connected automated driving.eu

EUROPEAN COMMISSION

HORIZON 2020 PROGRAMME

CALL TOPIC DT-ART-02-2018: SUPPORT FOR NETWORKING ACTIVITIES AND IMPACT ASSESSMENT FOR ROAD AUTOMATION

ARCADE

Aligning Research & Innovation for Connected and Automated Driving in Europe

Proceedings of workshop on Common Evaluation Methodology

Deliverable no.	Contribution to D2.4
Dissemination level	Public
Work Package no.	WP2, WP4
Main author(s)	Sytze Kalisvaart, Hala Elrofai, Johan Scholliers
Co-author(s)	Maarten Amelink, Sami Koskinen
Version number	1.0
Status (F: final, D: draft)	D
Keywords	Common evaluation methodology, workshop, proceedings, ARCADE, CCAM
Project Start Date and Duration	1-10-2018 / 30-9-2021

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 824251

Document Control Sheet

Main author(s) or editor(s):

Work area: WP2

Document title: D2.4 Joint CCAM Networking Report

Version history:

Version	Date	Author	Summary of changes
0.1	3-12-20	Sytze Kalisvaart	General structure
0.2	3-12-20	Sami Koskinen	Section on break-out 2
0.3	7-12-20	Satu Innamaa, Yvonne Barnard, Hala Elrofai	Sections on plenary and break-outs 1 and 4
0.4	6-1-20	Sytze Kalisvaart, Johan Scholliers	Section on break-out 3
0.5	14-1-20	Sytze Kalisvaart	4. Analysis, 5. Conclusion, Executive summary
1.0	3-3-21	Sytze Kalisvaart	Section on break-out 5

Approval:

	Name	Date
Prepared	Sytze Kalisvaart	3-3-21
Reviewed		

Circulation:

Recipient	Date of submission
EC	
Project consortium	
Public	

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 824251

Contents

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 824251

Executive Summary

On 23 November 2020, ARCADE arranged an online workshop with between 70 and 80 participants on Common Evaluation Methodology as defined in the CCAM Strategic Research and Innovation Agenda.

In a plenary introduction, project presentation, panel discussion and five break-out sessions, a large number of gaps in the current evaluation methodologies were identified and discussed. The need for a Common Evaluation Methodology was clearly confirmed.

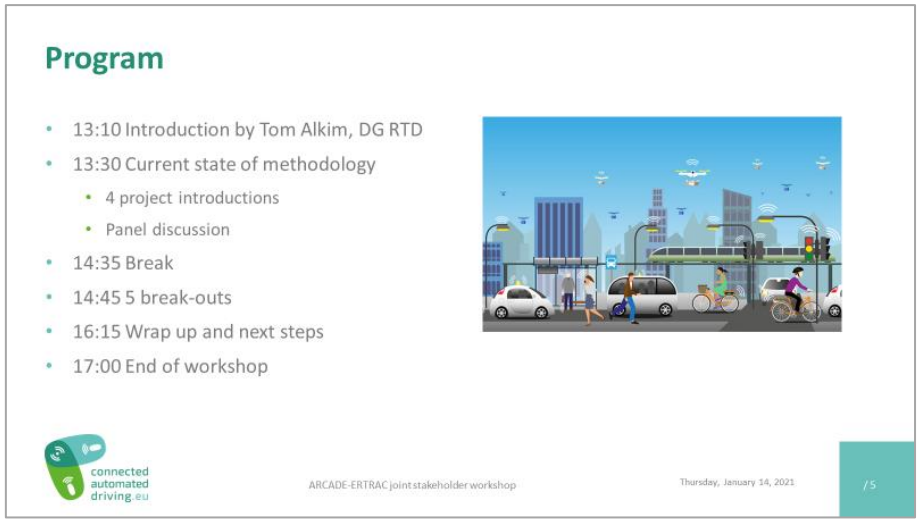
The major gaps were analysed after the workshop by arranging them according to FESTA V phases of carrying out a trial. The FESTA phases study design, data acquisition and impact assessment were prominent in the gaps.

The full list of gaps is given in section 4.1. The results will be taken up by the CCAM Platform and the ARCADE task on Common Evaluation Methodology.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 824251

1. Introduction

On 23 November 2020, an online workshop on Common Evaluation Methodology was held with between 70 and 80 participants. The agenda was as follows:



After introduction of the CCAM Working Group 2 work and the targets regarding Common Evaluation Methodology, the experience of the participants with regards to lessons learned from recent CAD projects and tests were reflected against methodology handbooks, such as FESTA and other frameworks, to identify gaps in current methodologies, guidelines and processes. For example, there have been discussion on how to include operational design domain aspects in different evaluation phases – and strong development of simulation tools ranging from safety validation of single scenarios to those that cover urban and regional planning.

Methodological gaps and lessons identified in the workshop will be input to an ARCADE knowledge base update on methodology and projects. They will also map the field for the upcoming CCAM actions. New evaluation activities will start in 2021.

1.1. Purpose of the document

This document is meant to capture the results of the workshop on 23 November 2020.

1.1. Intended audience

Experts, stakeholders and those interested in the field of evaluation in the domain of Connected and Automated Driving.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 824251

2. Plenary presentations

2.1. Welcome

The workshop was opened by the ARCADE project coordinator Stéphane Dreher (ERTICO) with a brief introduction to the ARCADE project and project contribution to CCAM and SRIA..

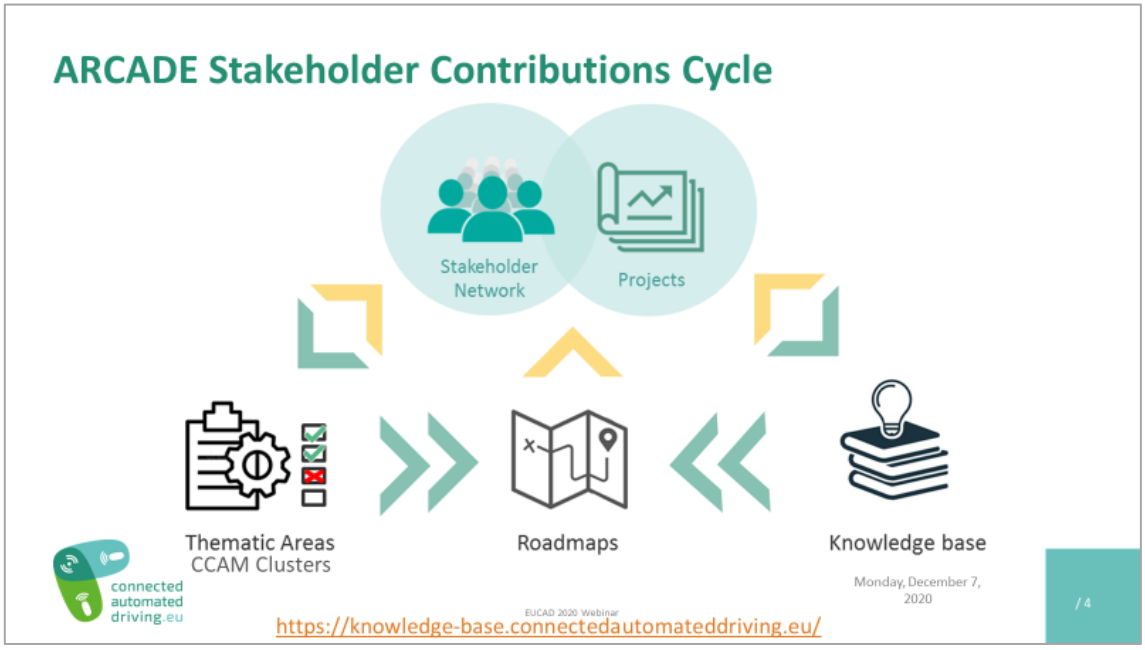


Figure 1: Stakeholder Contribution cycle presented by Stéphane Dreher

Then the workshop program and way of working and interacting during the workshop including Sli.do tool was explained by Sytze Kalisvaart (TNO), the workshop moderator.

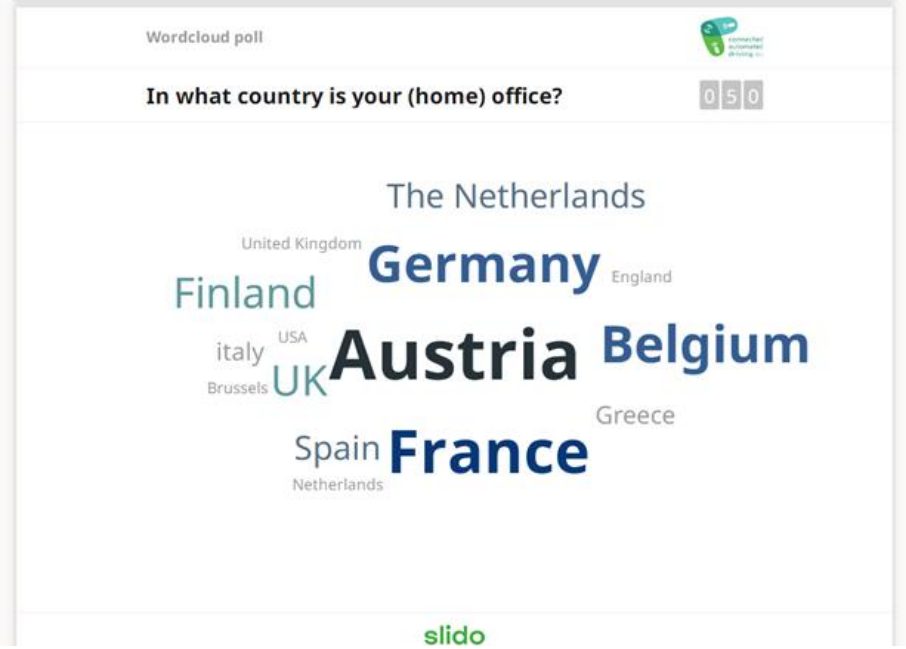


Figure 2: Sli.do results of the first question answered by the workshop participants

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 824251

2.2.Introduction of CEM

Tom Alkim of DG Research & Innovation introduced the CCAM Partnership and explained its objectives. Tom Alkim and Guus van de Schouw lead Working Group 2 (WG2) on Coordination and cooperation of R&I.

Next, he introduced the CCAM Platform and stressed the necessity of the testing. Therefore, one of the main goals of WG2 is to 'Develop common approaches for testing and assessing impacts' or in short 'Common Evaluation Methodology'.

Further he summarised the objectives and main findings of several workshops on EU Common Evaluation Methodology (EU-CEM) held by the members of WG2. The workshop efforts resulted in a draft paper that will become publicly available early next year. Also the elements of the envisioned EU-CEM and how it is supported by ARCADE Knowledge Base were presented as shown in **Figure 3**.

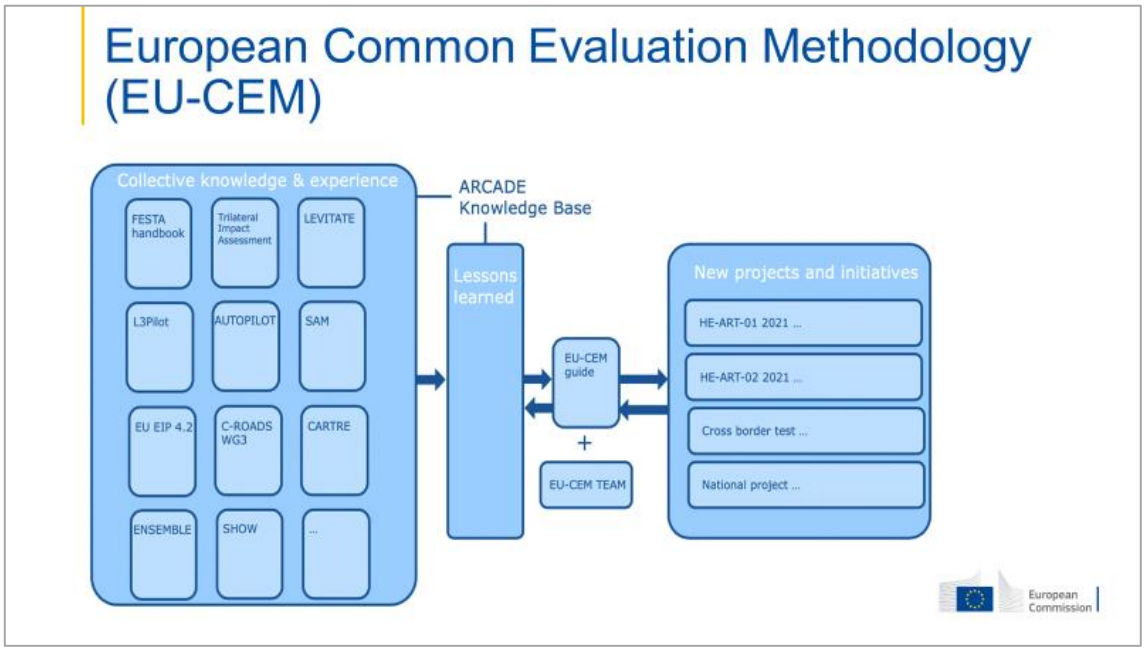


Figure 3: Elements of the envisioned EU-CEM

Tom Alkim finalised his presentation with the next steps of WG2. He showed where to find 'Guidelines and Evaluation Methodologies' on the ARCADE Knowledge Base.

2.3.Current state of evaluation methodology

This part of the program consisted of presentations. The first one was about Methodology baseline and known gaps. The other presentations introduced four projects, three EU projects and one national French project.

Methodology baseline and known gaps

Yvonne Barnard from the University of Leeds presented the current methodology baseline. She started by explaining the contents that are currently available in 'Guidelines and Evaluation Methodologies' section in the ARCADE Knowledge Base.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 824251

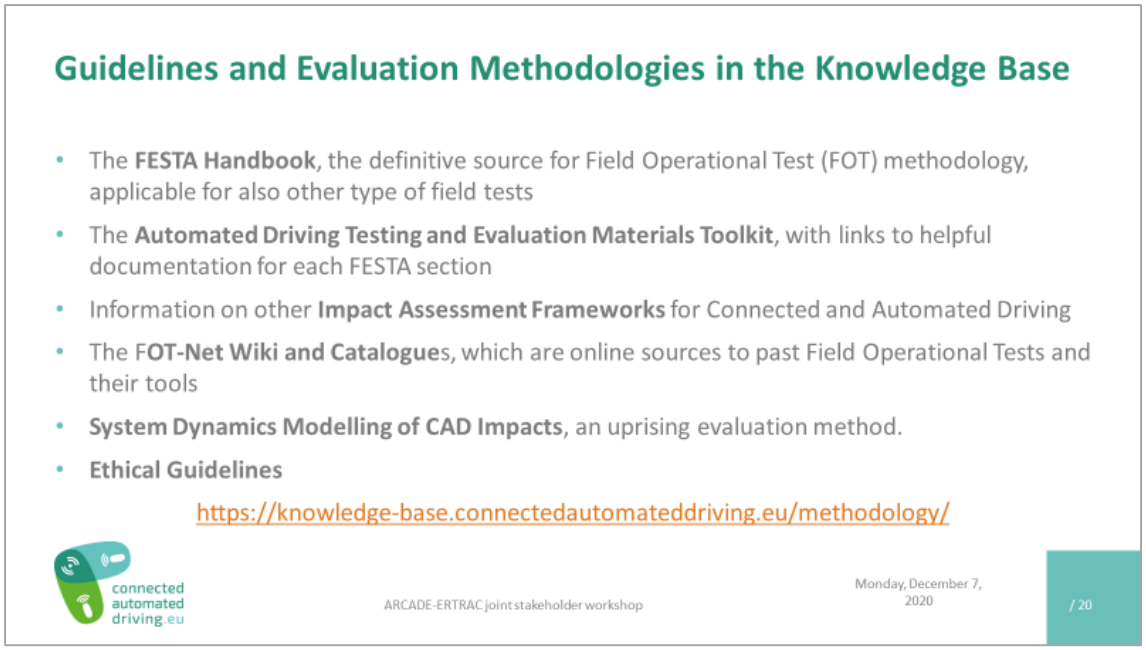


Figure 4: Content of 'Guidelines and Evaluation Methodologies' in ARCADE Knowledge Base

As identification of the methodological gaps of common evaluation frameworks is main objective of this workshop, Yvonne closed her prestation with addressing several of previously identified gaps resulting from different projects.

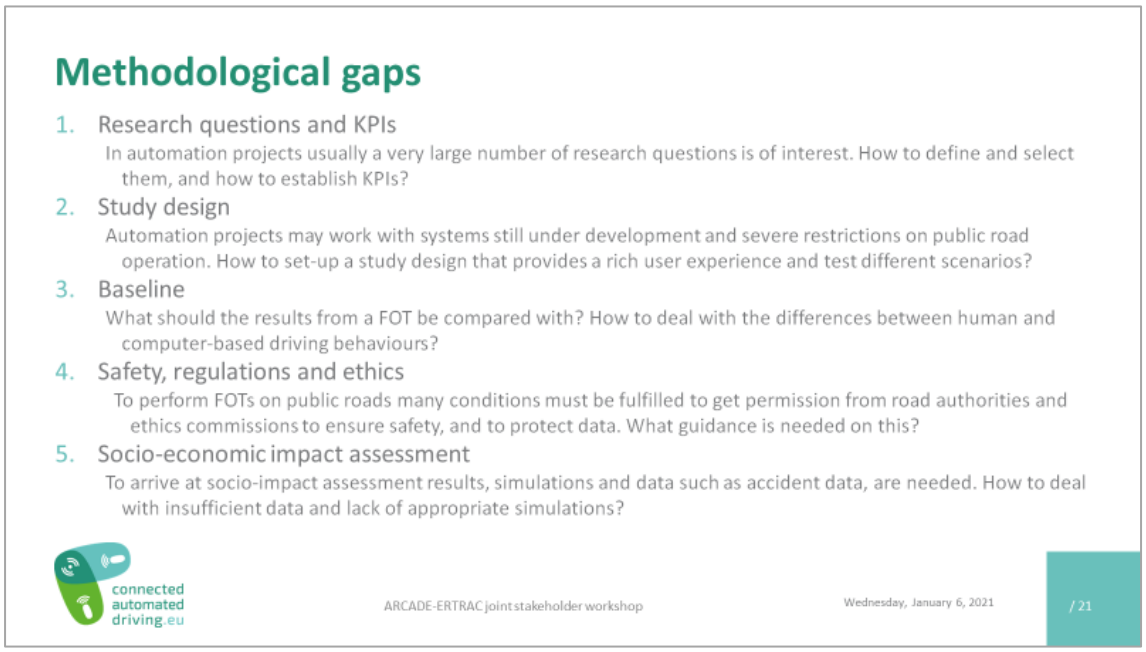


Figure 5 Previously defined methodology gaps

Project introductions

Each project presented its main goals, the used evaluation methodology and identified gaps for this methodology. There were several questions asked by the participants for these projects.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 824251

EU LEVITATE

Hitesh Boghani of Loughborough University presented the EU Horizon 2020 project LEVITATE[2]. One of the main goals of the project is to establish a multi-disciplinary methodology to assess the short, medium and long-term impacts of CATS (Connected and Automated Transport Systems). Evaluation methodology gaps of LEVITATE include common assumptions about CAVs, changes in human behaviour.

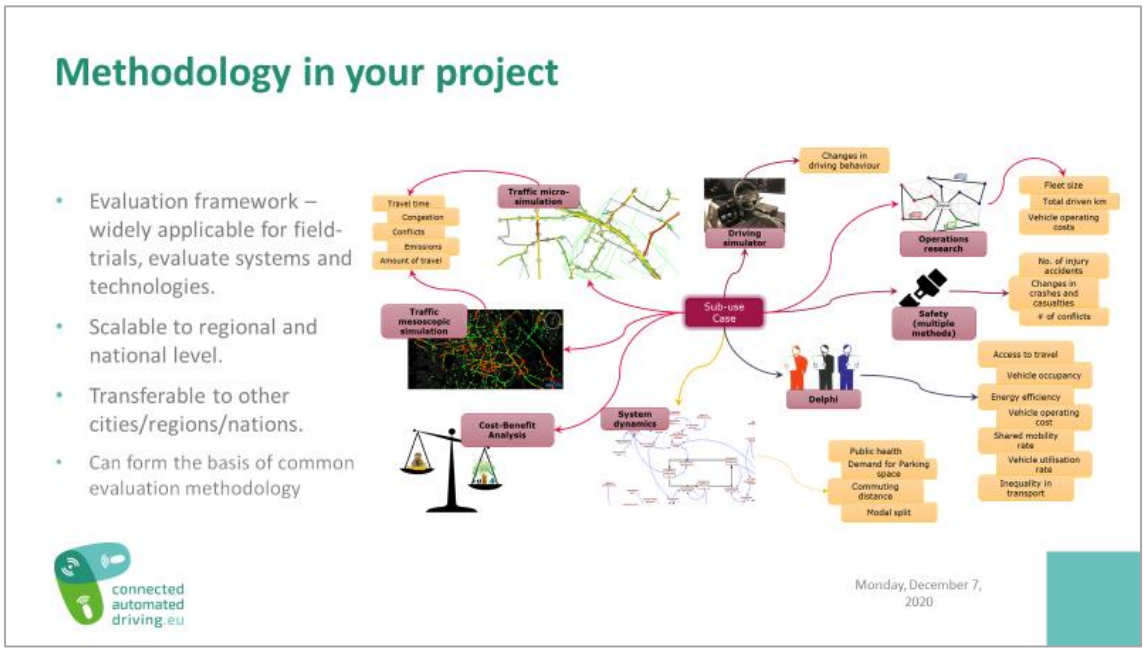


Figure 6: Developed Evaluation Methodology as presented by the EU project LEVITATE

FR SAM

Nadège Faul from Vedecom presented the three-year national French project SAM (Safety and Acceptability of autonomous Mobility[4]) and the largest project budget-wise. Main objectives of this project are to deliver the methodologies, data and evaluation results to evaluate the safety, acceptance and overall impacts of automated driving and mobility services. The overall approach to the project has been based on the FESTA methodology. The methodological gaps included references for coordinating multiple FOTs, lack of baseline for service evaluation, acceptance, etc. and assessment mapping model (data evaluation architecture).

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 824251

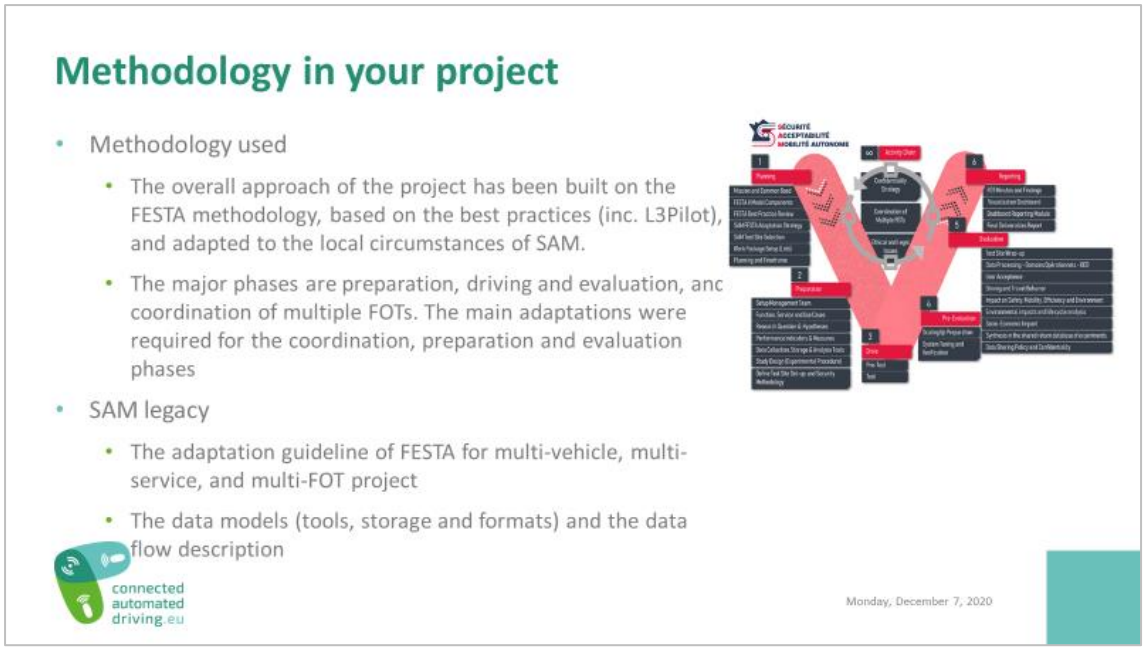


Figure 7: Evaluation Methodology used by SAM project

EU HEADSTART

Bernard Hillbrand from Virtual Vehicle, Austria (ViF) presented the three-year EU HEADSTART project[3]. The main goal of this project is to define testing and validation procedures of CAD functions. HEADSTART adopted a scenario-based safety assessment methodology. The methodological gaps of this approach included lack of standardized output format of scenario databases, lack of standardized format for ODD description and non-existence of a method to combine results evaluated from different tests.

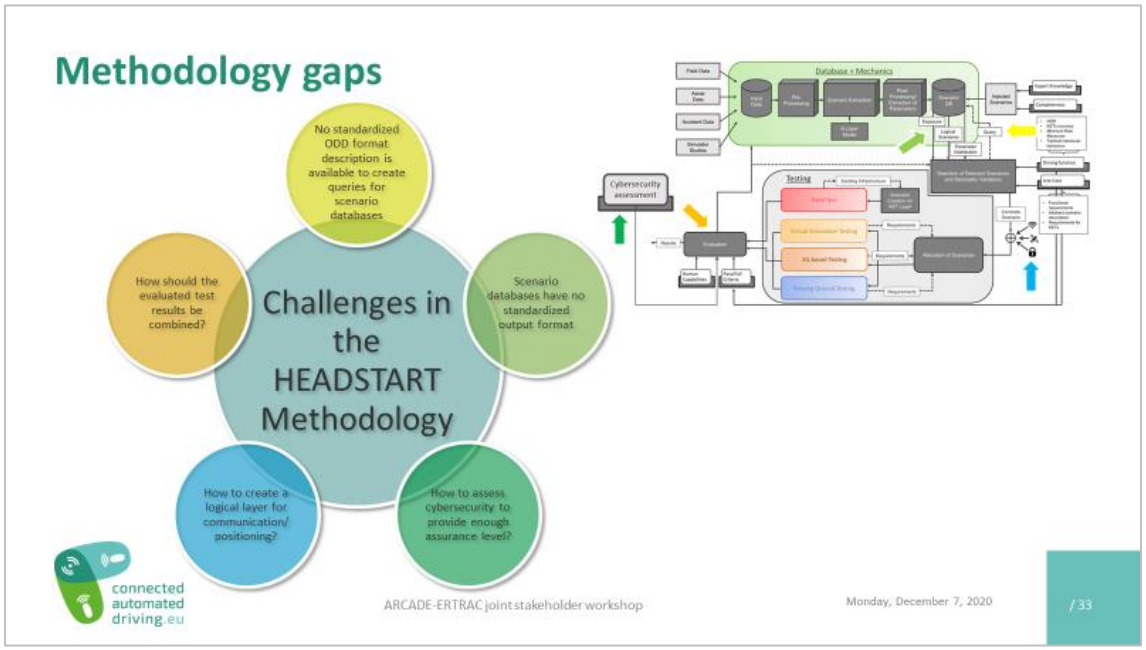


Figure 8: Methodology gaps as presented by the EU project HEADSTART

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 824251

EU L3Pilot

Satu Innamaa from VTT presented the 50-month EU project L3Pilot[5]. The main objective of this project is to assess the acceptance and the socio-economic impact of conditionally automated passenger cars by conducting real-world experiments in Europe with AD functions: motorway, urban, parking. This project also adopted FESTA as an evaluation methodology. Methodological gaps found for this project include a gap between prototype testing and the scenarios in impact assessment, principles for merging data from multiple pilot sites without compromising the privacy of OEMs and methods to predict future for the socio-economic impact assessment.

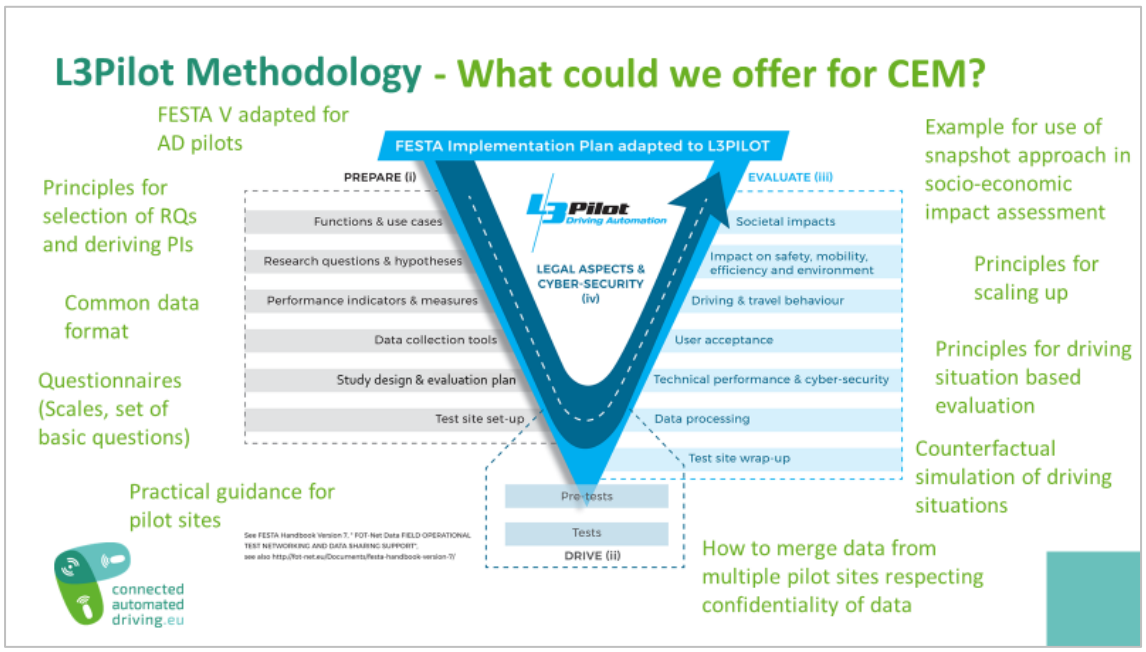


Figure 9: Evaluation Methodology as presented by the EU project L3Pilot

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 824251

2.4. Panel discussion

The panel discussion started with answering the questions addressed by the moderator. The panel consisted of Tom Alkim from DG RTD and the project representatives Hitesh Boghani, Nadège Faul, Bernard Hillbrand and Satu Innamaa.

Several questions/multiple-choice poll have been answered by the audience using Sli.do.

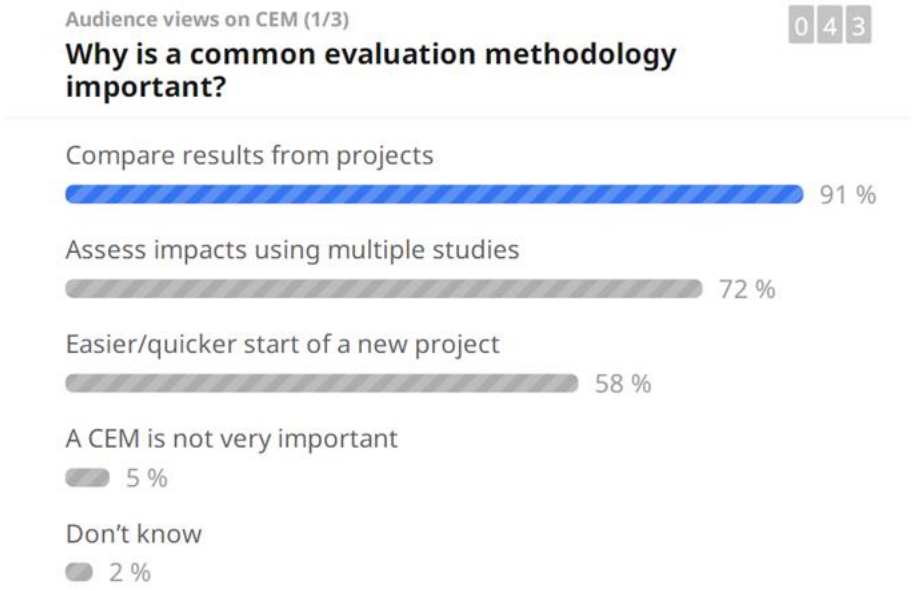


Figure 10:Example 1, multiple-choice question answered by the participants

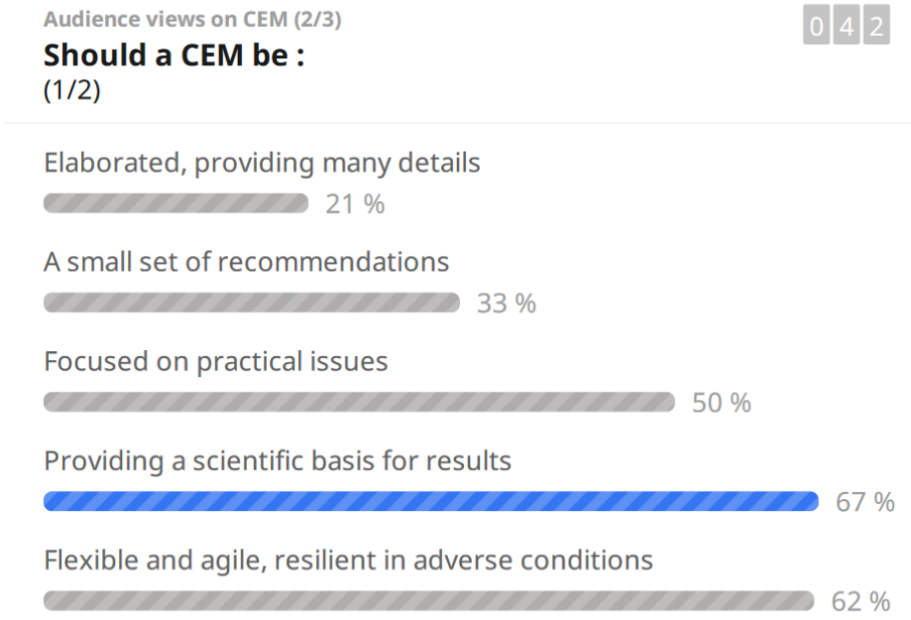


Figure 11:Example 2, multiple-choice question answered by the participants

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 824251

During the panel discussion the workshop participants had the chance to add comments and ask questions, using sli.do or MS TEAM chat, about the presented projects or evaluation methodologies in general.

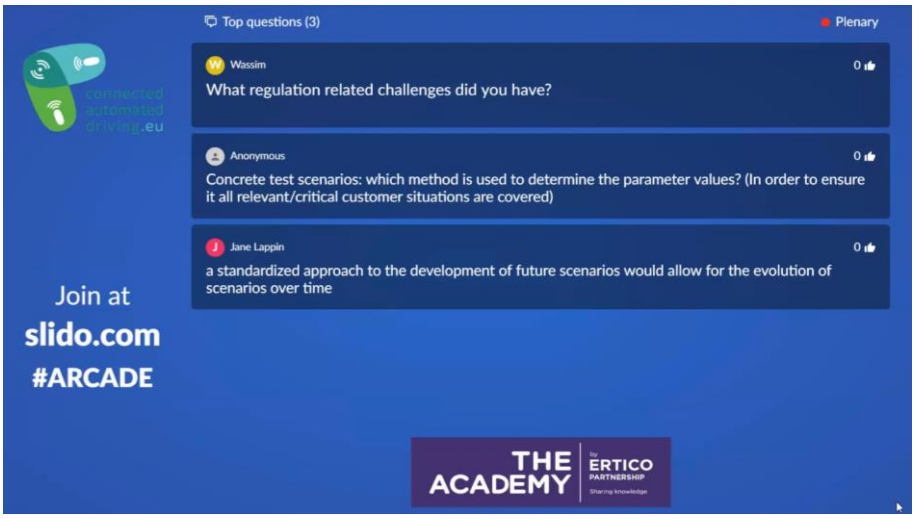


Figure 12: Questions asked by the participant during the panel session.

2.5.Wrap-up and next steps

The workshop was concluded with the wrap up of the breakout sessions (see the following chapter) and next steps. The participant's input regarding the next steps was collected using Sli.do.

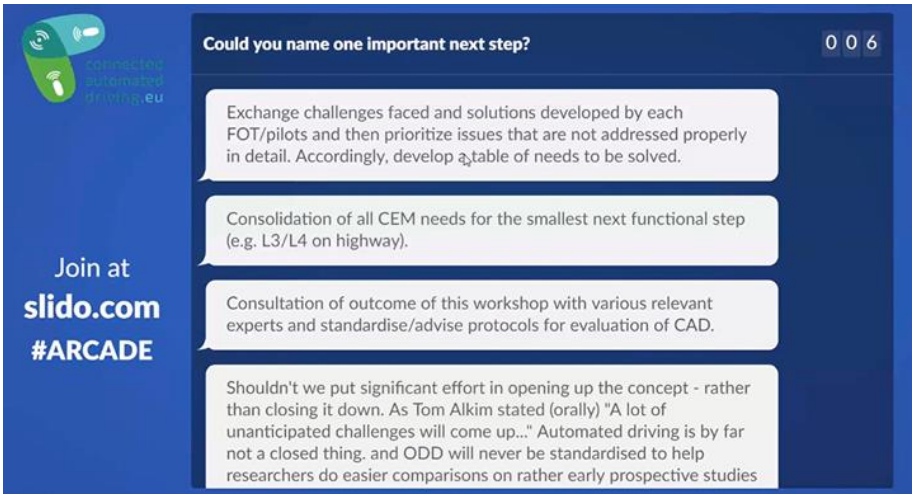


Figure 13: Examples of participant's input for the next steps.

Tom Alkim concluded with the next steps:

- Finalize CCAM report on WG2, publicly available Q1 2021
- Identify gaps in the current methodologies with input from this workshop
- Distribute work and activities between: CCAM Platform, CCAM Partnership, CSAs and projects
- Analysis of the results of this workshop by ARCADE

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 824251

3. Break-out sessions

3.1. Large AD pilots

This breakout session focussed on large AD pilots The session was moderated by Satu Innamaa (VTT) and Isabel Wilmink (TNO). It addressed the following questions:

1) What should be in the guidance given in CEM?
2) What should not be in the guidance given in CEM?
3) What can your project provide for CEM?
4) Which gaps we still have in methodologies that should be developed for CEM?

The participants provided their views with interactive Padlet tool. We also discussed the given input in the session.

The results are given in the tables below.

What should BE in the guidance given in CEM?

Topic	Reasons
Common minimum dataset	To have an overview on what usually is necessary
Common glossary	So that all projects share the same language
Structure and KPIs	A general common approach, areas of evaluation with some preliminary indication and a list of possible KPIs
Established evaluation results	So that we can identify areas that lack evaluation or use cases that have not been covered by evaluations
FESTA methodology structure	So that no steps are forgotten, e.g. formulating research questions
A list of common methodologies that could be utilised in projects	One size does not fit all so options for evaluation would be helpful.
Examples	For each step, examples that new projects can use as a starting point, for instance RQs and KPIs that are used regularly
Reporting structure	Provide a basic structure where to report results in order that it is easy to find the information you're are looking at
Lessons learned on options for baseline	Pros and cons of different alternatives, example how it is used in practice
Cost benefit analysis	Turning impacts into costs and benefits allow a comparison between different interventions/use cases, etc.
Addressing uncertainties and variabilities	Because there will always be those
Knowledge base	Common platform to share evaluation studies
User aspects	User behaviour, user acceptance
Lessons learnt	from previous projects: common pitfalls, things that did not go as planned for reasons you only find out during the project

What should NOT BE in the guidance given in CEM?

Topic	Reasons
Mandating instructions of methods	Better leave room for further development

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 824251

Too much details	Not many details should be in. No strict indication on how to evaluate. Not made the method too rigid otherwise it is not applicable
How evaluations should be conducted	Some evaluation methodologies are still research areas and new approaches could provide additional insights

What can your project provide for CEM?

Project	Item(s)
LEVITATE	Scalability and transferability of results - Being able to scale-up/down results and transfer them to different regions/cities would not only help them to use outcomes of various projects but also lead to less resource demanding activity and faster implementation.
C-Roads	C-Roads can contribute to a common evaluation methodology and provide evaluation studies on C-ITS from all Europe
UK national projects	UK might be able to share national methodologies developed up to now
L3Pilot	Common data format FESTA V for AD pilots Surveys Options for baseline Practical guidance for running the pilots

Which gaps we have in methodologies that should be developed for CEM?

Gap	Comments
Data sources for describing the traffic and accidents etc. on EU level, incl. all dimensions of ODD	
Largely agreed safety indicators (i.e. model between KPI and expected impact)	
Common approach	To have a decided common approach around Europe.
How to construct future scenarios	this is highly complex, what to take into account, what can be ignored?
Commonly agreed way of modelling AVs in commercial simulation tools	With examples + way how to describe how vehicles were modelled
Up-scaling	Many of these projects have to face a difficult up-scaling exercise from the experience with few vehicles to assess the impact at general level. This is really missing
Combining different impacts together from different methods	

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 824251

3.2.Small and national projects

The session was moderated by Sytze Kalisvaart (TNO) and Maarten Amelink (RWS). The session started with a tour de table of the participants, stating what were their expectations regarding the workshop.

The session started with a presentation of Karl Rehrl, Salzburg research, on the Digibus Austria[6]. This local project performed research and tested methods for traffic safe-operation of automated shuttles on open roads in mixed traffic on SAE level 3. The project developed a 7-step process model for the deployment of automated shuttles. The project also identified gaps in the evaluation methodology, as there is currently no common evaluation methodology for automated shuttles.

The session continued with an interactive survey, using the Padlet tool, including introductory questions regarding differences between small and large projects. The average duration of a small project is about 3 months. The participants can spend about 20-33% of the work to define the methodology. The project include only a small amount (10-20%) of routine tests. Major differences between national and large projects are that national projects are more focused, trying to answer a small number of research questions, with less administration than large scale projects. More flexibility and agility is needed to achieve the goal of the local project, whereas bigger projects have generally more slack.

Table 1 Answers to the first survey regarding differences between small and large projects

Question	Responses				
What do you consider a small test project?	1 week of work	1 month of work	3 months of work	1 year of work	
	0	1	**7**	0	
How much of your work can you spend to define the methodology?	5 %	10 %	20 %	33 %	50 %
	0	1	**4**	3	
Which share of your tests is routine test?	10 %	20 %	33 %	50 %	75 %
	3	**4**	0	0	0
How well do elaborate methodologies scale to small projects?	Very little	A bit	Same Principles	Mostly	Very well
	0	3	**4**	1	0
What are the main differences between large projects and small projects in methodology?	Small scale projects try to answer more focus on a select number of research questions. (3x)				
	Smaller budget, more focused, more concrete				
	Small project needs flexibility and agility to go more directly toward the goal				
	Need to be very pragmatic				
	Need to avoid risks				
	Smaller projects are often regional and not always aware of national larger pilots and findings, as well as the best methods for evaluation already established e.g. best way to measure impact etc. and use of common questionnaires. More focus on technical evaluation.				
	In a big project, it may be easier to get independent evaluation				
What are the main differences between	Focus on fixing local issues. So only piloting specific aspects of interest to the particular region				

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 824251

EU and national or local projects?	National: less administration, fewer language problems (3x)
	More focused work in national projects, less overhead
	National projects often lack the European scale
	EU Projects enable larger comparison

A second interactive survey addressed the methodology gaps in small projects. The results of the survey are shown in Table 2.

The main methodological problems in project relate to the confidentiality of the data, and the trust building with industrial partners. GDPR makes access to data more difficult, especially for camera data. Another issue which makes getting data difficult, is that due to technical issues, the main focus is on getting the service working and the time to collect data becomes smaller.

The Common Evaluation Methodology (CEM) should include KPIs (baseline and common KPIs), shared model for coming from measurement to evaluation, expert groups for Delphi estimation of topics which are hard to model and best practices. Technical evaluation and work already under standardisation should not be included.

The most important gap in the methodology to be addressed by the CEM is the access to sensitive data, and agreeing on the principles of data sharing among the industry and research. Other gaps to be addressed are the adaptability to different project scales, agile methods dealing with project issues, advice on common issues and the evaluation of safety benefits of AD and socio-economic issues.

Table 2. Methodology gaps in small projects

Question	Answers	Likes
What methodological problems do you see in your projects?	Working on national projects with multiple stakeholders (public transport operators, car manufacturers, shuttle providers), there a lot of competitive and sensitive topics. On these topics, it is difficult to get data for evaluations (for example socio-economical aspects)	**6**
	Getting across confidentiality of industrial partners involved	1
	In the context of this group is having regional trials aware of national trials and their agreed methods for evaluation - not reinventing the wheel so to speak. The power nationally is to combine any regional results. UK has a government related	2

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 824251

	body called Zenzic that is trying to look across all national CAV pilots. Regional projects are often with specific suppliers who might not share their data.	
	GDPR process to get data from users. Way to get evaluation feed-backs from other road users. To have enough users	
	Building trust with industrial partners	
	How to estimate changes in consumer behaviour with higher availability of CAD?	
	Availability of data, project time slip compressing evaluation period	
What can your project provide for the CEM?	Reference scenario database	
	Methodological material for evaluation	
	Connectivity in scenarios	
	Liaison with Zenzic regarding national pilot approaches (methods)	
	Process model to setup shuttle trials	
	Impact simulation	
	Road safety inspection and virtual risk assessment	
What should be included in the Common Evaluation Methodology?	KPIs. Ensure key ones for common evaluation are agreed as other countries may focus on others as well.	**4**
	Shared model for coming from measurement to evaluation	**4**
	Common data models	
	Step by step process model	1
	Expert group for Delphi estimations of qualitative / hard to model topics	1
	Example of information sharing principles and policies in order to foster trust building in the project to ease information sharing	2
	Best practices	
What should not be included in the Common Evaluation Methodology?	Technical evaluation - Should be a pre-cursor to impact evaluation and left to	

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 824251

	individual project evaluation leads.	
	Work already under standardisation.	
	The databases or tests themselves	
What gaps in methodologies should the CEM address?	Dealing with issues. Advice on common issues - like an FAQ list perhaps	1
	Evaluation of safety benefits of AD and other socio-economic issues	2
	Access to sensitive data or guidance how to arrange this with industry	1
	Agreed principles for data sharing among industry and research	**6**
	Adaptable to different test scales	3
	Agile methods, dealing with project issues	1

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 824251

3.3.Simulations and data

This breakout covered topics related to simulation tools and test databases. The format of the breakout was a normal teleconference, where everyone was free to weigh in. In the beginning, participants briefly introduced themselves.

The moderators Sami Koskinen (VTT) and Adrian Zlocki (FKA) had prepared five discussion topics:

- Gap 1: Accurate integration of safety strategies in existing traffic simulation software
- Gap 2: Edge cases
- Gap 3: Simulate automated mobility in the context of urban planning
- Gap 4: Could new projects share a common log data format?
- Gap 5: Video sharing is difficult between project partners and GDPR doesn't help.

Each gap – a difficult evaluation topic – was first introduced by the moderators for the group, to start the discussion. For every topic, the breakout was lucky to have a couple or more experts contributing personal experiences and recent news.

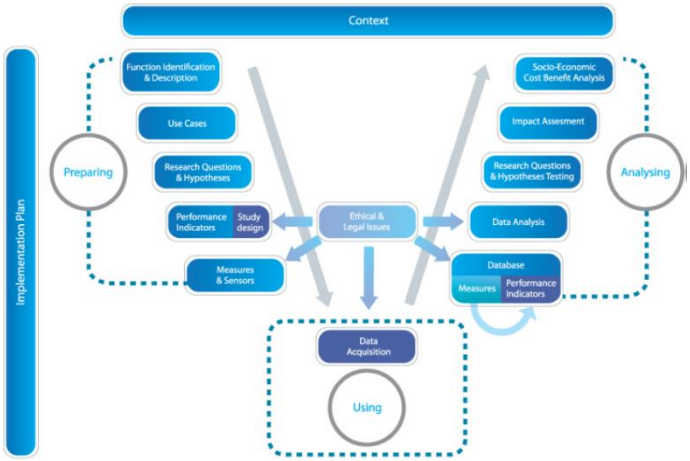
Regarding simulations, the group first discussed initial ongoing efforts to harmonize modelling of automated functions in current commercial tools. Analysing and sharing data from near-misses and edge cases was seen important to make the first simulation models more detailed. However, regarding such edge cases, there are no established definitions and classifications, yet.

Regarding data topics, the group considered experiences from recent projects sharing data and using a common log data format for several organisations to collaborate on evaluation. Despite of past work, new projects struggle with data formats due to having a new focus or other limitations to use existing tools. Modularity of a common log data format was suggested as a potentially easier way to work forward: enable projects to select parts of the format and related subset of tools.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 824251

3.4.FESTA Methodology overview

The session was moderated by Yvonne Barnard (University of Leeds) and Julie Castermans (ERTICO). In this break-out group we explained the FESTA methodology, and its main steps, for participants who are not familiar with the methodology or who do not have experience in Field Operational Tests or automation pilots.



The FESTA V distinguishes 3 phases, preparing the study, using the functions or systems on the road, gathering data, and analysing the data and determining the socio-economic impacts. The FESTA methodology is described in the FESTA handbook, version 7, available from the Knowledge Base. For all three phases, the Knowledge Base also contains detailed information and recommendations, as well as recorded webinars describing the different steps. FESTA is owned, developed and updated by the large FOT community.

There is additional information about the Data Sharing Framework and the Trilateral Impact Assessment Framework for Automation in Road Transportation.

FESTA has been updated in the last version of the handbook, addressing automation, but more is needed to arrive at a full evaluation methodology for automation projects. A common evaluation methodology enables a structured approach and ensures scientific rigour, providing evidence on the effects. This may support everyone involved in designing and conducting FOTs and pilots and provides a common vocabulary. What we want to learn from automation FOTs and pilots goes beyond a single project. FOTs and pilots should give insight in the behaviour of vehicle (systems), users, infrastructure, and services, and identify the opportunities and challenges for the future. We want to gain knowledge about the impact of road automation on society, users, and industry. What are the effects on mobility, safety, environment, and efficiency, but also on wider impact areas such as land use, health, and energy?

A common methodology is therefore important for being able to study the impacts of automation, combining all knowledge gathered from different kinds of FOT and pilot.

Some of the challenges and gaps for automation FOTs and pilots identified in FESTA:

FESTA step	Challenges and gaps
Context	For vehicle automation, we are talking about something that could be a societal game-changer, what are the wider lifestyle, transport, economic and social impacts?
Research questions and hypotheses	New research questions arise • How to prioritise them? • It may be difficult to formulate precise hypotheses

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 824251

	• Baseline question: What do we compare the outcomes with?
Study design	• Restricted scenarios or space mission approach? • Use cases cannot be only defined by what is desirable to evaluate, but also depends on what is allowed to be tested on the real road • Study design will need to be more flexible, allowing for exploration, and iteration and revision during the project • New methods are needed for measuring user acceptance, capturing user experience
Analysing data	• Large data sets are recorded and different sensor types provide different information and data quality • Need for automated processes for data and video analysis, and synchronization among all local data bases • Data privacy and data safety is to be ensured • Novel evaluation methods needed, e.g., visioning, scenario development, data mining, machine learning, automated scenario detection. • How to evaluate AI processes and decisions? • New role and new methods for simulations
Legal and ethical issues	New questions and dealing with ongoing and divers regulation need to be addressed

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 824251

3.5.Framework for efficient operation of automated fleets: the early days

With the large-scale development of automation, the number and size of pilots and Field Operational Tests (FOTs) is growing. For operation of professional fleets like taxis and delivery vans, various forms of monitoring are already in place. With the deployment of automated vehicles, in-use monitoring of vehicles may be used to optimise operation and safety. All this calls for a wider framework for operation. Since it is early days, this workshop focussed on outlining such a framework.

This break-out was moderated by Nadège Faul (VEDECOM) and Stéphane Dreher (ERTICO). Various presentations were given. Alvaro Arrue of IDIADA addressed the perspective from the perspective of the EU HEADSTART project[3]. He emphasised FOTs provide key real-world information to validate the safety assurance. In a stepwise approach, the complexity of the ODD can be expanded and the system improved for functioning within that ODD.

Within the context of evolving regulation and standards, particular CCAM use cases are the focal point of such FOTs. By carefully ranking CCAM use cases and their corresponding ODDs, step by step a comprehensive framework can be built up.

Henriette Cornet of UITP introduced the SHOW project[6], which pilots automation in 20 EU cities. It looks at services like Public Transport (PT), Demand Responsive Transport (DRT), Mobility as a Service (MaaS) and Logistics as a Service (LaaS). The use cases are automated mobility in cities, mixed mobility in cities and added-value services.

Diversity is the keyword and challenge in SHOW: diversity of partners, demo-sites, user needs and impacts (including sustainability and social impact) and legislation (testing and long-term). This calls for a higher-level management framework in addition to many local solutions. The participants in the break-out considered his diversity an advantage under COVID-19 epidemic, as it allows to switch locations and partners when needed. In a poll, a preference for centralised lessons-learned and structured concertation between the pilots was given.

Nadège Faul of Vedecom continued on the SAM project[4] (see also the Project introductions (Section 2.3). This involves multiple pilot sites, 12 evaluation domains and 60 levels of research questions. The focus is on L4 automation with various services, including new ones. The challenges:

- There is no reference or comparison base for some evaluations (like delivery service).
- Various levels of scale, from micro (user) to macro (ecosystem)
- Urgent need for unified description of ODD, use case and services (order defined by participants), possibly also vehicle configuration

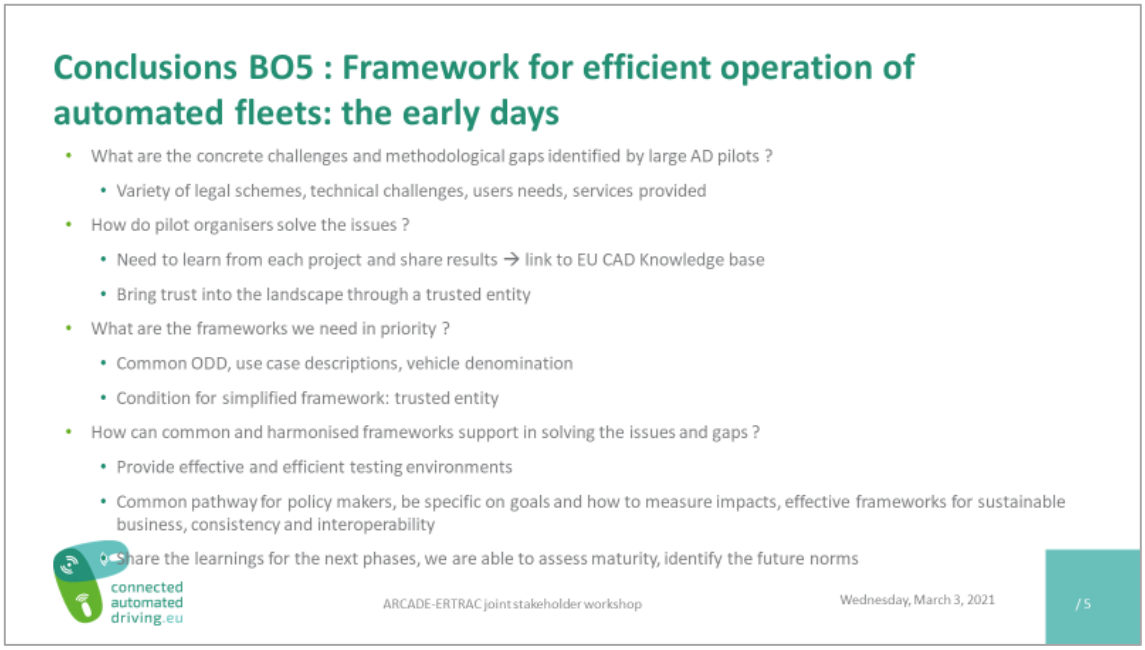
The participants saw as benefits of harmonization between projects and formats that the project becomes more agile and comparable in its decisions.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 824251

Martin Russ of AustriaTech addressed CCAM operations. He identified the need for transitional operation with an approval schema and continuous monitoring. In this way, risks can be manageable and be identified with monitoring.

He also proposed to create a trusted entity as caretaker for tests and deployment of prototype vehicles. This entity would consist of multiple stakeholders.
Though many roadblocks have already been identified at the High Level Meeting in Vienna 2018, harmonization is still much needed. The EU CAD knowledge base provides a start for that. A common pathway for policy makers is still missing. The process seems to move away from the targets and drift towards a complex legal body. An innovation friendly framework is needed, e.g. containing transitional operation and a trusted entity. The CCAM partnership will help to move from even larger pilots to operation.

The participants collected the following conclusions:



The gaps identified in this break-out session were:

FESTA stage	Topic	Gap
Data acquisition	Representativeness	The data collected needs to represent the current and future ODD. How to determine coverage?
Use cases	Priority	What use cases should have priority, given expected market introduction year, availability of test systems and permission on public road?
Study design	Reference	For new services, there is nothing to compare the pilot results with. What is better?
Research questions	Scale	How to manage many levels of scale in research questions?
Data acquisition	Coordination	How to manage many levels of scale in data collection?
Use cases	Description	Need for a common description of ODD, use cases and services
Ethical and legal issues	Diversity	Need for an innovation friendly framework for running pilots, FOTs and operation

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 824251

Ethical and legal issues	Roadmap	Need for common pathway for policy makers across EU
Implementation plan	Sharing	Shared learnings across projects and pilots to raise the common maturity

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 824251

4. Analysis of the results

Based on the inputs of the workshop, the ARCADE team made an analysis of the results to come to aggregated conclusions. An overview of the key gaps in the Common Evaluation Methodology is given below. For the break-out sessions, only the major gaps are listed based on participant support. It is sorted by topic along the FESTA V-cycle (Figure 14).

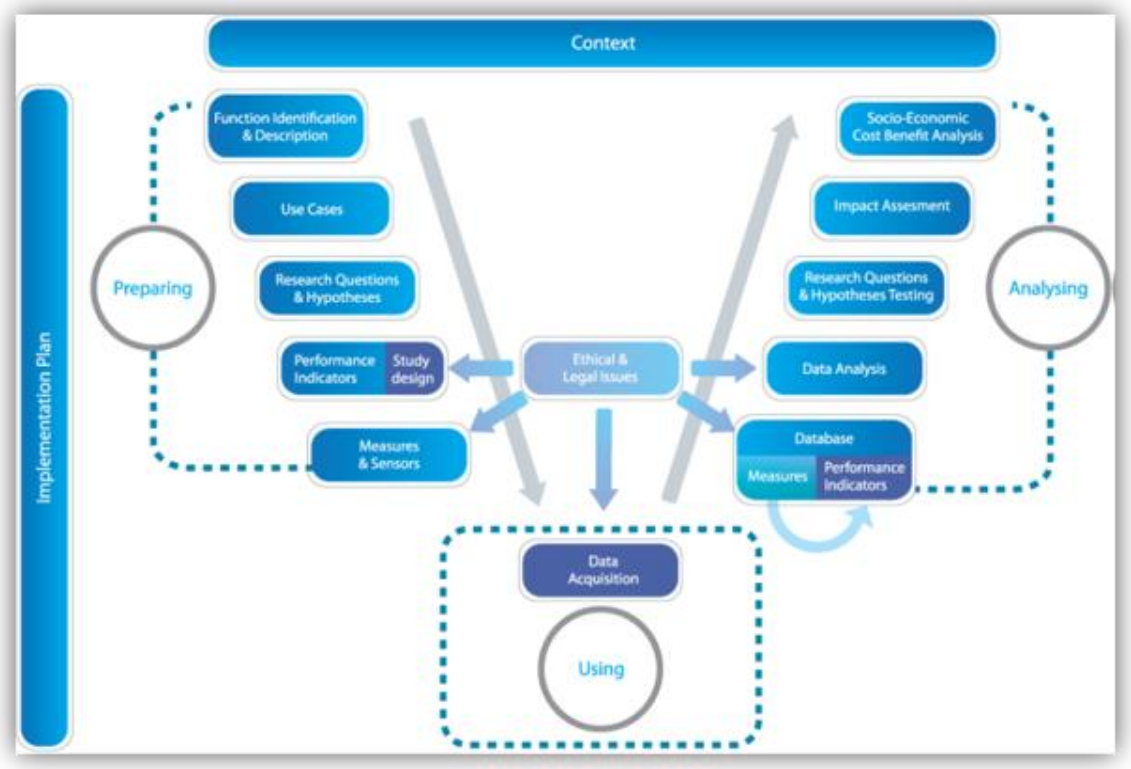


Figure 14 FESTA V

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 824251

Table 3 Gaps identified in CEM workshop sessions

Source	FESTA stage	Topic	Gap
SAM	Implementation plan	FOT management	Reference for coordinating multiple FOT (monitoring tools, confidentiality issues, legal issues..)
Break-out framework	Research questions	Sharing	Shared learnings across projects and pilots to raise the common maturity
Levitate	Function identification and description	Functional requirements	Common framework for functional requirements to AVs
SAM	Function identification and description	Terminology / ontology	Common description methodologies for vehicles, services, functions, use cases, etc…
HEADSTART	Function identification and description	Terminology / format	No standardized ODD format description is available to create queries for scenario databases
L3Pilot	Use cases	Input data and assumptions	Lack of single data source to define target accidents and travelling inside ODD of each AD function. Principles for combination of multiple sources
Break-out Simulation and data	Use cases	Edge cases	Defining, finding and using edge cases
Break-out framework	Use cases	Description	Need for a common description of ODD, use cases and services
Panel discussion	Research questions & hypotheses	Future scenarios	Need for standard approach for future scenarios
Break-out Large pilots	Research questions & hypotheses	Future scenarios	How to construct future scenarios (and traffic and driving scenarios)
Introduction	Research questions & hypotheses	Research questions	In automation projects usually a very large number of research questions is of interest. How to define and select them, and how to establish KPIs?
Break-out FESTA	Research questions & hypotheses	Research questions	Priority of research questions
Break-out framework	Research questions & hypotheses	Scale	How to manage many levels of scale in research questions?
Break-out Large pilots	Performance indicators	Shared safety indicators	Largely agreed safety indicators (how to go from KPIs to safety impact)

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 824251

HEADSTART	Performance indicators	Modelling Communication Positioning	How to create a logical layer for communication/ positioning?
Levitate	Study design	Human behaviour	Common assumptions about driving characteristics based on human driver models
Introduction	Study design	User experience	Automation projects may work with systems still under development and severe restrictions on public road operation. How to set-up a study design that provides a rich user experience and test different scenarios?
L3Pilot	Study design	Acceptance User behaviour Generalisation	How to assess acceptance and mobility impacts if pilots are conducted with safety drivers? Use of passenger seat participants, supplementing data with surveys and focus groups
Introduction	Study design	Baseline	What should the results from a FOT be compared with? How to deal with the differences between human and computer-based driving behaviours?
SAM	Study design	Baseline	Lack of baseline for service evaluation, acceptance, etc..
Break-out FESTA	Study design	Study design	Baseline question: What do we compare the outcomes with?
HEADSTART	Study design	Verification Cybersecurity	How to assess cybersecurity to provide enough assurance level?
SAM	Study design	Scale	Scale of experimentation in terms of catchment area, technical maturity, size of the fleet, target users
Break-out Large pilots	Study design	Scale	Upscaling
Break-out Small and national	Study design	Scalable methodology	Methodology adaptable to different test scales
Break-out framework	Study design	Reference	For new services, there is nothing to compare the pilot results with. What is better?
Break-out framework	Ethical and legal issues	Diversity	Need for an innovation friendly framework for running pilots, FOTs and operation.
Break-out framework	Ethical and legal issues	Roadmap	Need for common pathway for policy makers across EU
Break-out FESTA	Data acquisition	Test permission	How to handle tension between desired testing and allowed testing?
Introduction	Data acquisition	Guidance on public road permission	To perform FOTs on public roads many conditions must be

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 824251

			fulfilled to get permission from road authorities and ethics commissions to ensure safety, and to protect data. What guidance is needed on this?
SAM	Data acquisition	Data management	Definition of data flow architecture, data models and formats..
Break-out Simulation and data	Data acquisition	Data format	Share a common log data format for new projects
SAM	Data acquisition	Data sharing	Lack of data (operational data, willingness to share data, open data..)
L3Pilot	Data acquisition	Privacy	Principles for merging data from multiple pilot sites without compromising the privacy of OEMs. Use of 'driving scenarios'
Break-out Small and national	Data acquisition	Data sharing	Access to sensitive data (data sharing) or guidance how to arrange this with industry. Agreed principles for data sharing among industry and research
Break-out Simulation and data	Data acquisition	Privacy Video	Video sharing is difficult between project partners and GDPR doesn't help
L3Pilot	Data acquisition	Urban traffic	Good overview of traffic environments in urban environments
Break-out Framework	Data acquisition	Representativeness	The data collected needs to represent the current and future ODD. How to determine coverage?
Break-out Framework	Data acquisition	Coordination	How to manage many levels of scale in data collection?
Introduction	Data analysis	Simulation models	To arrive at socio-impact assessment results, simulations and data such as accident data, are needed. How to deal with lack of appropriate simulations?
Introduction	Impact assessment	Incomplete data	To arrive at socio-impact assessment results, simulations and data such as accident data, are needed. How to deal with insufficient data?
HEADSTART	Data analysis	Format	Scenario databases have no standardized output format
L3Pilot	Data analysis	AD modelling	Detailed modelling of AD functions in commercial simulation tools
Levitate	Data analysis	Simulation	Impacts on traffic microsimulation

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 824251

Break-out Simulation and data	Data analysis	Simulation AD in traffic	Simulate automated mobility in the context of urban planning
Break-out Simulation and data	Data analysis	Simulation Safety behaviour	Accurate integration of safety strategies in existing traffic simulation software
Break-out FESTA	Data analysis	Data diversity	Handling diversity in sensor types, data sources, data locations, analysis methods. Automating data access and analysis
Levitate	Impact assessment	Share of AV types on the roads	Common assumptions about phase-wise introduction of 'cautious' and 'aggressive' AVs.
Levitate	Impact assessment	Human behaviour	Estimating changes in human behaviour
Levitate	Impact assessment	VRU	Impacts on vulnerable road users
SAM	Impact assessment	Evaluation	Assessment mapping model (data evaluation architecture)
Levitate	Impact assessment	Merged evaluation	Merging of multiple evaluation methods
HEADSTART	Impact assessment	Evaluation	How should the evaluated test results be combined?
L3Pilot	Impact assessment	Evaluation	How to make the best use of many different experimental procedures? Understand all the experiments well, link all experiments to research questions for which they can contribute
Break-out Large pilots	Impact assessment	Evaluation	Combining different impacts together from different methods
SAM	Impact assessment	Scale	Approach to evaluation that is robust for large scale implementation
L3Pilot	Impact assessment	Representativeness	Gap between prototype testing and the scenarios in impact assessment. Concept of 'Mature ADF'
L3Pilot	Impact assessment	Future scenarios Generalisation	How to predict future for the socio-economic impact assessment? Possibly use snap-shot approach to assess only impact of automated driving, No prediction needed
L3Pilot	Socio-economic cost-benefit analysis	Future scenarios	Prediction of alternative futures to go beyond snap-shot approach in the socio-economic impact assessment
L3Pilot	Impact assessment	Accident evaluation	Methods for evaluating impacts on certain accident types

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 824251

Break-out Small and national	Impact assessment	Evaluation	Evaluation of safety benefits of AD and other socio-economic issues
Break-out FESTA	Impact assessment	Verification	How to evaluate AI processes and decisions?

As can be seen, there is a fair communality between the various sources of gaps.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 824251

4.1.Resulting gaps

The analysis above leads to the following long list of gaps for the Common Evaluation Methodology:

1. Implementation plan
   1.1. Framework for coordinating FOTs with multiple locations, OEMs, countries
2. Function identification and description
   2.1. Common description method (ontology, terminology, format) for ODD, use cases and services (and secondly requirements, vehicles, functions, accidents)
3. Use cases
   3.1. Common source for describing ODD in terms of driving behaviour, accidents, scenarios and edge cases
   3.2. Method to define, find and use edge cases
4. Research questions and hypotheses
   4.1. Method to define and prioritise research questions
   4.2. Method to define future scenarios
5. Performance indicators
   5.1. Common set of safety indicators with known relation to safety impact
   5.2. Accepted set of indicators or model for communication and positioning
6. Study design
   6.1. Approaches for achieving a realistic and rich user experience with prototype vehicles
   6.2. Method to compare human and automated driving
   6.3. Reference to compare new services to
   6.4. Method to define and measure a clear baseline for a FOT impact assessment. What is better?
   6.5. Shared assumptions on human driving or shared human driving models
   6.6. Method to validate cybersecurity
   6.7. Method to balance scale of experiment versus generalisation acceptable in impact assessment
   6.8. Methodology that can be scaled down for small projects or handles multiple scales of research questions
7. Ethical and legal issues
   7.1. Need for an innovation friendly framework for running pilots, FOTs and operation
8. Data acquisition
   8.1. Data sharing: approaches to handling lack of data and lack of willingness to share
   8.2. Guidelines for efficient and effective process to public road test permission
   8.3. Common solutions for data management (release, flow, models, formats)
   8.4. Agreed principles for data sharing among industry and research, respecting industrial sensitivity of the data
   8.5. Practical solutions for GDPR-compatible handling of video
   8.6. Overview of urban traffic environments
9. Data analysis
   9.1. Lack of accident data for socio-economic impact assessment
   9.2. Approaches to simulate at multiple levels of detail (sensor, AD function, vehicle, traffic, city), including effect of safety strategies at vehicle level on traffic level
   9.3. Standardisation of modelling of scenarios and AD functions in simulations
   9.4. Handling diversity in sensors, data sources, locations, formats – preferably in an automated process

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 824251

10. Impact assessment
    - 10.1. Shared framework to come from KPIs to assessment (data evaluation architecture, combining various test results)
    - 10.2. Shared assumptions to be used in impact assessment and generalisation
    - 10.3. Shared assumptions on changing human behaviour with higher share of CCAM
    - 10.4. Shared assumptions to estimate impact on VRUs
    - 10.5. Accepted method to come from test results with prototypes to impact assessment for mature CCAM and full scale
    - 10.6. Shared future scenarios for generalisation of impact assessment
    - 10.7. Methods for evaluating impacts on certain accident types
    - 10.8. Methods to evaluate AI processes and decisions

The FESTA stages of study design, data acquisition and impact assessment appear prominent in the results.

There is a strong relationship with specific CCAM R&I actions, such as Test Data Exchange Framework and European Framework for Testing on Public Roads. ARCADE will continue work on Evaluation Methodology in Task 4.3 and Data sharing in Task 4.4.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 824251

5. Conclusion

On 23 November 2020, ARCADE arranged a workshop with between 70 and 80 participants on Common Evaluation Methodology as defined in the CCAM Strategic Research and Innovation Agenda.

In a plenary introduction, project presentation, panel discussion and five break-out sessions, a large number of gaps were identified and discussed. The need for a Common Evaluation Methodology was clearly confirmed.

The major gaps were analysed after the workshop using the FESTA V. The FESTA stages of study design, data acquisition and impact assessment were prominent in the gaps.

Some examples of the prominent gaps from the full list section 4.1 are:

Study design

a) Approaches for achieving a realistic and rich user experience with prototype vehicles
b) Method to define and measure a clear baseline for a FOT impact assessment.
c) Method to balance scale of experiment versus generalisation acceptable in impact assessment

Data acquisition

a) Guidelines for efficient and effective process to public road test permission
b) Common solutions for data management (release, flow, models, formats)
c) Agreed principles for data sharing among industry and research, respecting industrial sensitivity of the data

Impact assessment

a) Shared framework to come from KPIs to assessment
b) Shared assumptions on changing human behaviour with higher share of CCAM
c) Shared future scenarios for generalisation of impact assessment

The results will be taken up by the CCAM Platform and the ARCADE task on Common Evaluation Methodology.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 824251

6. References

1. Strategic Research and Innovation Agenda (SRIA) V1.0. Proposed European Partnership under Horizon Europe, February 11, 2020.
2. EU Levitate project, https://levitate-project.eu/
3. EU Headstart project, https://www.headstart-project.eu/
4. SAM project, France, as introduced in https://www.ville-rail-transports.com/wp-content/uploads/2019/04/developpement-vehicules-autonomes.pdf
5. L3Pilot project, https://www.l3pilot.eu/
6. Digibus Austria project, https://www.digibus.at/en/
7. EU SHOW project, https://show-project.eu/
8. Padlet dashboard, https://padlet.com/dashboard
9. Sli.do online Q&A and polling tool, https://www.sli.do/

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 824251