Milestone 12 – Evaluation Framework

Version number:	1.0
Main author:	Bart Netten
Dissemination level:	Public
Lead contractor:	NMIE-R
Due date:	31/08/2017
Delivery date:	28/08/2017
Delivery date updated document:	

Co-financed by the European Union
Connecting Europe Facility

Grant Agreement No:
INEA/CEF/TRAN/M2015/1143833
Action No: 2015-EU-TM-0159-S

CONTROL SHEET

Version history			
Version	**Date**	**Main author**	**Summary of changes**
0.1	**30/12/2016**	**Bart Netten**	**First draft**
0.5	**22/01/2017**	**Bart Netten**	**Brussels meeting**
0.6	**28/03/2017**	**Bart Netten**	**Brussels meeting**
0.7	**17/08/2017**	**Bart Netten**	**Paris and Brussels meetings**
1.0	**28/08/2017**	**Bart Netten**	**Core Management Team, Advisory Committee, and General Assembly review & approval**
	Name		**Date**
Prepared	**Bart Netten**		**17/08/2017**
Reviewed	**Gary Crockford, Paul Wadsworth**		**25/08/2017**
Authorised	**Ronald Adams**		**30/08/2017**

Circulation	
Recipient	**Date of submission**
INEA	**31/08/2017**
InterCor consortium	**31/08/2017**

Authors (full list): Bart Netten, Isabel Wilmink, Paul Wadsworth, Eric Kenis.

Project Coordinator

Ronald Adams

Rijkswaterstaat

Office address: Toekanweg 7, 2035 LC, Haarlem (NL)

Postal address: Postbus 2232, 3500 GE, Utrecht (NL)

Mobile:+31 6 518 480 77

Email: ronald.adams@rws.nl

Legal Disclaimer

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

Terms and abbreviations

Term / Abbreviation	Definition
AC	Advisory Committee
AL	Activity Leader
ASR	Action Status Report
CMT	Core Management Team
EC	European Commission
GA	Grant Agreement
INEA	Innovation and Networks Executive Agency
IPR	Intellectual Property Right
CMT	Core Management Team
ML	Milestone Leader
MS	Member State
PC	Project Coordinator
TIC	Technical & Interoperability Coordinator

1 Executive summary

The objective of this document is to provide a framework for the InterCor project to coordinate the activities for pilot evaluation in Activity 4 and to align the activities with pilot preparation (Activity 2) and operation (Activity 3) from an evaluation perspective.

This evaluation framework defines "*what*" needs to be evaluated. It defines the artefacts for input and output for evaluation, such as the research questions and hypotheses to be tested and answered, and the required key performance indicators and measurements from pilot data. The evaluation framework also defines the structure for developing these artefacts, i.e. the logical order and responsibilities. Activities 2 and 3 provide the specifications of services, use cases and scenarios, pilot designs and test data. Activity 4 specifies the required input data for evaluation.

An important project objective for Activity 4 is to harmonize the evaluation activities across pilots such that similar services and systems are evaluated in the same manner and that evaluation results can be compared and aggregated. This requires the harmonisation of the artefacts and in a later stage also the evaluation methodologies, and sharing data and results. This report presents the structure and formats of the artefacts. The artefacts themselves are "living documents" on the InterCor web repository ProjectPlace, and attached in the annex.

The first year of the InterCor project focussed on the harmonisation of the Day-1 C-ITS services and ITS-G5 communication standards of the existing pilots in the Netherlands and France. The current evaluation framework has the same focus. To conclude a first iteration of evaluation, a first implementation of the evaluation framework has been trialled in the first InterCor TESTFEST performed in July 2017. This included the use of data log formats, data provisioning, data management and analyses.

In the next project phase, the evaluation framework will be refined and harmonised further with all pilot sites, and extended for new services and technologies (hybrid communication and security). The evaluation methodologies will be developed further in the respective evaluation subtasks for technical evaluation, impact assessment and user acceptance (see sections 9 - 13).

2 Introduction

The objective of this document is to provide a framework for the InterCor project to coordinate the activities for pilot evaluation and to align the activities with pilot preparation and operation from an evaluation perspective. The framework identifies the internal deliverables needed for developing and coordinating these activities. The framework also identifies the interdependencies of these deliverables, and assigns responsibilities of the deliverables to (sub) activities. The framework provides consensus on the deliverables, their scope and contents, and the responsibilities of activities to produce the deliverables.

The framework is based on the FESTA methodology for field testing and piloting [1]. Figure 1 shows the FESTA-V model of the subsequent steps to develop an implementation plan to prepare, operate and evaluate field tests.

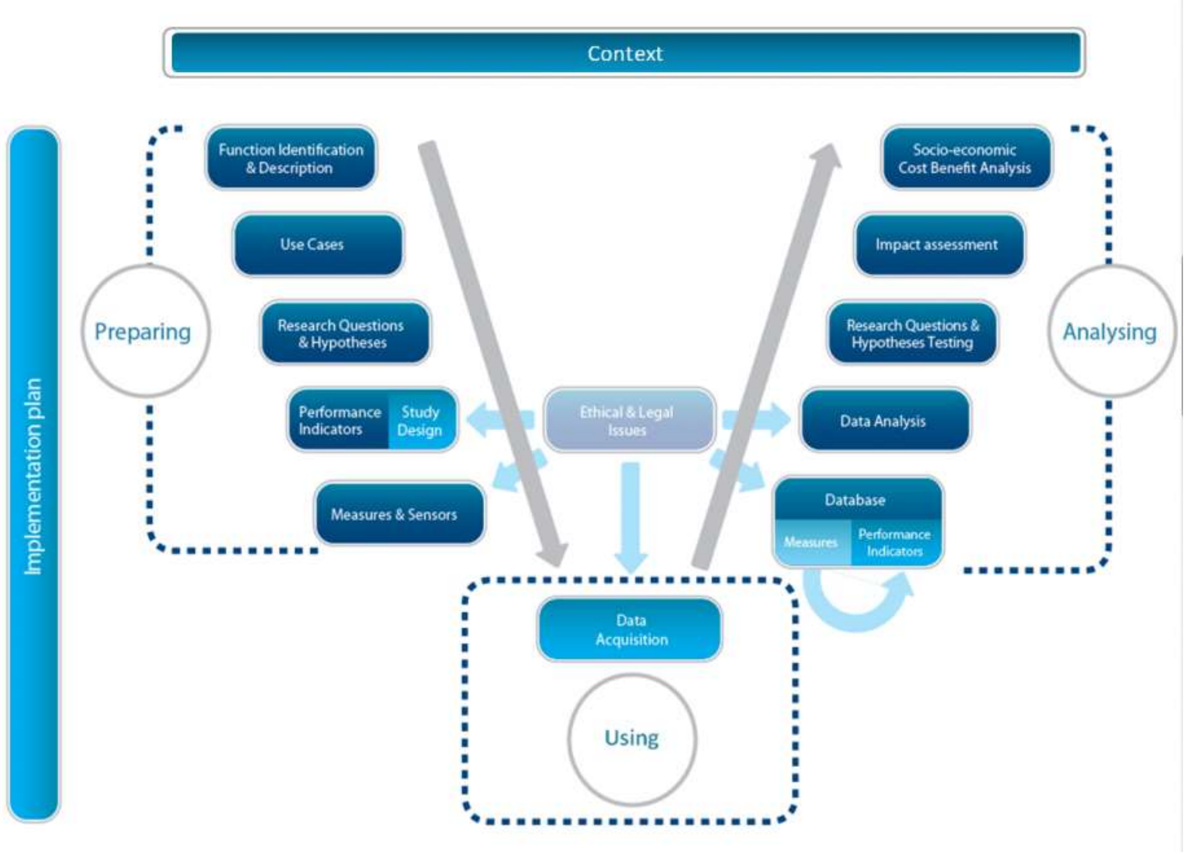

Figure 1: FESTA-V

The FESTA methodology can be broadly divided into following logical phases:

Preparing pilots:

Phase 1: Scoping the functionality to be developed, piloted and evaluated. This includes the formulation of the required functions of technologies and services to be developed, the use cases, situations and scenarios in which these functions will be piloted and evaluated.

Phase 2: Specification of research objectives, questions, hypotheses, indicators and measurements for piloting and evaluation.

Phase 3: Design of the pilot studies, including the formulation of test site infrastructure, test scenarios, test users, and test plans.

Operating pilots:

Phase 4: Development, procurement and roll-out of systems and services that have the specified measures and sensors to acquire the needed data.

Phase 5: Pilot operations executing the tests, including the collection, storage and management of data and the provisioning of data for analyses and evaluation.

Evaluating pilots:

Phase 5: Analyse the pilot data to test hypotheses and answer the research questions from phase 2.

Phase 6: Assess the impact on traffic, costs and society, based on the analysis results from phase 6.

FESTA is based on best-practices, attempts to be complete and cover all aspects that may become relevant at some time. Application and implementation of FESTA should be tuned to the objectives, needs and limitations in InterCor, using following guidelines:

- This document (and Activity 4.1) provides an evaluation framework, not the methodology for evaluation or piloting.
  - The evaluation **framework** defines "what" needs to be evaluated and provides the structure for developing and refining the artefacts for evaluation activities. **Artefacts** are the pieces of information needed as the input and output for developing evaluation methodologies and for executing evaluation activities.

Artefacts are for example the service descriptions, use cases and scenarios for piloting and testing, research questions, hypotheses and indicators. The framework defines common formats for these artefacts to enable the exchange and co-development with other activities and teams. The evaluation framework also structures the development and (iterative) refinements of artefacts in a logical order, and identifies the required input and output as dependencies for collaboration with other activities.

- The evaluation **methodology** defines "how" to evaluate, i.e. the principles, practices, guidelines, methods, tools, models, and analyses to implement the framework. The methodologies for technical evaluation will differ from those for impact assessment and user acceptance assessment for example. The methodologies for evaluating Day-1 C-ITS services may also need adaptation to also evaluate the hybrid communication and security solutions and the new services.

- This framework has a focus on the evaluation activities and elaborates on these phases. References are made to similar elaborations of the framework and methodologies for the other phases for preparation and piloting.
- Each step of the framework is covered in a separate section in this document.
- Many aspects are covered already in detail in the Project Plan, Grant Agreement [2], Description of Work, or activity Work Plans, and will only be referenced here. This applies in particular to:
  - Project objectives
  - Stakeholders, their collaboration and responsibilities
  - Project planning, including the planning of the pilots, preparations and evaluations
  - Project management and quality control.
- The framework gives an overview of the deliverables and does not contain the artefacts with contents for specific services, pilot sites, or evaluations. Such artefacts are managed by the activities as living documents on the InterCor web repository ProjectPlace and/or on other repositories. This framework document gives examples of the artefacts to be developed.

- Road Works Warning is one of the most deployed and well-known Day-1 C-ITS service and used as an example throughout this document. The examples are given as "*snippets*" from existing documents.

The current version is written from the perspective of the evaluation framework, and identifies the input needed for evaluation from other activities, and the output that evaluation will provide to other activities. The evaluation framework strongly depends on the piloting framework and the pilot preparations. The objective of this document is therefor to provide a starting point for other activities to complete and adapt the framework, and iteratively come to a project consensus. This includes further development of the methodologies in Technical, Impact and User Acceptance Evaluations.

2.1 Project Objectives

In the InterCor project, various Cooperative Intelligent Transport Systems (C-ITS) are being tested. InterCor builds on the various C-ITS corridor projects carried out in several EU member states. Member States involved in the InterCor project are: Netherlands, Belgium, France and the United Kingdom.

The InterCor project comprises a series of interlinked sub projects which together have common objectives, summarised as follows from the Grant Agreement [2]:

1. Demonstrating large-scale interoperable deployment of C-ITS,
2. Demonstrating cross border interoperability.
3. Providing C-ITS services on a broader scale by hybrid communication.
4. Extending the strategic cooperation between C-ITS front running countries and assisting other Member states to step-in;
5. Evaluating the life benefits of C-ITS applications by reports on technical evaluation, impact assessment and user acceptance.

These objectives provide the focus in the FESTA-V (Figure 1) for pilot preparation, pilot design and evaluation, from top level to the formulation of scenarios, research questions, hypotheses and pilot design.

2.2 The InterCor project setup

The project organisation is set up according to the Grant Agreement [2]. Figure 2 shows the set-up of activities.

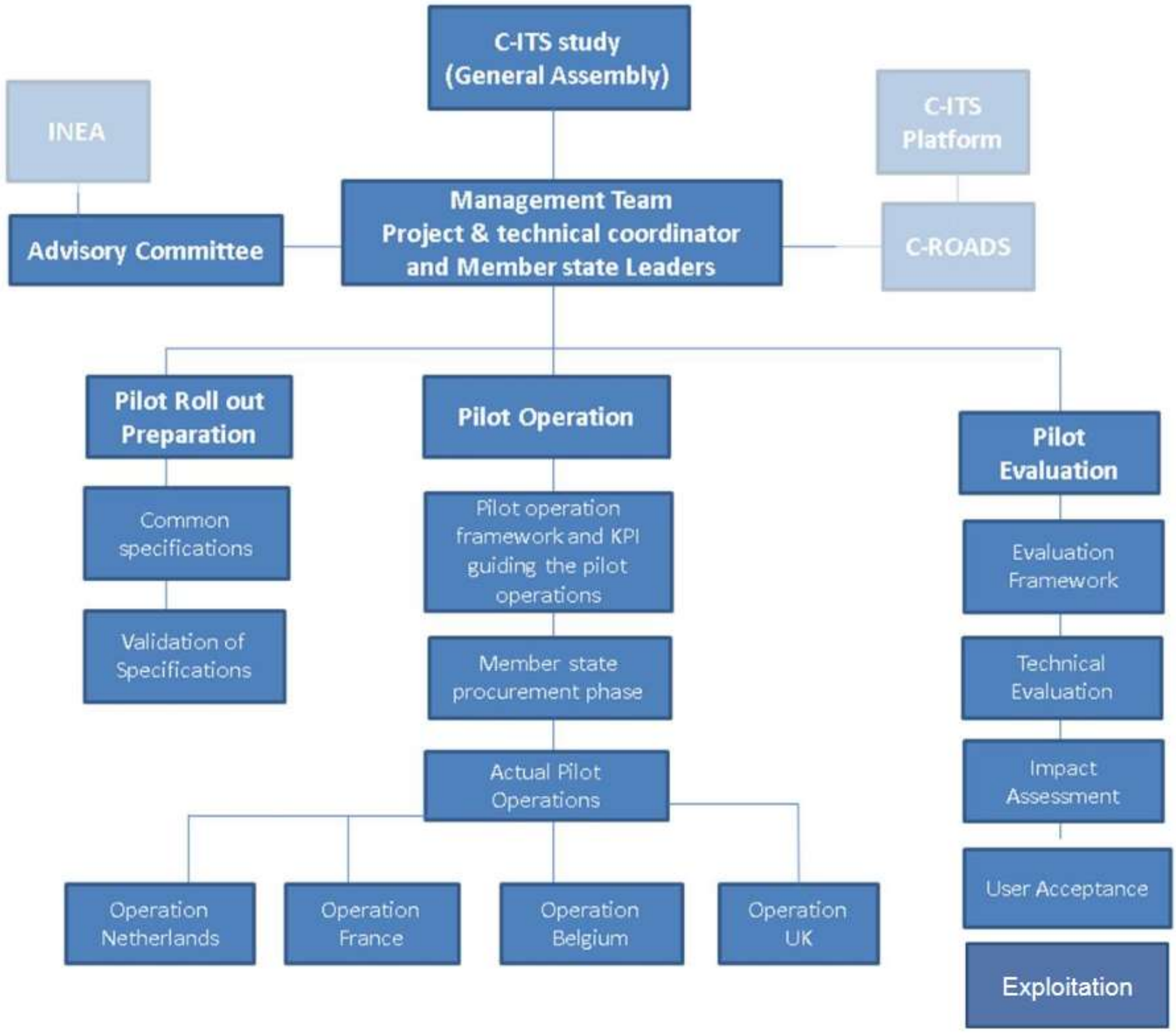

Figure 2: Project organisation

The project organisation is logically related to the steps and phases in the FESTA V-model (Figure 1). Although most steps will involve the cooperation of multiple activities, the primary responsibility can be allocated to a single activity or sub activity:

Activity 2 – Pilot roll out preparation – is mainly responsible for:

- Definition of Function Identification & Descriptions (Activity 2.1 a-d)
- Definition of Use Cases (Activity 2.1.a-d)
- Formulation of Research Questions & Hypotheses from the perspectives of the technologies, specifications, interoperability and services (Activity 2.1 a-d)

Activity 3 – Pilot Operation – is mainly responsible for:

- Pilot design, test scenarios and plans as part of the study design (Activity 3.1)
- Procurement of equipment that provide the sensors and measures (Activity 3.2)
- Data acquisition and management during piloting (Activity 3.3)

Activity 4 – Pilot Evaluation – is mainly responsible for:

- Data management for evaluation
- Data analysis
- Answering research questions and hypothesis testing
- Technical evaluation, impact assessment, and user acceptance evaluation
- Reporting of data analysis and evaluation results as feedback to pilot site operation and for gap analysis

There are also steps that cannot be clearly allocated to a specific activity. The main interaction steps between the InterCor activities, with shared responsibilities are:

- Situations and scenarios for piloting use cases
- Formulation of research questions and hypotheses for piloting and evaluation.
- Formulation of key performance indicators and data to be collected during piloting

The responsibilities and collaborations will be identified in each of the following sections. Figure 3 gives an overview and logical ordering following from Figure 1 and Figure 2.

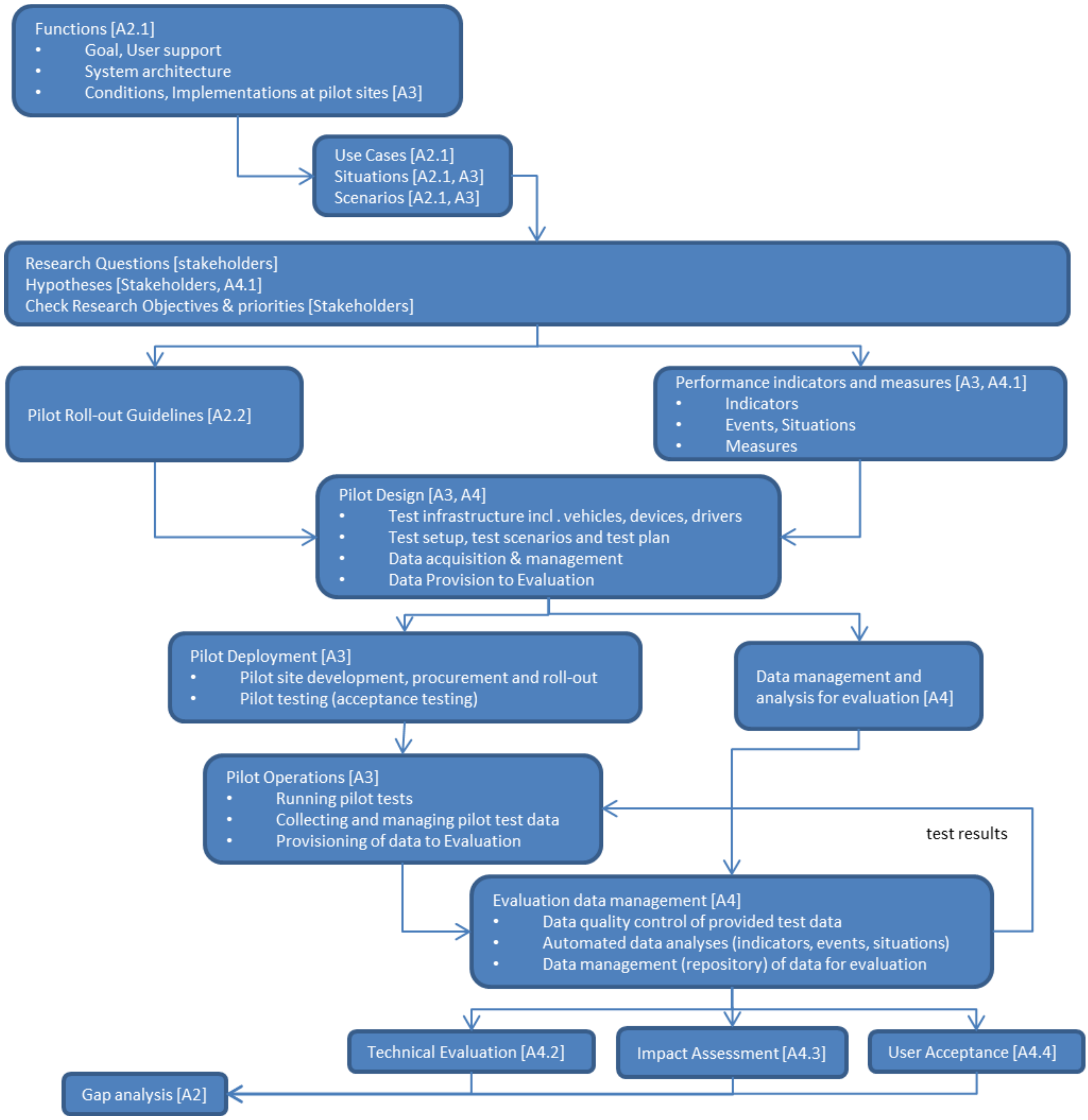

Figure 3: Flow of artefacts, dependencies and primary responsible activities [between brackets]

2.3 *Piloting framework*

The piloting framework will cover the pilot roll-out guidelines, design, deployment and operations. The piloting framework will be developed in InterCor Activity 3.

2.4 *Pilot Plan*

InterCor plans to roll out the following services on the following pilot sites (Table 1). The services are specified in Activity 2.1.d and described in more detail in section 3.

Table 1: Services per pilot site

Services	the Netherlands	France	Belgium	UK
In vehicle signage	x	x	(x)	x
Probe data	x	x		x
Road works warning	x	x	x	x
GLOSA	x	x	x	x
Multimodal cargo transport optimisation	x	x		x (freight slots for ports)
Truck Parking	x	x		x
Tunnel logistics	x		x	x

The project implements an iterative approach to piloting. The project will go through several cycles (Figure 4), starting with the piloting and evaluation of relatively established Day-1 C-ITS services in the first year in France and the Netherlands. These services will be enhanced and new services will be introduced and enhanced in subsequent cycles over the next two years. The pilot sites may plan intermediate updates and adaptations to tests. The iterative approach applies to all three project activities, even though not all sub activities may have to incorporate all iterations at every pilot site.

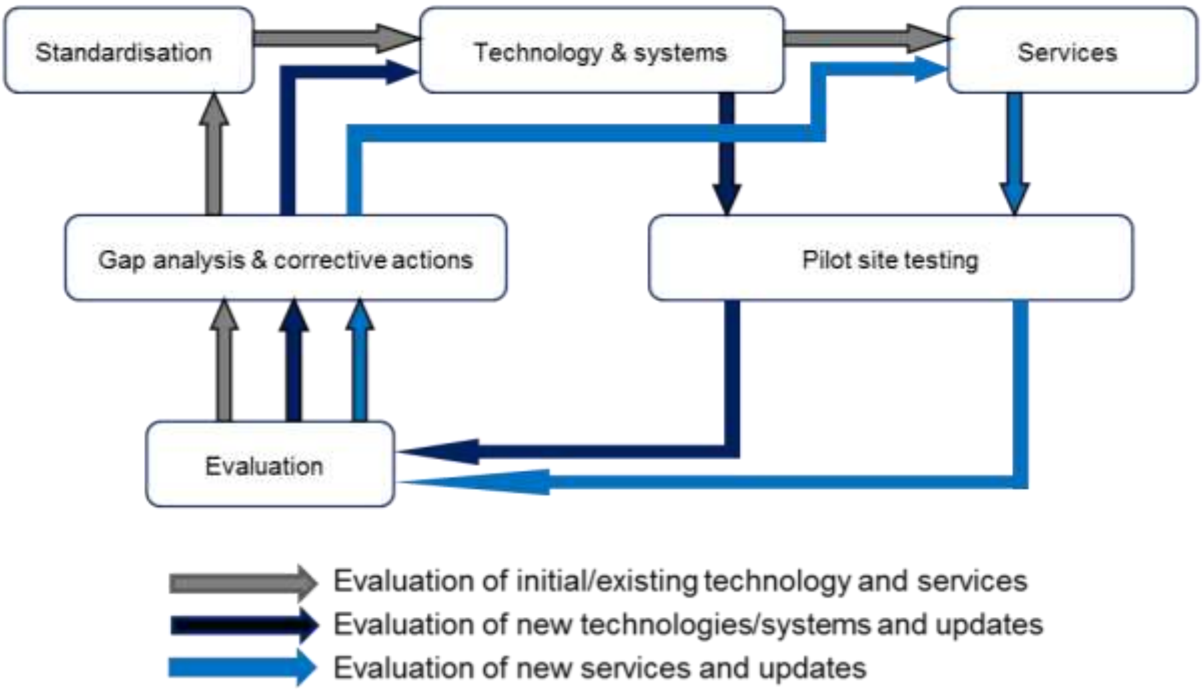

Figure 4: Iterative approach to piloting

2.5 Ethical and legal issues

Section 6 of the FESTA handbook [1] gives an overview of legal and ethical issues to be considered for piloting. Issues like participant selection, data protection, safety, approval for on-road use, etc. are primarily the responsibility for pilot operations in Activity 3.

Relevant to evaluation is the processing of private data and how privacy issues are resolved. It is assumed that any privacy sensitive information will already be removed or anonymized by the pilot sites before providing data to Activity 4 for evaluation. Note that the evaluation does not intend to give statistics for user ID's over events and tests.

Potentially privacy regulations may pose issues in the distribution of pilot data between public, research and private partners and between member states.

2.6 Evaluation Framework

The objective of the evaluation framework is to harmonize the evaluation activities such that similar services and systems tested under different situations and pilot sites are evaluated in the same manner. This requires a coherent distribution of evaluation activities and tooling, and sharing of data and results among the activities. The evaluation framework defines the structure for evaluation activities and tools within Activity 4, and also the required input and output and the processes to develop these in collaboration with other activities.

The second objective of the evaluation framework is to align the evaluation activities with the other activities, i.e. to align the evaluation activities with the piloting framework, pilot operations and to align the hypotheses and research questions for evaluation with the project and research objectives of the Member States and other stakeholders.

The objectives for piloting, and the services to be piloted, are the essential input for evaluation. Hence, the first sections of this report present the essential input and objectives for piloting, the service functions, Use Cases, and scenarios for piloting.

The plan for the first year of the InterCor project is the harmonisation of the Day-1 C-ITS services and ITS-G5 communication standards of the existing pilots in the Netherlands and France. Hence the implementation of the InterCor evaluation framework is also limited to these services and technologies. The evaluation methodologies for these services are being developed in the Netherlands and France. The current InterCor evaluation framework is a first step in the harmonisation of the evaluation frameworks.

In the second project year, new services and technologies (hybrid communication and security) will be specified. Harmonisation of the evaluation frameworks will continue and be adapted and extended for the new services and technologies.

3 Functions

Before any of the following artefacts for pilot preparation, operation and evaluation can be defined, the functions that are to be provided by services should be identified (i.e. Table 1) and described. In a later phase, these functions will be specified and defined in more detail, for example in use cases (next section).

In Figure 1 this first step is called “Function identification and description”. A function description of the services-to-be-tested should be extended and also include the technologies used for example for InterCor purposes. Function descriptions are part of the service specifications from Activity 2.1. The technology descriptions for services are a joint responsibility of Activity 2.1.a-d. Pilot operations (Activity 3) also have a responsibility to complement the service descriptions with product and service implementation details resulting from their pilot test setup and procurements.

Development and description of a new service initiates with a high level description of the function to create a basic and common understanding. The function description will be updated during the procurement process with descriptions of the functionality of products and services. The function or service can also be described as part of the system concept and use case definitions. Examples are given in the C-ITS Corridor system concept [3] and the Dutch Profiles for RWW [4], PVD [5], and GLOSA [6], and for SCOOP@F in [7]. A high level service description for In-Vehicle Signage harmonised in InterCor is available in [8].

At some point during preparation, the technologies, systems and products have to be identified and described for services. In InterCor, especially the communication and security solutions should be described for the services as well as for defining research questions and hypotheses for evaluation of these technologies.

Services may be applicable to specific users, locations, technologies or traffic conditions. The following subsections show the function descriptions needed for the services.

3.1 Goal

The goal or purpose of the function or service should be clear and provides the starting point for defining objectives and research questions to be tested. Typically, the purpose of a function is to provide support to users in the form of specific information or warnings.

> *The main aim of Road Works Warning RWW is to improve road safety. Construction sites and temporary maintenance working areas are accident black spots, because static traffic signs are ignored or realized too late. RWW aims at reducing the number of collisions with safety-objects near road works. RWW will alert the road user when*

approaching a dangerous road works zone and will simultaneously provide information on the changes in the road layout.

3.2 User support

The support provided to users should be described in more detail:

- Identify the users that receive the support.
- Identify the stakeholders and actors that are relevant to generate or provide the support.
- What information or support is provided?
- When is specific information provided, for example in specific traffic situations or locations? What are the trigger conditions to initiate or update information, and when is the information revoked?
- In what form is the information provided and presented to the user, for example as a hazard warning, or a time, distance, speed, lane, route or parking advice?
- If different devices or services have different ways of presenting information and warnings to the driver, please specify the various forms of presentation.

As an example from [4], I2V communication enhances the operational integration of local traffic management and in-car systems to improve safety, traffic efficiency and helps to protect the environment. A Road Works Warning message is sent by a roadside unit (or road works trailer) to approaching vehicles. Road Works Information is a related application used by road operators to inform road users – via service providers - on planned and actual road works for pre-trip and on-trip navigation. Three types of road works types are identified:

- *Short term mobile road works*
- *Short term stationary road works*
- *Long term stationary road works*

Figure 5 shows how, where and when the driver should be informed on the road works

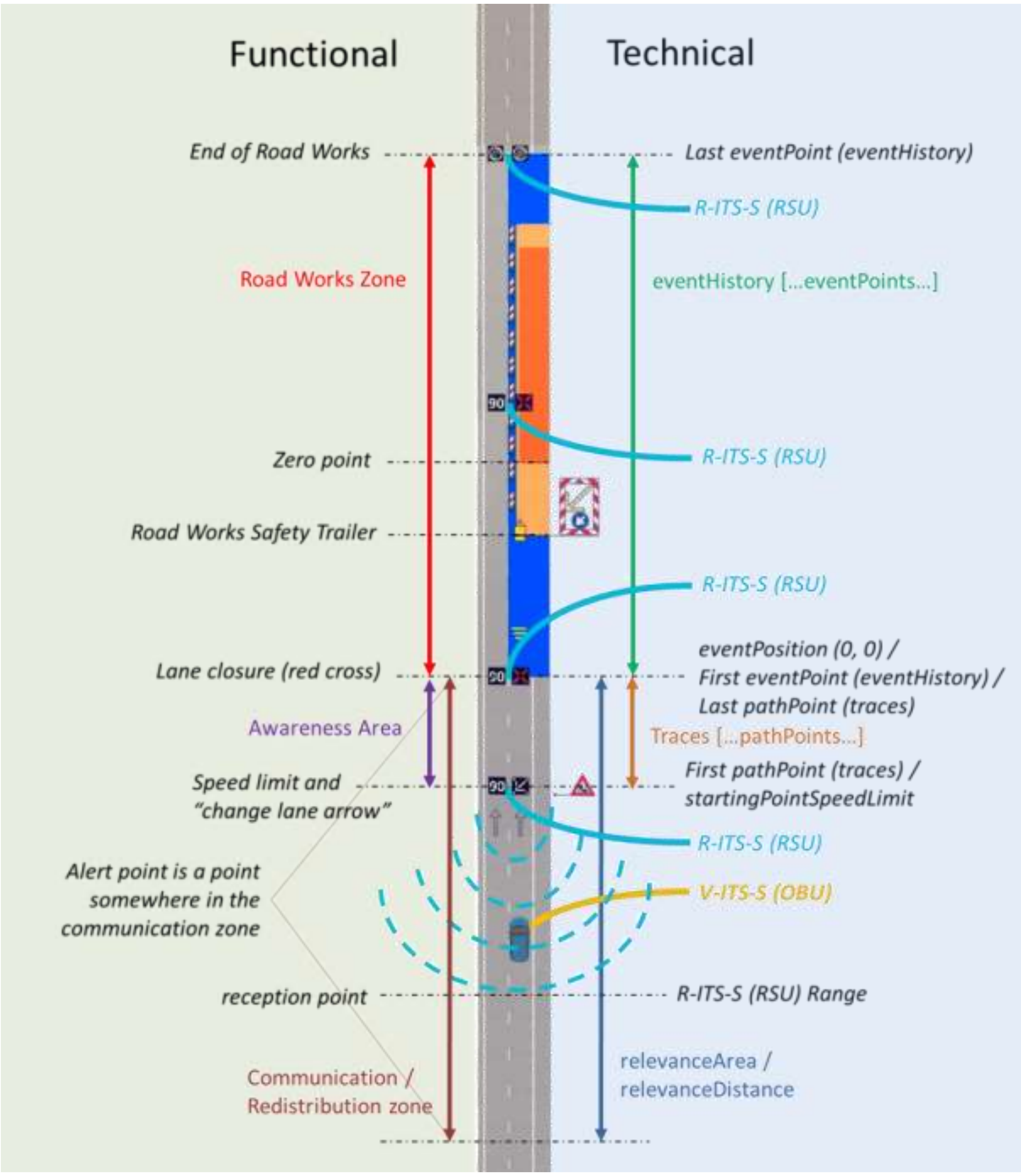

Figure 5: Road works scenario from [4]

3.3 System Architecture

What are the main components of the service to provide the function, what technologies are used, and what are the system/component interactions? The system architecture is described at functional level and at the technology level. It describes for instance the following:

- input data(bases) used for the service

- data flows and processing steps
- communication media and security
- the type of HMIs used

Usually, system diagrams are provided to visualise the functional and technical system architectures. An example for RWW can be found in [3]. Typically, a service could be deployed in one or more alternative system architectures, identifying alternative stakeholder organisations or technologies.

3.4 Conditions under which the services do (not) operate

Describe the conditions when the services do or do not operate, and when the function should or should not be provided. The study design (design of the tests, see step 7) should reflect this.

> *As a trivial example, a RWW only works when there are road works. On the other hand, keeping planned road works and road user information in sync with the actual road works in the field may not be so trivial. Some service may only work in specific regions (e.g. because input data are needed that are only available in some places). Some services don't work in low visibility conditions or weather conditions.*

3.5 Implementations at pilot sites

Implementations of the functionality may differ per pilot site. This has a large impact on the scenarios for piloting and hypotheses for evaluation and it is important to focus in an early stage. This input is provided by Activity 3.

> *Implementation choices may differ per pilot site for example for RWW on:*
>
> - *Type of road works (mobile, short term...)*
> - *Data and service providers*
> - *Laws and regulations for setting up road works (trigger conditions, configuration, …)*

4 Use cases, situations and scenarios

Use cases, situations and scenarios define how functions or services are intended to be deployed, tested and used, and are a refinement of the service descriptions in section 3. A use case defines the intended usage of the service to provide the function in typical situations and scenarios, and to define specific situational and trigger conditions. It should consider identified user needs (or preferences supporting an effective uptake) and requirements from both a system and user perspective.

Use cases, situations and scenarios are defined per function or services, and may also be defined for groups of related services.

> *Examples of use cases, situations and scenarios can be found in [4], [5] and [6] for RWW, PVD and GLOSA resp. A comparison of these Use Cases in InterCor is made in [9].*

Definition of these artefacts is a joined effort for activities 2 and 3 to match the specifications of technologies and services with the deployment and test plans of pilot sites. This is a bidirectional and iterative process:

- Use cases, situations and scenarios provide a concrete framework for the specification of technologies and services, including trigger conditions, profiles and deployment guidelines.
- Scenarios and deployment guidelines provide the reference framework for deployment of technology and services and for planning and setting up pilot test and evaluations.
- Practical situations and regulations at pilot sites also restrict possibilities for deployment and test scenarios, and may give rise to adaptations to specifications, profiles, or deployment guidelines.

Definitions of use cases, situations and scenarios at an early stage in the project are also important to identify (large) differences between pilot sites. The aim is to have common use cases, situations and scenarios at pilot sites, to make it possible to evaluate interoperability across pilot sites and borders. There is a link with the research questions and hypotheses (next step): if use cases or scenarios are not specific or downright vague, it is difficult to define concrete research questions or to agree on relevant research questions, and even more difficult to set hypotheses for testing and evaluation.

In practice, it is often difficult to separate descriptions of use cases, situations and scenarios.

- A **use case** describes the usage of the system that is aimed for, the assumed behaviour of the system, and the assumed behaviour of actors involved (e.g. the end user).
- A **situation** is a state of the environment or system that may influence the behaviour of the service.
- A **scenario** is made up of a use case and a situation. Scenarios are important when targeted testing is carried out, if it is important that use cases can be tested in specific situations. In naturalistic tests, it can be relevant to define 'scenarios' for selecting the relevant data from a large dataset.

4.1 Use cases

A use case describes a typical usage of the system, identifying the stakeholders and actors of the service, and describing the assumed behaviour of the system and the assumed behaviour of actors involved (e.g. the end user). The use case identifies the system architecture and configuration used (section 3.3), including the functional components, deployment model and technologies.

Different use cases are defined to identify alternative configurations of a service. Different use cases could be defined for example to differentiate deployments using ITS-G5 communication only, or hybrid communication.

Different use cases are also defined to identify alternative deployments and information flows, such as the mobile and short term RWW configurations.

Different use cases are also defined for alternative types of testing, such as specific controlled tests with test drivers on a test field or public road, or naturalistic driving tests with normal drivers on their daily commute. In the first case, a use case can be defined and tested precisely, where as in the latter case, events cannot be set up and need to be detected afterwards from the log data.

4.2 Situations

A situation is a state of the environment or system that is of relevance for the testing because it may influence the behaviour of the service. This paragraph describes the relevant external situations such as traffic conditions, traffic composition, traffic rules, weather, road and lane configuration, or the type of manoeuvre that may affect the behaviour and other users such as the traffic operator or services.

This paragraph should also describe external situations that are relevant for system and communication behaviour, such as the presence of obstacles for communication, penetration rates of vehicles equipped with communication systems and antenna mounting, spacing and transmission power of road side units, and areas with variable UMTS of LTE coverage.

Situations should also be identified to distinguish the behaviour of communication security solutions, such as the mix of multiple certificate authorities and certificate switching.

Situations differentiate the behaviour of a service or user in particular use case, and therefore differentiate the effect that can be expected. It is important to define situations accurately to distinguish situations in which an expected effect be expected and when not? Which situations should be included, excluded, or ignored in the piloting? Do situations occur frequently enough naturally, or do they have to be engineered?

Situations to be distinguished for testing road works could be:

- *Free flow and congested traffic to differentiate the freedom of choice for drivers to comply or disobey speed restrictions.*
- *Lane configurations to differentiate the driver compliance when the hard shoulder should be used as an extra lane or when lanes are narrowed.*

4.3 Scenarios

A scenario is made up of a use case and a situation. The combination of use cases and related situations provides an exhaustive set of test scenarios that pilot site could choose from. Vice versa, the capabilities of pilot sites limit the possible set of test scenarios. The sanity check of scenarios to be tested per pilot site may strongly limit the research questions and hypotheses that can be tested.

5 Research questions and hypotheses

Research questions are the stakeholders‘ questions that need to be answered in the InterCor project. Research questions express the a priori *uncertainty* of stakeholders, for example about the usage and benefits, impact on policies and legislation, whether and how to deploy new technologies or services, where is the business case? Getting answers to these research questions is the main reason for stakeholders to set up and participate in the InterCor project.

The next step is to formulate how a research question can or should best be answered. This is formulated in one or more hypotheses per question. A hypothesis defines *how to test a possible solution* to the research question. A hypothesis is stated as a proposition, possible explanation or tentative answer to the research question. The hypothesis can be tested, and the outcome of the test quantifies the measure of success, improvement or feasibility of the solution.

It is generally recommended to limit the number of research questions. Research questions could easily explode in the amount of work needed for testing and evaluation, as the total number of analyses for (scenarios * hypotheses * indicators * measurements * evaluation criteria * sample size) multiplies rapidly. Probably it will not be feasible to test all questions and hypotheses that are generated initially. These will have to be prioritised at some point. The project objectives will be used as a control-and-check-mechanism to match and prioritize research question.

Research questions and hypotheses are collected and organised in a spreadsheet as a living document on the InterCor web repository ProjectPlace (see Annex - Evaluation Artefacts). Research questions and hypotheses will be harmonised across pilot sites. Research questions and hypotheses originating from a specific pilot site, stakeholder, may be added. However, the research questions and hypotheses have to be valid for each pilot site, and should be agreed upon by all stakeholders for consistent evaluation later on.

5.1 Research questions

Research questions are the questions of stakeholders about uncertainties that need to be piloted and evaluated. Research questions should cover the project objectives (section 2.1), or extend the scope for piloting. Given the project objectives, the core research questions should focus on deployment of systems and services from the perspectives of road operators and service providers, the common set of specifications and on the cross-border interoperability. To deploy systems and services, research questions should also be raised about other practical and necessary conditions for success, such as the technical feasibility, impact and potential for scaling, costs and benefits.

Formulation of research questions primarily involves activities 2 and 3 to provide input from stakeholders' perspectives, for example:

- Deployment, operations & asset management, traffic rules and guidelines, and high level policy goals perspectives (Activity 3)
- Technical perspective on specifications of services technologies, functioning and performance of the systems (Activity 2.1.a-c)
- Technical perspective on interoperability and gap analyses on standards and profiles (Activity 2.1.a)
- Functioning of the service to the end users and their acceptance of the services or applications (Activity 2.1.d)
- Changes in behaviour of the end user (Activity 2.1.d)
- Impact of the services on traffic efficiency, safety, mobility, environment

5.1.1 Category

Stakeholders may refine the project objectives or define additional research questions. Most likely, there is significant overlap in the research questions from various stakeholders. Therefore, a differentiation of research questions is proposed in the following categories:

Technical research questions, such as:

- ITS-G5 communication (Activity 2.1.a)
- Hybrid communication (Activity 2.1.b)
- Security (Activity 2.1.c)
- Application related questions, such as the performance (accuracy, timeliness) of an end-user application

Service related research questions (Activity 2.1.d)

- Generic questions that are relevant to any services, e.g. whether a service is continued across borders, or policy, legal or privacy issues
- Service specific questions

Impact related research questions (Activity 3)

- Traffic safety

- Traffic efficiency
- Environment related questions on emissions, fuel consumption, etc.
- Mobility related questions, e.g. the impact on travel patterns
- Cost/Benefit

User related research questions (Activity 2.1.d)

- Usage of services and applications
- Behaviour and behavioural adaptations of users
- Acceptance of services and applications

These categories are chosen to guide the formulation of the research questions towards evaluation, for example to refine the questions into types of hypotheses and indicators. The categories are defined in Annex - Evaluation Artefacts, sheet “CATEGORY”.

5.1.2 Prioritisation

Research questions are prioritised by their relevance to project and research objectives and by the estimated feasibility to test and answer the research questions in the planned pilots. Each pilot site may give a priority to a research questions, e.g. as a percentage and a motivation.

The priorities for answering the research questions will ripple down to the prioritisation for pilot tests and the selection of measurements and evaluation analyses.

5.2 Hypotheses

For each research questions, one or more hypotheses are formulated to express how the question could be tested and answered. The purpose for defining hypotheses a priori is to set a feasible scope of testing: what needs to be tested to provide a solid answer to the question and what can be tested in the pilot. If the full answer cannot be provided anyway, then the research question and hypotheses should be reformulated. This avoids much unnecessary work in setting up pilots, logging and data collection.

The process of formulating hypotheses translates the general research questions into more specific and statistically testable hypotheses. A hypothesis is stated as a proposition or likely outcome of the research question. In [1] a hypothesis is defined as:

> *"a specific statement linking a cause to an effect and based on a mechanism linking the two. It is applied to one or more functions and can be tested with statistical means by analysing specific performance indicators in specific scenarios. A hypothesis is expected to predict the direction of the expected change."*

Testing the hypothesis results in the answer like true/false, a measure of success, the improvement or feasibility of the solution. Hypotheses for example can be defined for:

- Expectations of possible solutions and their performance or benefit, and how these can be measured or differentiated.
- Propositions about foreseeable issues for interoperability or performance degradations.
- Specific situations that may affect the service, technology or relevance to the user
- Effect of scaling, e.g. by penetration rate, density/coverage of road side equipment

Research questions may already be specific, or even formulated as a direct measure such as "what is the communication performance of Road side ITS-Stations". The corresponding hypotheses can then be formulated as statements for requirements for Road Side Unit deployment guidelines such as "Road side ITS-Stations are mounted and configured such that Vehicle ITS-Stations within a range of 500m in any travel direction receive at least 80% of the I2V messages".

Other research questions may be more generic or open, in which case alternative hypotheses can be formulated for alternative interpretations of the question for specific services, situations or conditions. Research questions may also be interpreted from different disciplines by formulating hypotheses from technical, impact, or usage perspectives for example.

Research questions can also be explorative in nature, with the objective to collect lessons learned from pilots, for example "What are the limitations for deploying / developing a service for ITS-G5, cellular communication or hybrid communication?" Such a question can only be re-iterated as a hypothesis to collect the experiences from testing, evaluation or user questionnaires.

Hypotheses are defined in cooperation with activities 2, 3 and 4. Activities 2 and 3 should express the relevant hypotheses and set the thresholds for accepting the outcome as an answer to their research question. Activity 4 should verify the testability of the hypotheses; e.g. whether sufficient data can be gathered to derive the answers.

6 Performance indicators and measures

The next step is to specify how the hypotheses can be measured in tests. This requires the matching of the possibilities of pilot sites on “what can be measured and tested” with “what is intended and needed” for the evaluation of hypotheses. If the data required for testing hypotheses cannot be provided, then the research objectives and questions needed to be adapted or the priority degraded.

Specification of how hypotheses can be measured consists of 3 steps:

1. Formulation of the performance indicators that quantify or qualify hypotheses
2. Formulation of the events and situations to differentiate test results
3. Measurements needed to calculate performance indicators and to detect events and situations.

6.1 Performance Indicators

The first step is to define indicators that measure the performance with respect to the hypotheses. A definition of an indicator is interpreted from [1]:

> *“A quantitative or qualitative parameter, derived from one or several measures, expressed as a percentage, index, rate or other value, that is monitored and can be compared to one or more evaluation criteria of a hypothesis”*

An indicator is defined as a relative measure with a denominator. The denominator makes the indicator comparable, for example per time interval, distance, location, or user group.

The hypothesis should have an evaluation criterion defined to accept or reject it eventually. This criterion is a threshold for the indicator.

Preferably, indicators are defined as quantitative parameters that can be measured or calculated from collected data. Indicators can also be defined as surrogate measures from literature, sensitivity analysis on the collected data, or as differences between reference and controlled tests. Alternatively, qualitative parameters can be defined that can be acquired for example from expert’s opinions, or user questionnaires.

The following is a list of examples of common indicators for the categories of section 5.1.1. The list is derived from [10] and C-ITS projects. The current list is defined in Annex - Evaluation Artefacts.

Technical / performance

- accuracy of warning distance and time
- end-to-end delay

- latency per processing step
- communication range
- detection range

Driving behaviour

- awareness of a speed limit or other sign
- average speed or change in speed
- time spent decelerating "harder" (harsh braking?)
- spot speed (in relation to sign/advice/stop line / traffic jam)
- amount of sudden manoeuvres
- number of times a secondary task was aborted (per unit of time / distance travelled?) OR becoming alerted OR level of alertness
- reaction time
- compliance with advice

Safety

- Injuries
- Fatalities
- Various surrogate safety measures

Mobility

- journey quality
- user uncertainty
- feeling of safety and comfort
- user stress
- journey duration
- travel pattern, e.g. number of trips
- share of intersections where driver has to stop for red light

Traffic efficiency

- travel time
- travel time variability

- delay
- average speed / variation in speed

Environment

- CO2
- NOx
- PM
- HC
- CO
- fuel / energy consumption

User acceptance

- usefulness
- ease of use
- reliability of information
- willingness to purchase
- reported change in behaviour
- alertness
- change in speed
- compliance

Scaling up / CBA

- Benefit-cost ratio
- Net present value
- Break-even point

6.2 Events

An event is defined from an evaluation perspective as a specific type of action from a user, driver, system or component. An event can also be defined as an activity during an interval of time. Events are specific for a service, a technology or a situation.

The objective for defining and detecting events in pilot data is to enable objective comparison of similar events, and reduce foreseeable bias in the evaluations. Indicators are calculated and evaluated per event. To detect and distinguish events during evaluation, events need to be logged or defined from measures (section 6.4).

Events can be defined in alternative ways from different perspectives, for example:

- From a functional perspective, an event is a period in which a service is active or relevant to the user. Such events can be derived from the different phases in the use cases in section 4.1.
- From a technical perspective, an event can be defined as a specific state, action or activity of the service, application, component or from communication.
- From a safety impact perspective, an event can be defined as the period of an unsafe situation or (near) crash, and the user is warned (or not).
- From a traffic efficiency perspective, an event may be defined as a period of heavy or congested and the user is informed (or not).
- From a user perspective, an event can be defined as the period in which information or warnings are presented and updated.

Alternative events can be defined from different perspectives and from alternative implementations of services, systems and components. It is important to explicitly relate the events in order to harmonize evaluations and to correlate events and performances from these different evaluation tasks to provide consistent answers to the research questions.

Figure 6 gives an example of an event for a Road Works Warning (RWW) and how this event can be defined from various perspectives:

- *From a driver behaviour perspective, the event is defined as the period during which a warning is presented on the HMI to the driver. The event may be refined for updates of the presented information and increasing warning levels*
- *From an application perspective, the event occurs while the vehicle position can be matched to the road works area. The event may be refined for matching to the trace, event history, detection or relevance zones defined in the I2V messages*
- *From a communication performance perspective, the event may be defined as the period that the Vehicle ITS-Station received specific I2V messages. Events could also be defined as actions for each individual update of an I2V message from an individual Road side ITS-Station, or even for every broadcasted repetition.*

- *From a traffic management perspective, the event can also be defined as the period in which the Traffic Control Centre or Central ITS-Station issues a RWW for the duration of the road works*

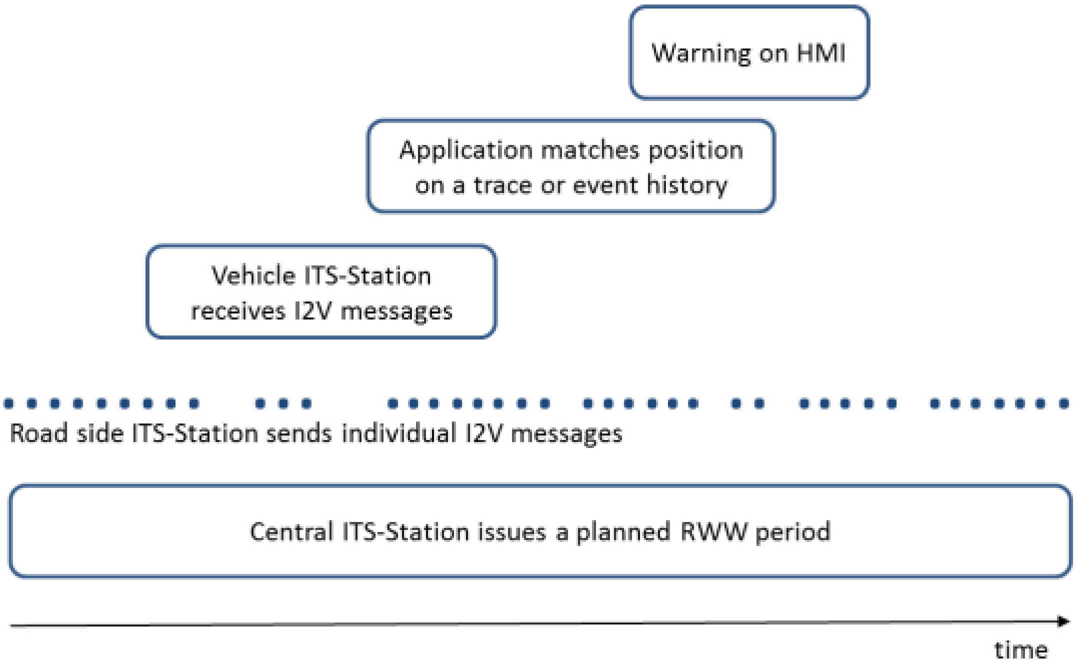

Figure 6: Example of events from various perspectives on a timeline

Various events may occur simultaneously causing interactions that may affect performance and the indicators. In the above example, DENM and IVI messages can be used for the same RWW, and consequently DENM and IVI events are generated. The periods of isolated events and combinations of events may have to be distinguished, for example for communication performance and driver behaviour.

From the perspective of the driver, a RWW service may interact with an IVS service and three types of events should be distinguished:

- *Driver receives only a RWW message on the display*
- *Driver receives only a IVS message on the display*
- *Driver receives both a RWW and IVS message on the display*

The driver may also experience a sequence of events if a RWW and IVS event partially overlap; e.g. first a RWW, then a RWW and IVS, and when the RWW ends only a IVS event.

6.2.1 Event

Events are defined by the following features:

- Trigger conditions that start an event
- Trigger conditions that end an event
- Characteristic measures of the event that distinguish the event from other events or combinations of services and events

6.2.2 Event model

An event model defines the set of events that belong together for a single evaluation perspective.

For technical evaluation of DENM applications for example, the following event models could be defined from perspectives of communication performance, application logic, and driver behaviour:

- *Generation of a DENM*
- *Reception of a DENM*
- *Relevance detection of a received DENM*
- *Awareness detection of a relevant DENM*
- *Presentation of the DENM warning on the HMI*

Events can be defined in event models from analogy with discrete event models or state machines in systems engineering. Within one event model:

- The service, application, device, system or component can only be in one event at a time
- Trigger conditions define transitions from one event to the next event

Figure 7 gives an abstract example of an event model with the events S1 – S4 as blue boxes and the transitions are grey dotted arrows.

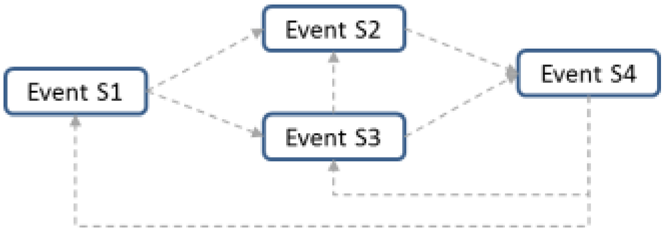

Figure 7: Event Model

An event model does not have to be a complete model. The event model should only define the events that are relevant for the evaluation perspective. The event model should also be agnostic to implementations:

- An event model makes no assumptions on the architecture and whether all events are detected in a single component or in multiple components.
- Implementations may ignore events, or do not log events.
- Implementation specific events remain hidden and will be ignored in evaluation
- Transition logic is implementation specific, especially if the implementation defines hidden states between the events in the event model.
- Event models are defined independently of each other. Interactions between event models are implementation specific and are ignored in the definition of event models.

The transitions do not have to be logged for evaluation and are therefore not defined in the event models. Instead, evaluation may reconstruct the transition logic for specific implementations from the occurrence and timing of logged events.

Figure 8 gives an example for the generation of DENMs in an ITS-Station. For evaluation purposes only the relevant events need to be defined in the event model; for example, only the actions taken by the DENM facility upon generation, updating and termination of a sequence of DENMs. Other states of the DENM facility that are irrelevant for evaluation, like the idle state (dotted blue box), may be omitted in the event. Evaluation will also reveal whether a specific type of device always terminated a sequence of DENMs.

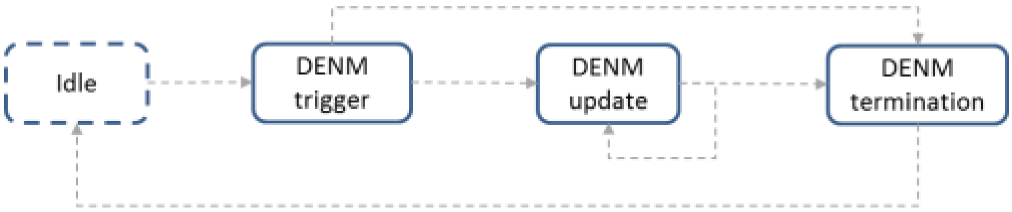

Figure 8: Event Model for generating DENMs

Figure 6 shows events from different models on a single timeline. Obviously, the event models from the perspectives of communication, application and HMI logic will be strongly related. An HMI warning will only be invoked when a DENM is received and the application matches the vehicle position on the RWW trace. These interdependencies are implementation specific, meaning that event occurrences, time delays, intermittent detections, etc. will differ per logging station. For evaluations, these interdependencies may be ignored, or the transitions may be reconstructed and analysed from the loggings for specific implementations.

6.3 Situations

Section 4.3 defines scenarios for piloting and testing that distinguish different situations for Use Cases. Situations should also be differentiated in the evaluations. The objective for distinguishing situations is to differentiate the performance of indicators within similar events and under similar situations. During evaluation, performance indicators will be calculated or aggregated per situation and event. The event model can be refined for situations; similar types of events can occur in different situations that may need to be distinguished for indicator calculations. Typically, situations are defined from situational variables. The following situational variables are widely used:

- Configuration of the road (lanes)
- Current speed limit
- Traffic conditions (density, flow, i/c ratio, etc.)
- Weather conditions (if extreme)
- incidents
- calamities
- Road construction
- Large events

- Season, holidays, weekend, etc.
- Rush hour

A choice needs to be made as to which of these will be collected or added to the database. Relevance of specific situational variables also depends on the services evaluated.

Situational variables are usually collected from external data sources, although they may also be measured or derived from measured parameters from central, road side, or vehicle systems.

6.4 Measures

Measures are the parameters that should be measured to calculate the nominator of a performance indicator. Measures are also defined to detect and distinguish events and situations. Measures are either directly measured by sensors, detectors or logged during pilots, or derived from direct measures.

The definition of a common set of measures can be divided in two steps:

1. Firstly, the parameters to be measured are defined by the physical property, unit, and minimum data quality criteria such as the sampling frequency, accuracy and precision.
2. Secondly, the format for collecting and storing the measures are defined.

Common definition of parameters in step 1 is essential to harmonise data analyses and evaluations. Obviously, unit conversions and resampling of data may be adequate for specific services and evaluations.

Ideally a common set of data formats are defined to be used by all pilots. Alternative formats may be used for specific devices or implemented services. The alternatives may differ in file or storage formats, subsets of (missing) parameters, and data qualities such as sample sizes and time synchronisation. The use of alternative formats necessitates the development of conversion tools. Currently two sets of measures are defined for usage in InterCor:

- SCOOP@F uses a log format defined in [12] and [13] for collecting log data
- An implementation of the event models for indicators is defined in [14], and the corresponding log formats in [15] and [16], that have been used in the first InterCor TESTFEST.

7 Pilot design

The design of pilots is developed in Activity 3, pilot operation [11]. The pilot design describes:

- Test site description
- Measurement plan
- Measurement period(s)
- Sample size and recruitment of participants
- Dealing with confounding variables

Confounding variables are situational variables that influence the functioning of the services outside the scope of the use case and are used to filter or eliminate “events” or “situations” from (parts of) the evaluation. For example: an incident causes congestion just before a road works site, thereby influencing the compliance of drivers to adapt their speed.

The pilot plan should also include baseline experiments that are needed for comparison and evaluation. A baseline could be collected from vehicles without driver advice systems, or in a pre-test without any services for example.

8 Pilot data collection and management

Activity 3 describes per pilot site, service or system, the process of data logging and collection of measurements including those needed for evaluation, as described in section 6.4.

It is assumed that pilot sites collect, store and manage their pilot site data in Activity 3. Pilot sites are also responsible for the protection of the data, including the implementation of privacy regulations and the protection of private data from test subjects and other stakeholders.

In a separate process, a pilot site provides the data that can be shared for evaluation. All data that a pilot site provides can be shared with other member states and partners. Pilot site data storage and data management is independent of the management of data for evaluations (section 9).

9 Evaluation data provisioning and management

To enable the harmonisation of evaluations across pilot sites and member states, data analysts and evaluators need to have harmonised access to all evaluation data. This is necessary to enable the cross assessments and comparisons of evaluation results as well as the quality assessment of the input data and analysis results such as calculated indicators, events and situations. It also facilitates the exchange and comparison of data analysis tools and scripts for example.

The next step in the evaluation framework is the definition of the processes and tools for managing and sharing the data for evaluation. This process is divided in three steps:

1. The provisioning of data by the pilot sites.
2. Assessment of the quality of the provided data.
3. Management of the provided data and evaluation results.

9.1 Data provisioning for evaluation

All data that a pilot site provides can be shared with other member states and partners. A common harmonised approach for providing data for evaluations needs to be agreed between Activities 3 and 4. The common approach for providing pilot site data includes:

- Common data formats for all data needed for evaluations. The collection of common data formats cover all measures defined for evaluation in section 6.4.
    - A proprietary format for logging and collecting pilot data needs to be transformed and provided in a common data format.
    - If alternative data formats for (subsets of) measures are defined as common, then transformations are provided, e.g. transformation tools or alternative data sources.
- Common data quality criteria are defined and used to assess the quality of measures. Data quality is required to assess the feasibility and reliability of evaluations and account for variations in data quality in the analyses and evaluations.
- A common process needs to be agreed for pilot sites to provide the data in a common format and quality, for example:
    - A pilot site can provide access to a data store for evaluation and analysis purposes.
    - A pilot site can upload the data regularly, i.e. after each test run or test series, to a central repository for evaluation as described in section 9.3.

A proposal for a common data format [14] and a central repository [17] for uploading and sharing input data and evaluation results has been implemented and tested during the first InterCor TESTFEST.

9.2 Data Quality Analysis

Upon provisioning of pilot site data, the data quality criteria need to be verified and analysed. Data quality criteria are defined for example for:

- Availability, frequency and value ranges, accuracy and precision of mandatory and optional parameters
- Plausibility of parameter values, e.g. for the consistency in location, speed and accelerations, and the time synchronisation of logging units.
- Availability, frequency and locations of events and situations, and comparison to the expectations defined in test scenarios.

A first implementation of data quality analysis has been tested during the first InterCor TESTFEST on the central repository data. This implementation is based on [18].

9.3 Data management

Activity 4 manages and shares data for evaluation and analyses results in a central repository. This repository manages access for uploading and downloading to all evaluation partners.

10 Technical evaluation

In the next period, Activity 4.2 will develop the methodology for technical evaluation, including the specification of the evaluation criteria and analysis for indicator calculation, event and situation detection, and data provisioning and management. Activity 4.2 will also refine the evaluation artefacts (Annex - Evaluation Artefacts) to include new technologies for hybrid communication and security, and new services.

11 Impact assessment

Drawing from the Evaluation Framework, Activity 4.3 will develop a detailed methodology for evaluating Impact at each of the pilot sites. The detailed methodology will, as far as is possible, use the same approach for each pilot site to maintain consistency across InterCor. The methodology will use and refine the Research Questions and the proposed services as basis for structuring the approach. This section will form a separate document to the Evaluation Framework.

12 User Acceptance

Activity 4.4 is to analyse elements and issues that influence the *penetration and effective use of C-ITS Services* and therefore, indirectly, the impact of these Services on traffic safety & traffic effectiveness in general terms.

Building on basic material considered to be available a Survey – launched widely or per Member State, and Focus Group Discussions are expected to provide the drivers for gaining insight and the basis for conclusions and recommendations.

13 Consolidation of InterCor Evaluation Results

General conclusions and recommendations do require a combination and synthesis of findings resulting from all evaluation 'sub activities' (technical – impact – user acceptance).

Such combination and consolidation shall be realized under the Activity 4 umbrella and leadership; in particular as it is reminded that the pilots (and therefore the findings relevant from an evaluation perspective) may be characterized by various levels of maturity and service deployment in the various partner Member States whereas also iterative developments of technologies, systems and services may occur.

Milestone 13, InterCor Evaluation Report, scheduled for delivery in August 2019, will comprise a concertation of Technical, Impact and User Evaluations from each of the pilot projects with an emphasis on assessment of Interoperability. This activity will be owned by the A4.0 lead. The final report will answer the research questions that form part of the Evaluation Framework using the detailed methodologies drafted under sections 0, 11 and 0 of this document. The InterCor Evaluation Report will be a separate document.

Bibliography

[1] FESTA Handbook, Version 6, 20 December 2016, http://fot-net.eu/Documents/festa-handbook-version-6/

[2] Grant Agreement under the Connecting Europe Facility (CEF) – Transport Sector Agreement No INEA/CEF/TRAN/M2015/1143833

[3] Description of the System Concept, Cooperative ITS Corridor System (ICS), Rijkswaterstaat, 4 juli 2016.

[4] Road Works Warning for the ITS-G5 context, Dutch Profile, Standaardisatie Tafel (NL), 2015-09-18

[5] ProbeData for the ITS-G5 context, Dutch Profile, Standaardisatie Tafel (NL), 2015-09-25

[6] Green Light Optimal Speed Advisory (GLOSA), Dutch Profile, Standaardisatie Tafel (NL), 2015-09-25

[7] French Use Case Catalog – Functional Descriptions, version 1.1- Draft, 18 May 2017.

[8] In-Vehicle Signage – High Level Service Description, file “Use case table template v10 - In-Vehicle Signage v1.docx” on the InterCor web repository ProjectPlace

[9] Use Cases Comparison, file “InterCor_A2.1_a_002_04052017_v0.6.docx” on the InterCor web repository ProjectPlace.

[10] Study on Key Performance Indicators for Intelligent Transport Systems, Final Report, in support of the implementation of the EU legislative framework on ITS (Directive 2010/40/EU), AECOM, February 2015.

[11] Milestone 8 deliverable (Sub Activity 3.1) – Pilot planning, file “InterCor M8 Pilot Planning V7.docx” on the InterCor web repository ProjectPlace

[12] SCOOP@F TLog log format definition, “SCOOP_2.4.1.3_CataOfDataTlog_v12.0_fr_1.xlsx”

[13] SCOOP@F Ulog log format definition, “SCOOP_151015)Livrable_2_4.1.3_Ulog_V1.0.xlsx”

[14] Data Logging for TESTFEST #1, “InterCor_DataLogging4TESTFEST_v1.0.docx”

[15] Communication log format definition, “InterCor_CommonCommunicationLogFormat_v0.5.xlsx”

[16] Application log format definition, “InterCor_CommonApplicationLogFormat_v0.5.xlsx”

[17] InterCor central repository for evaluation data, https://ada1.tno.nl/testfest/

[18] Wedemeijer et.al., Automating the analysis of field operational test data for the evaluation of cooperative systems”, ITS World Congress, Detroit, 2014

Annex - Evaluation Artefacts

The artefacts for evaluation from sections 0 and 0 are collected and maintained in a spreadsheet on InterCor web repository ProjectPlace. It defines:

- Research Questions (RQ) organised by evaluation task; Technical, Impact and User.
- Research Questions are prioritised and related to project objectives
- Research Questions are associated with Hypotheses and Indicators
- Indicators are collected on a separate sheet
- Categories of research questions, hypotheses and indicators are defined on a separate sheet

The spreadsheet is a living document that will be adapted and extended throughout the project for the new technologies and services, evaluation methodologies, and refinement during evaluations. Current version of the file is “RQ v 1.0.xlsx”.

