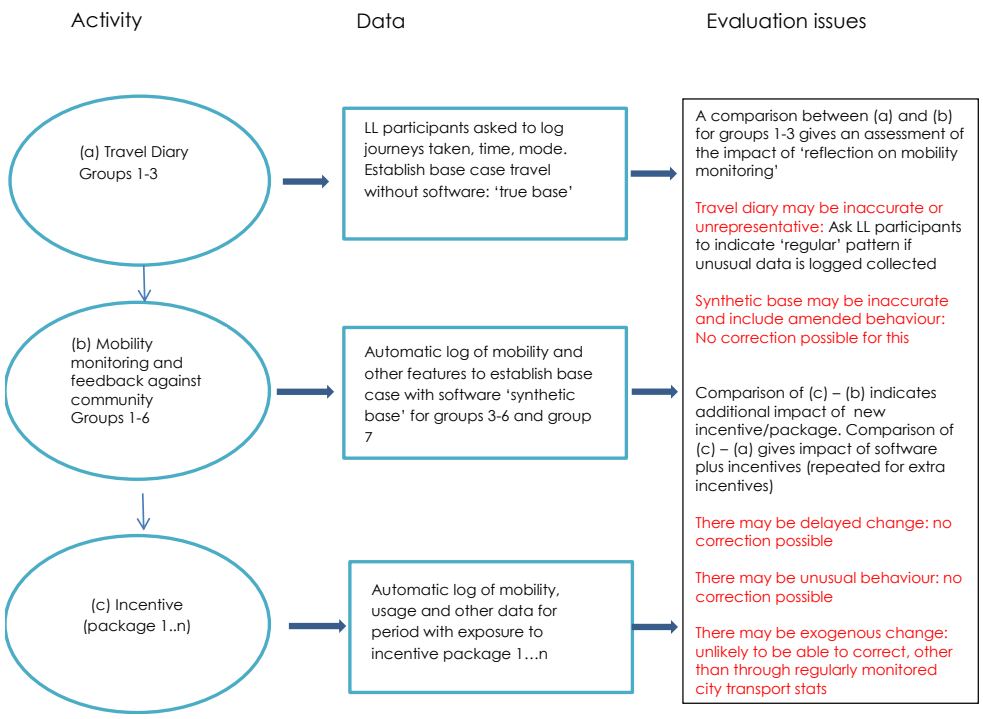

Figure 3.3 Evaluation schema

3.4.1 Behavioural data detected by tripzoom

To collect measurement data on the behavioural indicators, the sensing and recognition algorithms in the tripzoom app are utilised. For each trip that the user makes, three categories of data can be identified based on the collection method:

- *Automatically detected data*
  These are automatically collected by the sensing devices in the app/phone, such as time and GPS location data. Other data such as distance and location also fall into this category because they can be readily derived from the time and location data, with minimum post-processing necessary.

- *Data derived from pattern recognition:*
  These are the trip mode and trip purpose data. They are derived from the pattern recognition algorithms in tripzoom. D2.2 details the classification procedure of trips into different modes and purposes, based on the automatically detected time and location data. Basically, time and location data provide information on the speed profile and proximity to transport infrastructures along a trip's trajectory. These "hints" facilitate the classification of the trip into its most likely mode and purpose.

- *Data that require post-processing*
  Travel cost and emissions are not directly detected by the app. Instead, they are estimated based on the automatically detected data together with estimation models. D2.1 details how travel cost and emissions are estimated based on the available data. See Eq. (0.1) for an example of emission estimation model.

Based on the behavioural indicators discussed in §3.2, Table 3.11 lists the required behavioural data, together with the data collection methods, in order to evaluate the effectiveness of the SUNSET incentives.

Table 3.11 Trip level data detected by tripzoom

Category	Data	Unit	Collection method
Measurement	Number of trips	Count (#)	Detection
	Distance travelled	(km)	Detection
	Travel time	(min)	Detection
	Travel cost	(€, £, kr)	Estimation (post-processing)
	Emissions (CO2)	(kg)	Estimation (post-processing)
Characteristics	Mode	[car, bike, foot, …]	Pattern recognition
	Purpose	[commute, social, …]	Pattern recognition
	Location	[city centre, suburb, …]	Detection & recognition
	Timing	[peak, off-peak, weekends]	Detection
	Trajectory	[highway, corridor, urban roads]	Detection & recognition

3.4.2 Attitude data collected through experience sampling

Subjective attitude data cannot be automatically collected by the sensory capability of the tripzoom app. Instead, questionnaires and the experience sampling mechanism of tripzoom are used, to directly collect attitude data from the tripzoom users.

Based on the attitude indicators discussed in §3.2, Table 3.12 lists the example questions that can be asked. Of course, not all questions should be sent to each single tripzoom user as some may be irrelevant and this could lead to disaffection with tripzoom. It is therefore recommended that participants are only asked the most relevant questions. Relevance can be determined here by checking Table 3.7 and Table 3.8. Table 3.8 shows which group is exposed to which (combination of) incentives, and Table 3.7 lists (for each incentive) the attitude indicators that the incentive is most likely to have impact on. A user should therefore only receive questions which correspond to the attitude indicators that are appropriate for the incentives that the user has been exposed to.

Table 3.12 Attitude data collection via electronic questionnaire and experience sampling

Category	Indicators	Example questions on attitude change	Repetition
Self-awareness	Mobility pattern	*With tripzoom I am more aware of my mobility pattern.* *With tripzoom I know my travel time/cost better.*	QRM
	Mobility impact	*With tripzoom I know more of the impact of my mobility on the environment.*	QRM
Alternative-awareness	Feasibility	*Tripzoom makes me more aware of alternative routes.* *Tripzoom makes me more aware of alternative modes.*	QRM
	Performance	*With tripzoom I know better the travel time/cost of the alternative routes.* *With tripzoom I know better the travel time/cost of the alternative modes.*	QRM
Societal-awareness	Congestion	*Tripzoom makes me more aware of the traffic congestion problem in my city.* *Tripzoom makes me more aware of the economic consequence of traffic congestion in my city.*	QRM
	Environment	*Tripzoom makes me more aware of the environmental problems our society is facing (e.g. global warning).* *Tripzoom makes me more aware of the cost of environmental problems in our society.*	QRM

	Externality	*Tripzoom makes me aware of the environmental damages caused by traffic.* *Tripzoom makes me more aware of the environmental cost of traffic.*	QRM
Self-identification	Categorisation	*I am a driver/cyclist/walker/other.* *I do/not avoid the peak hour.* *I travel little/okay/a-lot.*	QR1
	Preferences	*I've chosen the [mode] because it is the quickest/cheapest /greenest/healthiest/…*	XP
	Satisfaction	*Tripzoom helps me achieve my desired mobility.* *With tripzoom I have more travel freedom.* *Tripzoom improve travel comfort for me.*	QRM
Social attitude	Perception	*Traffic congestion/safety/emissions/noise is a big/no issue.* *Traffic congestion/safety/emissions/noise is/not damaging the livelihood of our community.*	QRM
	Opinion	*Fuel tax should go up/down.* *More roads/railways/cycle-paths should be built.* *Bus should always receive priority at traffic lights.*	QRM

3.4.3 Attribution of changes to the SUNSET incentives

In order to be able to attribute the behavioural and attitude changes to the SUNSET incentives, several requirements need to be met:

- The consideration of the impact of external factors such as weather (e.g. people cycle more in the summer than in the winter), which is based on exogenous data and may be difficult to address fully rigorously
- A sufficiently long experiment period with each measurement period covering at least 2~4 weeks (so that travel behaviour has stabilised)
- A sufficiently large number of participants in each user group (e.g. at least 40 for the main LL, around 20-25 for the reference LL)
- An equal distribution of social-economic characteristics between comparison groups, or inclusion of these characteristics as independent variables in data analysis

In task 6.3 (reported in D6.2), the interpretation of the requirements into the context of the specific LL will be given. This will clarify the exact number of groups to be used, the time period for the lab and the number of participants.

Several other issues also need to be addressed which are discussed in details below.

Measurement errors and data inaccuracy

Data inaccuracy mainly concerns the behavioural data automatically collected by tripzoom. Two types of inaccuracy are distinguished:

- Random errors
- Biases (systematic errors)

A measurement error can be split into a random component and a systematic component. The random error has a zero expected value, but it enlarges the variance in the sample dataset. The systematic error mainly affects the mean value of the sample dataset.

Systematic errors are believed to not significantly impact on the evaluation. This is because the evaluation focuses on the changes between a before period and an after period. The

systematic errors affect the measurement data in both periods. It remains a question whether the magnitude of the system errors changes between the two periods or not. If it does change, then the systematic errors affect the measurement data of the two periods in different ways. Comparison bias is then present if the evaluation does not take this into account. Here, the monitoring of external data can be deployed to address the trend in systematic errors.

The presence of random errors is, however, potentially problematic for the evaluation. The risk here is that the margin of error in measurement may be larger than the actual changes caused by the incentives. The various data analysis and hypothesis testing tools in §3.3 already have provisions for how to deal with random errors in sample data. To be able to attribute the behavioural changes to the SUNSET incentives, a prerequisite knowledge here is the size of the random errors. The use of focus group generated data is intended as a means to estimate the size of these random errors, which can then be accounted for in the evaluation on effectiveness of incentives.

Table 3.13 summarises the known measurement errors in tripzoom. Four points of attention arise:

- The random error in travel cost may affect the evaluation of incentive types (1/3/4).
- The random error in emissions may affect the evaluation of incentive types (1/3/4/5/6/8/9).
- The random error in mode may affect the evaluation of incentive types (1/2/3/5/6/8/9).
- The random error in purpose may affect the evaluation of incentive types (1/2/5/6/7/9).

For mode and purpose, it is not yet known how severe the random errors are. However, tripzoom supports the functionality of allowing users to manually correct their trip mode and purpose. It is therefore expected that the errors in mode and purpose would cause less of a problem for the evaluation (if most users actively correct mode and purpose).

Table 3.13 Random and systematic errors of tripzoom behavioural data

Category	Data	Random error	Systematic error	User correction
Measurement	Number of trips	minor	Missing trips Ghost trips	Only remove trips
	Distance travelled	minor	Missing first segment	N/A
	Travel time	minor	Missing first segment	N/A
	Travel cost	major	?	N/A
	Emissions (CO2)	major (compounding with mode and distance)	?	N/A
Characteristics	Mode	?	?	Possible
	Purpose	?	?	Possible
	Location	minor	minor	N/A
	Timing	minor	minor	N/A
	Trajectory	minor	minor	N/A

Note: ? — unknown yet.

Missing data

Missing data may occur in the behavioural data due to missing detection by tripzoom, or due to participants dropping out. Missing data in the attitude data are the results of non-response from users who have been asked the experience sampling questions. There is the need to provide a financial reward for participation in the LL in order to reduce dropouts, and to ensure a high level of motivation to manually correct travel data and answer the questionnaires.

When dealing with missing data, imputation techniques can be used, which substitute the missing data by certain values so that the dataset is complete. Multiple imputations can be applied to improve reliability. Alternative, deletion techniques can be used, which reduces the dataset. There are two methods (Allison 2001):

- Listwise deletion, which excludes an entire record from the entire analysis if any single value is missing
- Pairwise deletion, which only excludes a record from the current analysis if the record has a missing value in the variables currently being analysed.

Listwise deletion method suffers from reduced statistic power due to the removal of data points. However, pairwise deletion has the problem that each statistic may be based on a different subset of dataset.

For the SUNSET evaluation, the choice between data imputation and listwise/pairwise deletion will be made based on several factors:

- The reason of the missing data: if not random, this would cause the listwise deletion method to introduce selection bias;
- The scope of missing data for a particular participant: if the majority of data are missing for a participant, it may be optimal to exclude the user from its group;
- The size of the dataset: when the sample size is limited, data imputation may be adopted in order to not reduce the statistical power.

Since these factors co-exist, the above choice would be made on an ad hoc basis, under the discretion of the LL coordinator/evaluator.

3.5 Recommendations for SUNSET

3.5.1 Behavioural and attitude indicators

The nine types of incentives in SUNSET are expected to have different effects on the participants. Some incentives may be more effective in invoking behavioural changes, while other incentives may be more powerful in making an attitude shift.

Indicators for measuring the behavioural changes due to incentives are defined by two qualifiers of the traveller's mobility pattern:

- Measurement refers to the trip level performances;
- Characteristics refer to the context and means of each trip.

Table 3.14 highlights for each incentive its most likely areas of impact in the measurements. An indicator can be just an overall measurement, summing up all trips. Or, it can be restricted to trips of certain (combinations of) characteristics. If an incentive is believed to affect only certain sub-categories of trips (based on the characteristics, e.g. only leisure trips by car in the off-peak), then these characteristics should be taken into account in classifying the trips. It can also be the case that an incentive has impact over all or several sub-categories of a characteristic but with different levels of significance, then a more representative indicator would be the distributional statistics over the characteristics.

Take incentive (5) "points without exchange value" for example. Its indicator should be based on the measurements of trip number, travel time, and emissions, and focus on the distribution of these measurements over different modes and trip purposes.

Table 3.14 Indicators for measuring behavioural change

Indicators		Type of incentives (numbered as in Table 3.1)								
		(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Measurement	Number of trips	✓	✓	✓	✓	✓	✓	✓	✓	✓
	Distance travelled				✓					✓
	Travel time	✓	✓	✓	✓	✓	✓			
	Travel cost	✓		✓	✓					
	Emissions	✓		✓	✓	✓	✓		✓	✓
Characteristics	Per mode	✓	✓	✓		✓	✓		✓	✓
	Per purpose	✓	✓			✓	✓	✓		✓
	Per location							✓		
	Per timing		✓							
	Per trajectory								✓	✓

Indicators for measuring the travellers' attitude change due to incentives are listed in Table 3.15. The incentive types (3/4/7/8) are expected to raise self-awareness, types (1/2/5/6/7/9) may raise alternative-awareness, and types (1/2/7) can raise societal-awareness. Self-identification is affected by the incentive types (3/4/8). Social attitude can be changed via the incentive types (1/2/3/8).

Table 3.15 Indicators for measuring attitude change

Indicators		Type of incentives (numbered as in Table 3.1)								
		(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Self-awareness	Mobility pattern			✓						
	Mobility impact			✓	✓			✓	✓	
Alternative-awareness	Feasibility	✓	✓					✓		✓
	Performance	✓	✓			✓	✓	✓		
Societal-awareness	Traffic congestion	✓	✓					✓		
	Environment		✓					✓		
Self-identification	Categorisation			✓	✓					
	Preferences			✓	✓				✓	
	Satisfaction	✓		✓	✓				✓	
Social attitude		✓	✓	✓					✓	

3.5.2 Living lab design

For testing the effectiveness of incentives in the SUNSET living labs, the panel survey method is used where participants are divided into groups. Each group receives a different combination (packages & sequences) of incentives, divided over the periods of living lab operation. An experimental design is set up to manage the issuance of incentives over time and over the groups, together with the data collection methods.

Experimental design

It is recommended to divide the living lab operation into 6 equal-length periods. Users are recruited and classified into six panel groups (Group 1~6), plus a free-form group (Group 7). The six panel groups should have similar social-demographic characteristics, so that no systematic bias is introduced to the evaluation.

The detailed LL design for the main LL is presented in Table 3.16. In D6.2 this will be interpreted into the specific local context for each LL, with a flexible interpretation for the reference labs that are expected to be smaller and focused around particular incentives. Here, various incentives (numbered according to Table 3.1) are offered, depending on the time period and the user group. Data are being collected across all six time periods and for all groups. Quantitative data are automatically collected by tripzoom, while qualitative data are collected through questioning the user (thus requiring the action of the LL coordinator). More recommendations on data collection are presented in the following subsection.

Table 3.16 Participant groups and data collection for the living lab operation

LL operational design		Time period					
		1	2	3	4	5	6
LL user groups	1	Travel diary	*(3&4)*	*(1-5-6)*		*(2&7-8-9)*	
	2	Travel diary	*(3&4)*	*(2&7-8-9)*		*(1-5-6)*	
	3	Travel diary	*(3&4)*				
	4	*(3&4)*	*(1-3-6)*				
	5	*(3&4)*	*(2&7-8-9)*				
	6	*(3&4)*	*(1-5-6) – (2&7-8-9)*				
	7	*(1&2&3&4&5&7&8)*					
Data	Tripzoom	*--- detection & recognition & estimation ---*					
	Question	*--- QR1 / QRM / XP ---*					

Note: numbers in brackets denote the types of incentives; "&" represents a package, while "-" represents a sequence.

Sample sizes

Ideally, each user group should consist of at least 40 users. For a smaller living lab, we can use as few as 10-15 individuals in each group but will then need to amend the statistical approach for testing. Instead of focusing on specific objectives, we may focus on the extent to which tripzoom as a whole contributes to a change in behaviour and/or attitude.

It is also advisable that, during the formation stage of groups in the experiment, the LL coordinator should follow a prioritization plan, in order to ensure that at least the most important groups contain sufficient sample size.

Design of challenges

The type (5/6) incentives are challenges with points rewarded to the user. Table 3.17 provides a list of proposed challenges that can be used during the LL operations. For users who are already responding to incentives and making more sustainable choices, Table 3.18 further lists some "re-enforcing" challenges.

Table 3.17 Proposed challenges for testing in SUNSET

Challenges	To whom	Personal objective	System objective
If you are using the bus again tomorrow, why not get off a stop earlier and walk instead? (750 points)	Regular bus users who walk <500m at destination	Health	Maximise use of PT capacity, reduce overcrowding, health of citizens
If you travel again tomorrow, why not leave the car at home and walk to the station? You will save money, be healthier and help the environment. (1,000 points)	Regular car users, with station < 1.0 km of origin, who don't normally walk to station	Health, travel costs, Eco-life	Maximise use of parking capacity, health of citizens, energy use and environmental
If you travel again tomorrow, you could reduce your costs and help the	Regular car users who don't also	Travel costs, Eco-life	Reduce congestion, energy and

environment by using the bus or train - the timetable link is here (*insert link!)(*2000 points)	take train or bus on regular journey		environment
If you travel again tomorrow why not try walking? You could reduce your costs, be healthier and help the environment. (1,000 points)	Car, or bus users with regular destination within 2 km of origin	Health, travel costs, Eco-life	Reduce congestion, energy and environment
Fed up of delays on the (*insert target route*) – why not use the (*insert diversion route*) instead? (750 points)	Car users with regular route	Fuel use, improve travel time and reliability	Reduce Congestion, energy and environment
If you travel again tomorrow, instead of the taking the (*insert target route*) why not use the (*insert diversion route*) to avoid the (*hospital/school*) traffic? (750 points)	Car users with regular route past hospital, school and no children	Reduce delays	Noise reduction, safety and environment in sensitive areas
There are likely to be road works on the (*target route*) why not use the (*insert diversion route*) if you travel (insert *tomorrow, in the next week, fortnight*)? (750 points)	Car users with regular route affected by road works	Fuel use, improve travel time and reliability	Reduce Congestion, energy and environment
There has been an incident on the (*insert route*); we suggest you try the (*insert diversion*) instead if you travel today? (750 points)	Car users with regular route affected by accident	Fuel use, improve travel time and reliability	Reduce Congestion, energy and environment
A pollution hot-spot in (*insert location*) is building up – if you travel today could you use the (*insert diversion*) as an alternative route? (1,000 points)	Car users with regular route through pollution hotspot	Eco-life	Reduce Congestion, energy and environment

Table 3.18 Re-enforcing challenges for testing in SUNSET

Challenges	To whom
If you reduce your car use by another 5 miles (?km) this week, you will save energy, save fuel costs and help the environment (1,000 points)	Any car users who have reduced their mileage on regular trips over the past week (stayed at home, walked, travelled by other modes, done a rideshare)
If you increase your walking activity this week by another 2km, you will be healthier, help the environment and save travel costs (1,000 points)	Travellers by any mode who have increased their walking activity related to regular trips in the past week
If you use the bus or train more and your car less again this week, you will save travel costs and help the environment (1,000)	Car users who used reduced car use and increased bus or train use in previous week

3.5.3 Data collection

Data are collected from three sources: tripzoom, travel diary, and questionnaire. Tripzoom and travel diary record the user's trip data, while electronic questionnaire and experience sampling are used to collect data on user attitude. To mitigate the problem of missing data, it is recommended that the LL coordinator should provide a financial reward for participation, so that:

- A larger group of participants can be attracted, and also, user retention would no longer be a big problem;

- Users are more motivated to correct their travel data;
- Users are more willing to answer questionnaires.

Detection by tripzoom
For each trip that the user makes, the following data are collected by tripzoom:

- Measurements: trip distance, time, cost, and emissions
- Characteristics: mode, purpose, location, timing, trajectory

When a user makes multiple (repeated) trips, these measurements are summed up to form the user's mobility profile. Trip characteristics are used to classify the trips into different categories, so that changes in behaviour can be compared for the same category of trips before and after. Thus the trip measurements also need to be divided into the categories (as defined by the pre-selected characteristics). Typical categories are:

- Car trips
- Commute trips
- Commute trips on the car
- Trips in the peak hour
- Etc.

Collection through experience sampling
From time to time, electronic questionnaires are deployed to examine the (changes in) user attitude. All awareness-related attitude indicators, including self-, alternative-, and societal-awareness, can be measure by asking the corresponding experience sampling questions once (i.e. QR1) at the end of the measurement period. For opinion-related indicators,

- Self-categorisation questions should be asked only once (QR1);
- Preference questions should be asked frequently (at trip ends) (XP);
- Satisfaction questions need to be asked two or multiple times (QRM);
- Social attitude questions need to be asked two or multiple times (QRM).

3.5.4 Evaluation methods

The nature of the SUNSET project is a panel-like survey allowing a longitudinal study to evaluate the effectiveness of incentives. Therefore, an experimental design is recommended to monitor participants of the experiment for a longer period of time and introducing them to the various incentives and incentives groups developed in SUNSET. Moreover, the participants of the experiment are divided into 7 groups to be able to isolate the effects of specific (groups of) incentives. During the experiment the travel behaviour of the participants is registered by the tripzoom application, collecting the trips and its characteristics (e.g. trip distance, travel time, purpose, modality, departure time, route, etc.). Simultaneously, information about personal characteristics and attitudes affecting travel choices are gathered via questionnaires. This approach allows for the analysis of all factors influencing travel choice as presented in Figure 3.1.

The analysis of behavioural change is conducted in two stages. Firstly, the overviews of behavioural indicators described in §3.2 are compared for the consecutive time periods (i.e. base cases and the cases where various incentives or groups of incentives are offered). This will provide an aggregate overview on the effectiveness of Tripzoom in for example reducing the amount of vehicle kilometres travelled during rush hours and reducing the amount of CO2 emission. Secondly, it is recommended that a tour-based approach is used to investigate changes on trip and trip sequence level. Tour-based analysis offers a method to capture 'average/regular' travel behaviour in a number of tour types and allows for studying sequential

nature of travel behaviour. Moreover, the method can also incorporate factors influencing travel choice into the tour types. However, there is a trade-off between the number of variables incorporated in the tour types and the actual number of tours per type to analyse. The tour-based analysis will support to the more aggregate overviews of behavioural change.

In the data analysis, the random errors in the measurement data need to be taken into account. For missing data, imputation and/or deletion techniques can be used, although it is more preferable that the LL coordinator attempts to reduce the scale of missing data by encouraging participants to manually correct missing data and to answer all experience sampling questions.

4. Conclusions

This deliverable aims to provide methodologies for evaluating the operational success of tripzoom, and the effectiveness of the SUNSET incentives. Operational success is defined here by four indicators:

- The ability to meet travellers' individual goals (IG)
- The success of the social network concept (SN)
- The tripzoom functionality as intended (Fun)
- The usability of tripzoom (Usa)

Effectiveness is defined by the ability of an incentive to invoke changes in:

- Travellers' behaviour (i.e. their mobility profile) (Beh)
- Travellers' attitude (Att)

For both evaluation approaches, measurements are taken as input to derive indicators for success and effectiveness. It is inevitable that certain measurement contributes to several indicators but this potential for double-counting is not problematic as long as the indicators are not aggregated to composite indicators. As an overview, Table 4.1 lists the required measurements in order to derive the value of an indicator.

Table 4.1 Measurements and indicators for success and effectiveness

Measurements		**Operational success**				**Effectiveness**	
		IG	SN	Fun	Usa	Beh	Att
User profile	Social-economic	✓				✓	✓
	Mobility constraint	✓				✓	✓
tripzoom	Mobility profile	✓				✓	
	Friends		✓				
Travel diary	Mobility profile					✓	
Questionnaire	Self-categorisation	✓					✓
	Preferences	✓					✓
	Awareness						✓
	Satisfaction	✓			✓		✓
	Rating of tripzoom	✓	✓	✓	✓		
Testing	Functionality			✓	✓	✓	
	Usability			✓	✓		✓
LL operation	Participants		✓			✓	✓
Google Analytics	Portal usage		✓				
	App usage		✓				
Radian6	Sentiment		✓				

The measurement data are collected from various sources:

- User administration data, including user profile data on their social economic characteristics and mobility constraints (if any)
- Data collected by tripzoom, including the user's mobility profile (movement, context, and impact) and the profile of their tripzoom social network (e.g. number of friends)
- Mobility profile collected through travel diary

- Data collected through electronic questionnaires and experience sampling, including attitude and rating data
- Data collected during the testing phases, on both functionality and usability
- Administration of participants during the SUNSET living labs, especially on the number of participants and how they get to know tripzoom
- Google Analytics and Radian6 data, on users' browsing behaviour on the portal and the app

4.1 Success of the tripzoom app

There are several travel goals that individual travellers may wish to gain when making a trip, such as to minimise or maximise travel time, costs, and pleasure. To understand how well tripzoom helps an individual achieve the desired travel goals, personal objectives are first identified. Users are then asked about their perception on how well tripzoom app helps in achieving those goals. The following set of **subjective indicators** is proposed as Key Performance Indicators for **operational success**:

- travel time, distance and cost by trip purpose and distance
- scheduling effort by trip purpose and distance
- experience with social networks (e.g. motivational support, feedback, and satisfaction)
- output (the added value of tripzoom app in relation to each of the desired mobility)

In order to evaluate the operational success based on the above mentioned KPIs a number of additional indicators are required to scrutinize and explain the results based on groups of individuals

Prerequisites

- Indicators on socio-demographics and household characteristics (e.g. age, gender)
- Preferred mode and type
- Accessibility to modes
- Mobility constraints
- Number of trips by trip purposes (work related or leisure)
- Number of trips by distance (short vs. long)
- Total kilometres travelled
- Total travel time by trip purpose and distance
- Objective indicators on social networks (e.g. number of friends)
- Attributes related to travel dislike
- Attitudes on pro-environmental policy
- Attitudes on commute benefit
- Attitudes on travel freedom
- Attitudes on pro-high density
- Attitudes on travel stress
- Personality
- Lifestyle
- Excess travel
- Desired mobility (with regard to travel goals by trip purpose and distance)

To define the **success of the social network concept** in SUNSET, the following Key Performance Indicators are proposed:

- The number of unique visitors to the web portal
- The number of people that register for tripzoom

- The number of people that agree to participate in a living lab
- The number or percentage of participants that are recruited using a LL/city-FB site
- The number or percentage of participants that are recruited using a FB-add
- The number or percentage of participants that are recruited via a friend on an external social network
- The number or percentage of participants that are recruited via a friend invite by email
- The number or percentage of participants engaged in the tripzoom social network
- The number or percentage of participants linked in as 'friends' in the tripzoom social network either actively or passively
- The number of mutual friends within the local tripzoom social network
- The number of messages posted on external social media that relate to tripzoom
- The percentage of positive messages on social media that relate to tripzoom

For a **successful product** with respect to the **functionality** and **usability** from an end-user's point-of-view, it is recommended that a design approach is applied with the following characteristics:

- End-users are involved during every stage of development, providing feedback to concepts and prototypes regarding functionality and usability
- The design process proceeds in iterations
- Evaluation (partly) occurs in real-life contexts, increasing ecological validity

For discovering user needs and evaluating concepts in an early stage of development, it is recommended that feedback is gathered in two ways:

- Quantitative user data from large user groups applying surveys
- Qualitative user data from smaller groups applying focus-groups and interviews

During development, it is recommended that evaluation is undertaken with several iterations in the following ways:

- Ad-hoc functionality testing by developers
- Functionality testing by non-developers (in the role of potential users of the system) whilst executing a structured task list, whereby developers are notified of bugs and not properly working functions
- Usability testing by non-developers in the role of usability experts and executing the same structured task list, by applying evaluation techniques such as heuristic evaluation and cognitive walkthrough, whereby usability issues are reported to developers
- Functionality and usability evaluation with real end-users, by applying evaluation techniques such as think-aloud protocol, interviews, focus groups, field testing, experience sampling, and surveys

The system is **successful from a functionality perspective** if the requirements set by the technical work packages in the SUNSET project are met. Inevitably these requirements will be adjusted during the development process. The functionality that the SUNSET system offers is then a mixture of those functions described in the DoW, feedback from evaluations, and feasibility (i.e. whether it was possible to develop during the project lifetime).

Regarding basic, general functionality that apply to most apps, it is recommended that the system should provide acceptable levels for the end-users regarding:

- Battery usage of the mobile phone
- Interference with other uses of the mobile phone

- Robustness –ability to work or at least not crash with network problems, ability to recover unobtrusively, ability to signal to a user that there is a problem to contact the developer
- Security – ability to maintain confidentiality and integrity of data
- Privacy – have sufficient procedures in the operation to support privacy

The extent to which acceptable levels have been reached should be evaluated by the end-users, who can give scores on acceptability using a psychometric scale. How acceptable levels of security and privacy from an objective point-of-view are supported is described elsewhere.

Specific to the web portal, it is recommended that:

- The information on the website provides end-users with an understandable view on what SUNSET is about, and what registering and installing the app will mean to them
- The website provides helpdesk/FAQ to support users in the interaction with the system
- The website provides clear and usable instructions for how to join
- The process of joining (registering, installing the app) should be user-friendly
- Interaction login using other social media (Facebook, Twitter) is provided in a user-friendly way
- Users can communicate with other users in a user-friendly way

Specific to the tripzoom app, it is recommended that from and end-user's point-of-view:

- Interaction with the app is user-friendly (easy to learn, efficient, effective, satisfying) regarding all specific functions
- The functionality that the app offers provides added value to the user
- Interaction login and account links when using other social media (Facebook, Twitter) is provided in a user-friendly way
- Trips are detected with acceptable accuracy (locations, routes, modalities)
- Communication regarding inaccuracies is presented in a user-friendly way
- Overviews of travel behaviour are presented in a user-friendly way
- Comparisons of the user's travel behaviour with the travel behaviour of other users is presented in a user-friendly way
- Any communication regarding challenges is handled in a user-friendly way (see also Chapter 3 of this document)
- Any communication regarding experience sampling questions is handled in a user-friendly way
- The user has control over settings regarding privacy
- Interactions with friends is supported in a user-friendly way

Specific to the city dashboard, it is recommended that from the point-of-view of a LL co-ordinator (or researcher):

- It is possible to interact with the (end-)user database in a user-friendly way
- Challenges can be issued in a user-friendly way to specific groups of users
- Overviews of responses to challenges can be viewed in a user-friendly way
- Experience sampling questions can be issued in a user-friendly way to specific groups of users
- Overviews of responses to Experience sampling questions can be viewed in a user-friendly way

4.2 Effectiveness of incentives

The nine types of incentives in SUNSET are expected to have different effects on travellers. Some incentives are more effective in invoking behavioural changes, while other incentives are more powerful in making an attitude shift. Indicators for measuring the **behavioural changes** due to incentives are defined by two qualifiers of the traveller's mobility pattern:

- Measurement refers to the trip level performances;
- Characteristics refer to the context and means of each trip;

Key performance indicators are

- number of trips
- distance travelled
- travel time
- travel cost
- CO2 emission

Where all trips are characterized by mode, purpose, timing, route and location.

Key performance Indicators for measuring the travellers' **attitude change** due to incentives are summarised below:

- Awareness of (the impact of) the personal mobility pattern;
- Awareness of the existence and/or performance of alternatives (modes, routes, etc.);
- Awareness of the societal impact of traffic (externalities);
- Self-categorisation, preferences, satisfaction, social attitude.

For testing the effectiveness of incentives in the SUNSET living labs, the panel survey method is used where participants are divided into groups. Each group receives a different combination (packages & sequences) of incentives, divided over the periods of living lab operation. An experimental design is set up to manage the process of issuing incentives over time and over the groups, together with the data collection methods. Ideally, each user group should consist of at least 40 users. For a smaller living lab, we can use as few as 10-15 individuals in each group but will then need to amend the statistical approach for testing. Instead of focusing on specific objectives, we may focus on the extent to which tripzoom as a whole contributes to a change in behaviour and/or attitude. It is also advisable that during the formation stage of groups in the experiment the LL coordinator should follow a prioritization plan, in order to ensure that at least the most important groups contain a sufficient sample size.

Data are collected from three sources: tripzoom, travel diary, and questionnaires. Tripzoom and travel diary record the user's trip data, while electronic questionnaire and experience sampling are used to collect data on user attitude. To mitigate the problem of missing data, it is recommended that a financial reward should be provided for participation in the LL, so that a larger group of participants can be attracted (also, user retention could be less of a problem), and users are more motivated to manually correct their travel data and to answer questionnaires.

The analysis of behavioural change is conducted in two stages. Firstly, the overviews of behavioural indicators are compared for the consecutive time periods (i.e. base cases and the cases where various incentives or groups of incentives are offered). This will provide an aggregate overview of the effectiveness of Tripzoom in for example reducing the amount of vehicle kilometres travelled during rush hours and reducing the amount of CO2 emission.

Secondly, it is recommended to use a tour-based approach to investigate changes on trip and trip sequence level. Tour-based analysis offers a method to capture 'average/regular' travel behaviour in a number of tour types and allows for studying sequential nature of travel behaviour. Moreover, the method can also incorporate factors influencing travel choice into the tour types. This is especially useful for SUNSET when the effectiveness of challenges and when to offer them are concerned. However, there is a trade-off between the number of variables incorporated in the tour types and the actual number of tours per type to analyse. The tour-based analysis will support to the more aggregate overviews of behavioural change.

In the data analysis, the random errors in the measurement data need to be taken into account. For missing data, imputation and/or deletion techniques can be used, although it is more preferable that the LL coordinator attempts to reduce the scale of missing data by encouraging participants to manually correct missing data and to answer all experience sampling questions.

References

Ajzen, I., 1991. The theory of planned behaviour. *Organizational Behavior and Human Decision Processes*, 50(2), 179-211.

Aleahmad, T., Balakrishnan, A.D., Wong, J., Fussell, S.R., Kiesler, S., 2008. Fishing for sustainability: The effects of indirect and direct persuasion. Conference on Human Factors in Computing Systems, 3021-3026.

Allison, P. D., 2001. *Missing Data*. Sage University Papers Series on Quantitative Applications in the Social Sciences, 07-136. Thousand Oaks, CA: Sage.

Asur. S and Huberman, B, A., 2010. Predicting the Future with Social Media. eprint arXiv:1003.5699. 2010arXiv1003.5699A.

Axhausen, K.W. & Gärling, T., 1992. Activity - based approaches to travel analysis: conceptual frameworks, models, and research problems. *Transport Reviews*, 12, 323-341.

Axhausen, K.W., 1997. Presenting and Preserving Travel Data. In Transport Surveys: Raising the Standard. Grainau, Germany: Transportation Research Board. Available at: http://pubsindex.trb.org/document/view/default.asp?lbid=686594 [Accessed June 2, 2009].

Axhausen, K.W. et al., 2002. Observing the rhythms of daily life: A six-week travel diary. *Transportation*, 29(2), 95–124.

Bekkers, V., Moody, R. & Edwards, A., 2011. Micro-Mobilization, Social Media and Coping Strategies: Some Dutch Experiences. Policy & Internet, 3(4). Available at: http://www.psocommons.org/policyandinternet/vol3/iss4/art6 [Accessed October 10, 2012].

Binsted, A. & Hutchins, R., 2012. The role of social networking sites in changing travel behaviours, UK. Available at: http://www.trl.co.uk/online_store/reports_publications/trl_reports/cat_traffic_and_the_environment/report_the_role_of_social_networking_sites_in_changing_travel_behaviours.htm [Accessed July 5, 2012].

Brög, W., Erl, E., Ker, I., Ryle, J. & Wall, R., 2009. Evaluation of voluntary travel behaviour change: Experiences from three continents. *Transport Policy*, 16(6), 281-292.

Centola, D., 2010. The Spread of Behavior in an Online Social Network Experiment. *Science*, 329(5996), 1194–1197.

Ceron, A., Curini, L., Iacus, S. & Porro, G., 2012. "Every tweet counts? How sentiment analysis of social networks can improve our knowledge of citizens' policy preferences. An application to Italy and France" UNIMI - Research Papers in Economics, Business, and Statistics. Statistics and Mathematics. Working Paper 58. Available at: http://services.bepress.com/unimi/statistics/art58 [Accessed October 10, 2012].

Chatterjee, K., 2009. A comparative evaluation of large-scale personal travel planning projects in England. *Transport Policy*, 16, 293-305.

Consolvo, S. & Walker, M., 2003. Using the Experience Sampling Method to Evaluate Ubicomp Applications. *Pervasive Computing*, 2(2), 24-31.

Constantine, L.L., 2009. Human Activity Modeling: Toward A Pragmatic Integration of Activity Theory and Usage-Centered Design. In A. Seffah, J. Vanderdonckt, & M. C. Desmarais, eds. Human-Centered Software Engineering. Human–Computer Interaction Series. Springer

London, pp. 27–51. Available at: http://www.springerlink.com/content/wrk62572w6516004/abstract/ [Accessed October 10, 2012].

EC-ISM European Commission, 2009. Living Labs for User-driven Open Innovation – An Overview of the Living Labs Methodology, Activities and Achievements, Luxembourg: Office for Official Publications of the European Communities. Available at: http://ec.europa.eu/information_society/activities/livinglabs/docs/brochure_jan09_en.pdf.

ENoLL, 2012. European Network of Living Labs. Available at: http://www.openlivinglabs.eu/ [Accessed October 10, 2012].

Ericsson, K.A., & Simon, H.A., 1993. *Protocol analysis: Verbal reports as data*. MIT Press, Cambridge, MA.

Eriksson, M., Niitamo, V. & Kulkki, S., 2005. State-of-the-art in utilizing Living Labs approach to user-centric ICT innovation - a European approach, Available at: http://www.vinnova.se/upload/dokument/verksamhet/tita/stateoftheart_livinglabs_eriksson2005.pdf [Accessed October 10, 2012].

Fenner, Y. et al., 2012. Web-Based Recruiting for Health Research Using a Social Networking Site: An Exploratory Study. *Journal of Medical Internet Research*, 14(1), p.e20.

Frees, E.W., 2004. *Longitudinal and panel data: analysis and applications in the social sciences*, Cambridge University Press.

Gärling, T., Gärling, A. & Johansson, A., 2000. Household choices of car-use reduction measures. *Transportation Research Part A: Policy and Practice*, 34, 309-320.

Gerhardt-Powals, J., 1996. Cognitive engineering principles for enhancing human - computer performance. *International Journal of Human-Computer Interaction*, 8(2), 189-211.

Golob, T.F., 1986. A nonlinear canonical correlation analysis of weekly trip chaining behaviour. *Transportation Research Part A: General*, 20, 385-399.

Gray, D., 2009. Doing research in the real world, 2nd edition, London: SAGE.

Handy, S., Weston, L. & Mokhtarian, P., 2005. Driving by choice or necessity? *Transportation Research Part A: Policy and Practice*, 39(2-3), 183–203.

Herfindahl, O.C., 1950. *Concentration in the steel industry*. Columbia university.

Huff, J. O. & Hanson, S., 1986. Repetition and variability in urban travel. *Geographical Analysis*, 18, 97-114.

Hunecke, M., Haustein, S., Böhler, S., Grischkat, S., 2010. Attitude-based target groups to reduce the ecological impact of daily mobility behaviour. *Environment and Behavior*, 42(1), 3-43.

ISO 9126-2/3, 2003. ISO 9126-2/3. Available at: http://www.iso.org/iso/iso_catalogue/catalogue_tc/catalogue_detail.htm?csnumber=22749 [Accessed May 25, 2011].

ISO 9241-11, 1998. ISO 9241-11. Available at: http://www.iso.org/iso/catalogue_detail.htm?csnumber=16883 [Accessed May 25, 2011].

ISO 13407, 1999. ISO 13407. Available at: http://www.iso.org/iso/catalogue_detail.htm?csnumber=21197 [Accessed May 25, 2011].

Jarvis, S., Barbarena, L., 2008. Focus group, In Ed. Lavrakas, P., *Encyclopedia of survey research methods*, Vol.1, London: SAGE.

Joh, C.-H., Arentze, T., Hofman, F. & Timmermans, H., 2002. Activity pattern similarity: a multidimensional sequence alignment method. *Transportation Research Part B: Methodological*, 36, 385-403.

Joh, C. H., Arentze, T. & Timmermans, H., 2001. Pattern recognition in complex activity travel patterns: comparison of Euclidean distance, signal-processing theoretical, and multidimensional sequence alignment methods. *Transportation Research Record: Journal of the Transportation Research Board*, 1752, 16-22.

Kitamura, R. & Hoorn, T., 1987. Regularity and irreversibility of weekly travel behavior. *Transportation*, 14, 227-251.

Kraut, R.E., Resnick, P., Kiesler, S., Ren, Y., Chen, Y., Burke, M., ... & Konstan, J., 2012. *Building successful online communities: Evidence-based social design*. The MIT Press.

Krizek, K.J., 2003. Neighborhood services, trip purpose, and tour-based travel. *Transportation*, 30, 387-410.

Kwak, H., Choi, Y., Eom, Y.H., Jeong, H., & Moon, S., 2009. Mining communities in networks: a solution for consistency and its evaluation. In Proceedings of the 9th ACM SIGCOMM conference on Internet measurement conference (pp. 301-314). ACM.

Mankoff, J., Fussell, S. R., Dillahunt, T., Glaves, R., Grevet, C., Johnson, M., ... & Setlock, L., 2010. StepGreen. org: Increasing energy saving behaviors via social networks. In Proceedings of the Fourth International AAAI Conference on Weblogs and Social Media (pp. 23-25).

Melville, P., Sindhwani, V. & Lawrence, R., 2009. Social media analytics: Channeling the power of the blogosphere for marketing insight. Proc. of the WIN.

Mokhtarian, P., 2005. Travel as a desired end, not just a means. *Transportation Research Part A: Policy and Practice*, 39(2–3), pp.93–96.

Nielsen, J., 1993. *Usability Engineering* 1st ed., San Diego: Morgan Kaufmann.

Nielsen, J., 1994. Heuristic evaluation. In J. Nielsen & R. L. Mack, eds. *Usability Inspection Methods*. New York: Wiley.

Nielsen, J., 2008. Usability ROI Declining, But Still Strong. Available at: http://www.useit.com/alertbox/roi.html [Accessed October 10, 2012].

Nielsen, J. & Mack, R.L. eds., 1994. *Usability Inspection Methods* 1st ed., New York: Wiley.

Næss, P., 2003. Urban structures and travel behaviour. Experiences from empirical research in Norway and Denmark. *European Journal of Transport and Infrastructure Research*, 3, 155-178.

Norman, D., 1990. *The Design of Everyday Things*, New York: Doubleday Business.

Norman, D., 2005. "Human-Centered Design Considered Harmful." *Interactions*, 12(4), 14-19.

Novick, D., Hollingsed, T., 2007. "Usability inspection methods after 15 years of research and practice" (2007). Departmental Papers (CS). Paper 16. http://digitalcommons.utep.edu/cs_papers/16 [last accessed on 5/11/2012].

O'hEocha, C., Wang, X. & Conboy, K., 2012. The use of focus groups in complex and pressurised IS studies and evaluation using Klein & Myers principles for interpretive research. *Information Systems Journal*, 22, 235–256.

Ortúzar, J.D.D., Armoogum, J., Madre, J.-L. & Potier, F., 2010. Continuous Mobility Surveys: The State of Practice. *Transport Reviews*, 31, 293-312.

Ory, D.T. & Mokhtarian, P., 2005. When is getting there half the fun? Modeling the liking for travel. *Transportation Research Part A: Policy and Practice*, 39(2-3), pp.97–123.

Pas, E.I., 1983. A flexible and integrated methodology for analytical classification of daily travel-activity behavior. *Transportation Science*, 17, 405-429.

Pitts, M.G. & Browne, G.J., 2007. Improving requirements elicitation: an empirical investigation of procedural prompts. *Information Systems Journal*, 17(1), 89–110.

Pontes, T., Vasconcelos, M., Almeida, J., Kumaraguru, P., & Almeida, V., 2012. We Know Where You Live: Privacy Characterization of Foursquare Behavior. UbiComp '12, Sep 5-Sep 8, 2012, Pittsburgh, USA.

Prochaska, J. & DiClemente, C., 1983. Stages and processes of self-change of smoking: Toward an integrative model of change. *Journal of Consulting and Clinical Psychology*, 51(3), 390–395.

Prochaska, J. & Velicer, W., 1997. The transtheoretical model of health behavior change. *American journal of health promotion: AJHP*, 12(1), 38–48.

Redecker, C., Ala-Mutka, K., Bacigalupo, M., Ferrari, A. & Punie, Y., 2009. Learning 2.0: The Impact of Web 2.0 Innovations on Education and Training in Europe – Final Report, Available at: http://ftp.jrc.es/EURdoc/JRC55629.pdf [Accessed November 14, 2012].

Richardson, A., Seethaler, R. & Harbutt, P., 2004. Design issues for before and after surveys of travel behaviour change. *Transport Engineering in Australia*, 9, 103-118.

Saunders, M., Lewis, P., Thornhill, A., 2009. *Research Methods for Business Students*, 5th edition, Pearson Education Ltd, FT Prentice Hall: Harlow.

Schlich, R. & Axhausen, K.W., 2003. Habitual travel behaviour: Evidence from a six-week travel diary. *Transportation*, 30, 13-36.

Schönfelder, S., 2006. *Urban rhythms: Modelling the rhythms of individual travel behaviour*. ETH.

Schönfelder, S. & Axhausen, K.W., 2002. *On the variability of human activity spaces*. ETH.

Shackel, B., 1991. Human factors for informatics usability. In B. Shackel & S. J. Richardson, eds. *Human factors for informatics usability*. Cambridge: Cambridge University Press, pp. 21–37. Available at: http://dl.acm.org/citation.cfm?id=117829.117833 [Accessed September 3, 2012].

Shoval, N. & Isaacson, M., 2007. Sequence Alignment as a Method for Human Activity Analysis in Space and Time. *Annals of the Association of American Geographers*, 97, 282-297.

Sluzki, C.E., 2010. Personal social networks and health: conceptual and clinical implications of their reciprocal impact. *Families, systems & health: the journal of collaborative family healthcare*, 28(1), 1–18.

Spencer, R., 2000. The streamlined cognitive walkthrough method, working around social constraints encountered in a software development company. In Proceedings of the SIGCHI conference on Human factors in computing systems. CHI '00. New York, NY, USA: ACM, pp. 353–359. Available at: http://doi.acm.org/10.1145/332040.332456 [Accessed October 10, 2012].

Spitsmijden, 2009. *Leerervaringen Spitsmijden*, Den Haag.

Stopher, P., Clifford, E., Swann, N. & Zhang, Y., 2009. Evaluating voluntary travel behaviour change: Suggested guidelines and case studies. *Transport Policy*, 16, 315-324.

Stopher, P. & Greaves, S., 2007. Guidelines for samplers: measuring a change in behaviour from before and after surveys. *Transportation*, 34, 1-16.

Stopher, P. & Zhang, Y., 2011. Repetitiveness of Daily Travel. *Transportation Research Record: Journal of the Transportation Research Board*, 2230, 75-84.

Stopher, P., Zhang, Y. & Jiang, Q., 2010. Tour-based analysis of multi-day GPS data. In: 12th World Congress on Transport Research, 2010 Lisbon, Portugal.

Tertoolen, G., van Kreveld, D. & Verstraten, B., 1998. Psychological resistance against attempts to reduce private car use. *Transportation Research Part A: Policy and Practice*, 32(3), 171–181.

Tomitsch, M., Singh, N, Javadian, G., 2010. Using diaries for evaluating interactive products: the relevance of form and context. *Proceedings of the Conference of the Australian Computer-Human Interaction* (OZCHI'10), ACM Press.

UXPA, 2012. *What is User-Centered Design?* Available at: http://www.usabilityprofessionals.org/usability_resources/about_usability/what_is_ucd.html [Accessed October 28, 2012].

Weinschenk, S. & Barker, D. 2000. *Designing Effective Speech Interfaces*, Wiley.

Wharton, C. et al., 1994. The cognitive walkthrough method: A practitioner's guide. In J. Nielsen & R. L. Mack, eds. *Usability Inspection Methods*. New York: Wiley.

Wismans, L., van Berkum, E., Bliemer, M. (2011). Modelling externalities using dynamic traffic assignment models: A review. Transport Reviews 31: 521-545.

Appendix A. Hypothesis testing

For indicators where two subsets of data are available (i.e. a 'before' set and an 'after' set), hypothesis testing techniques can be applied to assess the statistical significance of changes in travellers' behaviour and attitude.

Regarding the indicators of change, two types of variables are distinguished:

- Changes in numeric value of the variable, e.g. the user makes fewer car trips now than before;
- Changes in proportions, e.g. the proportion of car trips has decreased (among trips of all modes, whilst the absolute number of car trips may have increased, decreased, or remained the same).

Different hypothesis testing techniques may be employed, as discussed below.

Testing changes in proportion
Suppose that we have two samples (perhaps a 'before' sample and an 'after' sample), with

$$X_1 \sim \mathbf{B}_i(n_1, p_1),$$
$$X_2 \sim \mathbf{B}_i(n_2, p_2),$$

If n_1 and n_2 are both sufficiently large to apply the Normal approximation to each sample, then a two sample test of proportions may be used. We say that the number n is sufficiently large if it satisfies the following condition:

$$n > 9 \cdot \max\left[\frac{p}{1-p}, \frac{1-p}{p}\right].$$

A two-sample test of proportions may be developed based on the result that, approximately:

$$\frac{X_1}{n_1} - \frac{X_2}{n_2} \sim \mathbf{N}\left(p_1 - p_2, \frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}\right).$$

Estimating the variance term by substituting $\hat{p}_1 = x_1/n_1$ and $\hat{p}_2 = x_2/n_2$ in place of p_1 and p_2, then to test a hypothesis that $H_0 : p_1 = p_2$ (i.e. $H_0 : p_1 - p_2 = 0$), we have that under H_0, approximately

$$\frac{X_1}{n_1} - \frac{X_2}{n_2} \sim \mathbf{N}\left(0, \frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}\right).$$

This suggests a logical test statistic to be (based on a Z-transformation):

$$z = \frac{x_1/n_1 - x_2/n_2}{\sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}},$$

with the statistical significance of the value then deduced from standard Normal tables.

Testing changes in means and variances

For testing changes in the variance for indicators of interest (such as journey time), we can use an F-test. The F-statistic is given by

$$F = \frac{S_A^2}{S_B^2},$$

with S_A^2 and S_B^2 being the sample variances. The statistical significance of the F-value can then be deduced from F-tables.

For testing any differences in means (e.g. mean journey time), a two-sample z- or t-test can be used. There is a need to test the assumptions on equality of variance first. The two-sample z-statistic is given by

$$z = \frac{(\bar{x}_A - \bar{x}_B)}{\sqrt{\frac{\sigma_A{}^2}{n_A} + \frac{\sigma_B{}^2}{n_B}}}.$$

The two-sample *t*-statistic is given by

$$t = \frac{\bar{x}_A - \bar{x}_B}{\sqrt{\hat{\sigma}^2\left(\frac{1}{n_A} + \frac{1}{n_B}\right)}},$$

with

$$\hat{\sigma}^2 = \frac{(n_B - 1)s_B{}^2 + (n_A - 1)s_A{}^2}{n_B + n_A - 2}.$$

The correct use of the F-test, the two-sample t and two-sample z test will of course depend on satisfying the standard statistical assumptions for these tests to apply.

Appendix B. Longitudinal data analysis techniques

There are numerous methods to analyse the sequential changes in individual travel behaviour. This appendix discusses in details a number of these methods:

- Similarity indices
- Tour-based analysis
- Sequence alignment method
- Hazard modelling and survival analysis
- Herfindahl Hirschmann Index (HHI)
- Location-based analysis
- Discrete choice modelling

Similarity indices

Schlich & Axhausen (2003) provided an overview of a number of similarity measures in an attempt to provide a procedure to identify similarities of activity/travel patterns over long periods. Similarity measures introduce a sequential component in travel behaviour research, which is lacking in regular statistical methods. Over time various researchers have introduced similarity indices in behavioural research (e.g. Kitamura & Hoorn 1987, Schönfelder & Axhausen 2002, Schlich & Axhausen 2003).

The measure of repetition (R) was first introduced by Huff & Hanson (1986). This measure indicates the extent of repetitiveness as the sum of deviations from a uniform distribution of behaviour. Trips are categorised according to their characteristics (i.e. mode choice, trip purpose, trip distance, trip arrival time and trip destination). People normally only perform a small number of these possible trip combinations, indicating a high repetitiveness. However, this measure is limited to comparing single trips. Introducing the number of trip per day as the unit of measurement led to the following similarity index:

$$Similarity = \left[1 - \frac{1}{2}\sum_{k}\left|P_{ik} - P_{jk}\right|\right]\frac{n_i}{n_j}, \quad n_i \geq n_j.$$

With: P_{ik}: share of trips in cell k of the contingency table at day i
i, j : index for the days to compare (with $i, j = 1, 2, \ldots, n$, if $i \neq j$)
n_i, n_j: number of trips on day i and j

A value of 1 for the index indicates identical travel patterns on the consecutive days. The disadvantages of this measure are that only two attributes can be taken into account. Moreover, the days are compared at an aggregate level instead of as a sequence of trips.

A slightly more advanced measure is the similarity index introduced by Pas (1983), comparing the trips of a day pairwise based on primary and secondary attributes. The primary attribute describes whether or not a trip is executed and the secondary attributes include purpose, mode, destination, etc.:

$$S_{ij} = \alpha n_{\min} + \frac{\beta\left|\sum_{t=1}^{n_{\min}}\sum_{k=1}^{n_{sa}}\omega_k m_{ijks}\right|}{\alpha n_{\max} + \beta n_{\min}}.$$

With: S_{ij}: similarity index for two daily activity patterns i and j
t: trip within a daily activity pattern (1, 2, . . . , n)
α: relative weight of primary attribute (trip n performed?)
β: relative weight of secondary attribute (0 – 1) and ($\alpha + \beta = 1$)
n_i: number of trips in daily activity pattern i
n_j: number of trips in daily activity pattern j
n_{max}: maximum number of daily activity pattern i and j
n_{min}: minimum number of daily activity pattern i and j
m_{ijks}: function for matching of 2 attributes; 1 in case of a match, otherwise 0
n_{sa}: number of secondary attributes
ω_k: weight of secondary attribute k in relation to other secondary attributes
where: $\sum_{k=1}^{n_{sa}} \omega_k = 1$

Depending on the important of the number of trips and the values of the secondary characteristics the parameters α, β and ω can be adjusted. This allows for flexibility in this approach.

Tour-based analysis
Stopher *et al.* (2010) used a tour-based analysis technique to investigate travel behaviour, based on previous work of among others Golob (1986) and Krizek (2003). Firstly, a list of tour type classifications is proposed. The classifications used by Stopher *et al.* (2010) and Stopher & Zhang (2011) are presented in Table B.1.

Table B.1 Tour type classifications

Tour type number	Tour description	Sequence*
1	Simple work tour	h – w – h
2	Simple education tour	h – e – h
3	Simple shopping tour	h – s – h
4	Simple other tour	h – o – h
5	Complex work tour (including composite and multi-part work tours)	h – (w/o) – (– w/o –) – (w/o) – h
6	Complex education tour (including composite and multi-part education tours)	h – (e/o) – (– e/o –) – (e/o) – h
7	Complex shopping tour (including composite and multi-part shopping tours)	h – (s/o) – (– s/o –) – (s/o) – h
8	Complex work and education tour	h – (w/e/o) – (– w/e/o –) – (w/e/o) – h
9	Complex education and shopping tour	h – (e/s/o) – (– e/s/o –) – (e/s/o) – h
10	Complex work and shopping tour	h – (w/s/o) – (– w/s/o –) – (w/s/o) – h
11	Complex work, education and shopping tour	h – (w/e/s/o) – (– w/e/s/o –) – (w/e/s/o) – h
12	Multi-part other tour	h – (o) – (– o –) – (o) – h

Note: h — home, w — work, e — education, s — shopping, o — other; (– x/x/x/x –) — (multiple) intermediate tours.

By classifying the trips of an individual into the groups as defined above and accounting for trip characteristics such as distance, departure time, travel time, modality and location, the travel behaviour can be captured and analysed. However, the larger the number of tour types the smaller the mass within the groups will be and the more difficult it will be to detect a behavioural change of an individual.

Sequence alignment method

The sequence alignment method (SAM) was introduced in transport research to conduct a systematic exploration of sequential dimension of human spatial and temporal activity. The fundamental concept of sequence alignment is to find the minimal number of operations to equalise two sequences and has its origin in the analysis of DNA sequences (Shoval & Isaacson 2007, Joh *et al.* 2002). The three elementary operations are: insertion, deletion and substitution (Shoval & Isaacson 2007). The more operations needed, the longer the 'distance' between the sequences. A sequence in the case of transport can be a list of activity locations or trips.

The SAM has the potential to find patterns in spatial behaviour and can be used to find similar purposes and common sequences that appear in a large database of spatial behaviour (Shoval and Isaacson, 2007, Joh et al., 2002). An extension to a multi-dimensional SAM offers additional possibilities to use SAM in transport research. Trips with their accompanying characteristics (i.e. the additional dimensions) can now be included in the analysis (Joh et al., 2001, Joh et al., 2002).

Hazard modelling and survival analysis

Both hazard modelling and survival analysis aim to describe the temporal component of travel behaviour. It is concerned with the probability of the occurrence of an event (i.e. making a trip with a specific purpose) within a specific time span ($t + \Delta t$), provided that the event was not observed by time t (Schönfelder, 2006). The occurrence of the event also depends on numerous other variables, such as personal or household characteristics and external factors. The methods enable a study of the intervals between separate occurrences such as trips with a specific purpose and modality.

Herfindahl Hirschmann Index

The Herfindahl Hirschmann Index (HHI) was first introduced by Herfindahl (1950) as an index to reflect the loyalty of an individual towards a certain brand. In the context of travel behaviour, it may be defined as the sum of squares of the market shares of each individual modality. As such it can range from 0 to 1 moving from equal distribution of all available modalities to a domination of one travel modality for the combination of location and purpose. The HHI is given by:

$$HHI = \sum_{i=1}^{n} \left(s_i^2 \right).$$

With: s_i: the modal share of modality I in the modal split
n: the number of modalities considered

In travel behaviour analysis it may reveal changes in the stability of choices, such as modality and departure time for specific trip purposes (Schönfelder, 2006).

Location-based analysis

This method uses the concept of continuous representation of space, based on the observed behaviour which can determine the locations likely to be visited and the part of a region that is used according to the needs and preference of an individual. Schönfelder (2006) described three methods of capturing human behaviour in space:

- confidence ellipses; represent the extent of the activity space. This method may be used to investigate the spatial dispersion of individuals. In human activity research the centre of the ellipse is assigned to the home location. It can also indicate change in spatial activity choice of one individual over time.

- kernel densities; represent the clustering of activities of an individual based on the amount of times a location is visited or activity duration. It converts locations of an individual over time into a histogram-like representation.

Discrete choice modelling

In general, panel survey data, similar to the travel behaviour data gathered through tripzoom, can be used to estimate choice models. These models are generally estimated using regression techniques. The advantage of such data and models is that behaviour of individuals and the causality in behaviour can be modelled (Frees, 2004). The models are generally based on

$$y_{ij} = \alpha_i + \beta_i x_{ij} + u,$$

but come in numerous varieties related to model assumptions. In the context of travel behaviour research, y could represent the weekly number of trips, kilometres travelled, time spent, total cost, total emissions, depending on x which represents for example personal or household characteristics.