

Summary of the inventory of existing methodologies and best practices

The inventory has brought many aspects, either best practices (things that have turned out different from planned, and things to avoid) to light that are experienced in the field of CCAM evaluation. The inventory highlights the need for a common evaluation framework that provides guidance, structure, and best practices (from the field). Furthermore, the inventory provides additional input and context to the chapters in the CEM. Below presented are the main findings.

Main findings from the inventory

From several interviews, many have indicated that many evaluations are very (or too) ambitious given the proposed timelines and/or budgets. This likely has to do with getting the project granted by the European Commission (EC). Advice would be to create more realistic project plans, and more critical reviews by the EC could potentially avoid projects that overpromise considering budget/time. One of the consequences is that projects become increasingly less flexible when encountering setbacks. If the project process is delayed, any deviations from the original project plan become increasingly difficult to execute. Any delays may cause the projects to run out of budget or time to fulfil the original evaluation plan and deliverable. Qualitative results can be an alternative when some quantitative evaluations prove not feasible (whatever the reason).

Several interviews and the FAME workshop indicate that alignment between the evaluation team and the pilot sites is not always there, although is crucial. Several questions were raised such as what sites must evaluate, what methods and sensors they could use and what (K)PIs they had to measure. Pilot sites have their own ways of working, and according to one interview, providing these sites with the freedom to come up with ways to tackle problems and measure PIs proved beneficial.

Too much control can hamper the process of measuring and/or the quality of the results. According to another interview, the research question is the goal of what we want to answer; the methodology (proposed to the sites) is pragmatic, so it does not constrain the experiment. Though it is important to state that the evaluation team presents the PIs the sites must measure, with (if needed) negotiations about how to measure the PIs. Furthermore, sometimes evaluation teams are not yet ready with an evaluation plan when data work packages are already determining what data should be collected. It is important that the pilot sites are clearly (and in time) notified what PIs they must measure. A frequent dialogue between evaluation team and pilot sites can provide a clear picture of the status of measurements and which PIs can be delivered.

A job for the CEM is to find a balance between PIs desired by the evaluation team and what the sites can deliver. This balance is not easily found, though having frequent contact between the evaluation team and the sites can help. This is something mentioned in interviews, that visiting sites and partners in person benefits the overall process, as does allocating budget to team building. Visiting many sites requires finding a balance with the available project budget.

The template analysis highlights that most evaluations focus on relatively small pilot level such as on street level. Furthermore, the time scales are relatively short. Results of these tests and evaluations must be scaled up to assess effects on national or international scale (see also Aspects not yet covered in the CEM). Other main points are the large and diverse number of direct and indirect impact areas mentioned. The filled-in templates indicate that more direct than indirect impact areas are included. Socio-economic aspects have a relatively low frequency of being included in evaluations, such as health, equity, and labour market.

From the template analysis, in many projects' steps of FESTA are used, either following the FESTA V structure, or in comparable steps while not following FESTA directly. Some have indicated that the entire framework works for them, following every step. Others have indicated that the framework can prove relatively rigid, and they mention that taking several steps from FESTA and using them in a different order works for them. This diverse image of FESTA use is confirmed from the template analysis, from which also can be seen that steps such as ethical and legal issues, study design, and social economic CBA are relatively less often used in evaluations.

The sharing or handling of data between stakeholders is a subject which is named numerous times. A common data format, consolidated databases and collaborative code editing seem to have helped parties in data processing. By doing so, teams can be agile in their ways of working and new data or coding can be implemented directly into a common format.

According to one interview, it was a key enabler to speed up the evaluation process. From interviews it was gathered that difficulties are experienced when data is not easily shared between partners or when data remains absent. Sensitive data (either commercial or personal) forms a barrier to the exchange of information. In the template analysis became apparent that data specification as a process was always there, though in practice (from interviews) seems to be often quite troublesome, indicating that there could be a misbalance between the proposed plan and actual execution of the evaluation process.

Aspects not yet covered in the CEM

The CCAM technology has usually low TRL levels and is relatively immature. This can prove challenging in for instance conducting reliable pilots on the road, simulating traffic interactions, and conducting user evaluations.

A wider socio-economic impact assessment needs a scale-up from pilot level to the requested higher geographical level (for instance city, region, nation) and a scale up towards a higher temporal level (for instance a year). This impact assessment also needs scenarios with mobility services including business models and a baseline (or target value) to compare.

Scaling up is a challenge for evaluations about technologies that have low TRL levels. What experiments on the road do you need to do to get suitable input for simulations for scaling up? Street level alone is not enough to achieve a wider socio-economic viewpoint, so scaling up in some form is always needed. A baseline is needed for among others the research question, hypothesis, target values and comparison between CCAM technologies and non-self-driving vehicles. Methods for scenario development is also mentioned. The same goes for using AI in evaluations. Not a lot of examples exist how AI is used in CCAM evaluations.

Another important subject is the way evaluation methodologies and results are reported. Most of the time they are reported in a precise and technical way, which could turn out challenging to comprehend for non-technical people or people not directly involved in CCAM, such as policy makers. Reporting in a more accessible way may increase the impact of evaluations.

Brussels, 2 November 2023