

Evaluation While Doing: Verfijning van FESTA

Kerry Malone – TNO – Kerry.malone@tno.nl

Eline Jonkers – TNO – eline.jonkers@tno.nl

Han Zwijnenberg – TNO – han.zwijnenberg@tno.nl

**Bijdrage aan het Colloquium Vervoersplanologisch Speurwerk
19 en 20 november 2015, Antwerpen**

Samenvatting

Evaluation While Doing: Verfijning van FESTA

Het voegen van intelligentie toe aan voertuigen en infrastructuur (Intelligente Transport Systemen of ITS) brengt grote beloften voor veiliger, schoner en betrouwbaarder verkeer. Dit wordt erkend door beleidsmakers op allerlei niveaus. Denk bijvoorbeeld aan het ITS-actieplan van de Europese Commissie en nationale plannen zoals het Nederlandse verkeersmanagement beleid zoals beschreven in het 'Beleidskader Benutten' of 'Beter Geïnformeerd op Weg'. Voor beleidsmakers is het belangrijk om te weten op welke manier ITS bijdragen om beleidsdoelstellingen op de doorstroming, de verkeersveiligheid en het milieu, en wat de gevolgen zijn (kosten, business modellen, samenwerking etc) zijn om het besluitvormingsproces te bereiden; overheden maken meestal investeringsbeslissingen op basis van deze aspecten.

ITS deployment wordt versneld, bijvoorbeeld door het Beter Benutten Programma en Praktijk Proef Amsterdam. In korte periodes wordt heel veel neergezet en voortgang geboekt. Ervaring in een aantal projecten in het binnen- en buitenland laat zien dat evaluatie met zijn resultaten van de neergezette dienst relatief laat in het veldtest komt. Het boodschap in dat geval dat de dienst de beloften voor verbetering van de doorstroming, veiligheid om milieu effecten niet bereikt heeft, komt hard aan. Ervaring in veldtesten leert ook dat de voorgestelde dienst betere effecten zouden kunnen bereiken als experts in bijvoorbeeld de verkeerskunde of human factors bij de dienst ontwikkeling betrokken worden. Inzichten vanuit deze expertises op het ontwerp of prototype van de dienst, die verwerkt worden in de volgende fase, kunnen een hogere acceptatie niveau van de dienst in de veldtest realiseren. Een betere dienst versnelt deployment van ITS en leidt tot hogere socio-economische baten-kosten ratio's en betere business cases voor stakeholders.

Tot nut toe zijn er relatief kleinschalige veldtesten geweest, op Praktijk Proef Amsterdam In-car na. Het verzamelen van gegevens, controle en analyse is grotendeels een handmatig proces. Kijkend naar de projecten van de toekomst, waarin de niveaus van 4000, 20.000 en 100.000 deelnemers verwacht worden, een handmatig proces is niet langer haalbaar. Daarom is er behoefte aan een deel-geautomatiseerd proces. Evaluatie outcomes (effecten op bijvoorbeeld veiligheid, verkeer efficiëntie en milieu) zouden nog meer waard zijn als het sneller beschikbaar komt na of tijdens de veldtest, of, beter nog, tijdens de ontwerp- en ontwikkelingsfase van de dienst, waardoor het gebruik van het verkeerskunde en human factors inzichten om de effectiviteit van het vertrouwen in en / of de naleving van de dienstverlening te verbeteren. Daarom wordt in dit document een nieuwe aanpak genaamd "Evaluation While Doing" wordt gepresenteerd. Evaluatie While Doing is gericht op het uitvoeren van de evaluatie eerder in het proces, gericht op meer betrokkenen (bijvoorbeeld ontwikkelaars / industrie en overheid) en op een efficiëntere manier (door het versnellen van het verzamelen van gegevens, verwerking en analyse door middel van automatisering) om de snelheid van het evaluatieproces te verhogen. Door "data", bedoelen we zowel objectieve als subjectieve gegevens.

1. Introduction

1.1 Background

Adding intelligence to vehicles and infrastructure (Intelligent Transport Systems or ITS) brings great promises for safer, cleaner and more reliable traffic. This is recognized by policy makers on all kinds of levels. Think for example about the ITS Action Plan [ITS Action Plan] of the European Commission, and national plans such as the Dutch traffic management policy as described in 'Beleidskader Benutten' [Beter Benutten, 2008] or 'Beter Geïnformeerd op Weg' [Beter Geïnformeerd op Weg, 2013]. For policy makers it is important to know in what way ITS contribute to policy goals on throughput, traffic safety and environment and what the implications are (associated costs, business models, collaboration etc) are in order to prepare the decision making process; governments usually make investment decisions based on these aspects.

When we talk about the evaluation of ITS, we mean the process to determine the effects of an ITS, e.g. an in-car system, a cooperative system, a traffic management measure or a mobility management measure, in simulation, under specific controlled conditions (Controlled Tests) or even in real world circumstances. "Effects" can mean different things for different stakeholders: for a driver the ITS may be about comfort, for a car manufacturer it may be about safety, for a policy maker it may be about traffic efficiency. There can be various reasons for doing evaluation and not all evaluations serve the same purpose; some evaluations focus on determining the quality of the ITS under research and whether goals are being met, other evaluations serve a monitoring function or feedback function (e.g. on the design of the ITS). Evaluation provides stakeholders with insights in the progress of the development process or in reaching certain policy goals and information that takes the deployment process forward. Evaluation can include effects on safety, traffic efficiency, environment, mobility, impacts on an organization as well as investment, operational and maintenance costs that can be used in a CBA or CEA, as required for governmental parties.

Evaluation is an important part of designing, developing and testing innovations. It provides feedback on the performance of the innovation, enabling developers to improve the current design and settings. The scope of the design improvements depends on the phase the development process is in. Feedback in the early design phase most likely allows modifications of a wider scope than, say, feedback on a product that is about to be launched. Currently, evaluation or impact assessment is usually carried out at the end of a project, indicating what the impacts of a measure or measures were on specific indicators. While this is very valuable information, the assessment is available only after the field test or pilot. This lag period can be several months after the data collection, due to the data processing and analysing steps, limiting the usability of the evaluation in the redesign of the product or service.

Evaluation outcomes (impacts on e.g. safety, traffic efficiency and environment) would be even more valuable if it were available sooner after or during the FOT or pilot, or, even better, during the design and development phase of the service, thereby using traffic science and human factors insights to improve the effectiveness of, trust in and /or compliance with the service. Therefore in this document a new approach called

“Evaluation while doing” is presented. Evaluation while doing aims at performing evaluation earlier in the process, directed at more stakeholders involved (e.g. developers/industry and government) and in a more efficient way (by accelerating data collection, processing and analysis through automation) to speed up the evaluation process. By “data”, we mean both objective and subjective data.

1.2 Approach for FOTs and evaluation: FESTA

A methodologically sound approach to setting up and evaluating Field Operational Tests (FOTs) and pilots was developed in the FESTA (Field opERational teSt supportT Action) project and updated in 2014 [FESTA, 2014]. Evaluation includes both technical evaluation and impact assessment in the areas of safety, traffic efficiency, environment, acceptance and socio-economic cost-benefit analysis. Tests can take different forms, as shown in Figure 1: computer simulation, lab, proof-of-concept, field test, pilot. FESTA defined the “FESTA-V” (also called “FOT-chain”) of evaluation steps based on the FESTA methodology. Figure 2 shows the FESTA-V. The left side of the “V” describes the steps taken to prepare the evaluation of the FOT. The right side of the “V” describes the steps taken to carry out the analysis of the FOT. The left and right-hand sides of the V at the same horizontal level are connected to each other. That is, the preparation of the research questions and hypotheses has a direct link with (and impact on) the analysis of the research questions and testing of hypotheses. An important contribution of the FESTA approach is that it made the explicit link between (1) what has to be evaluated and (2) the preparation of the FOT. A consequence of this link is that evaluation issues were brought to the discussion earlier in the execution of the FOT, in theory guaranteeing that the data necessary for impact assessment is collected during the FOT.

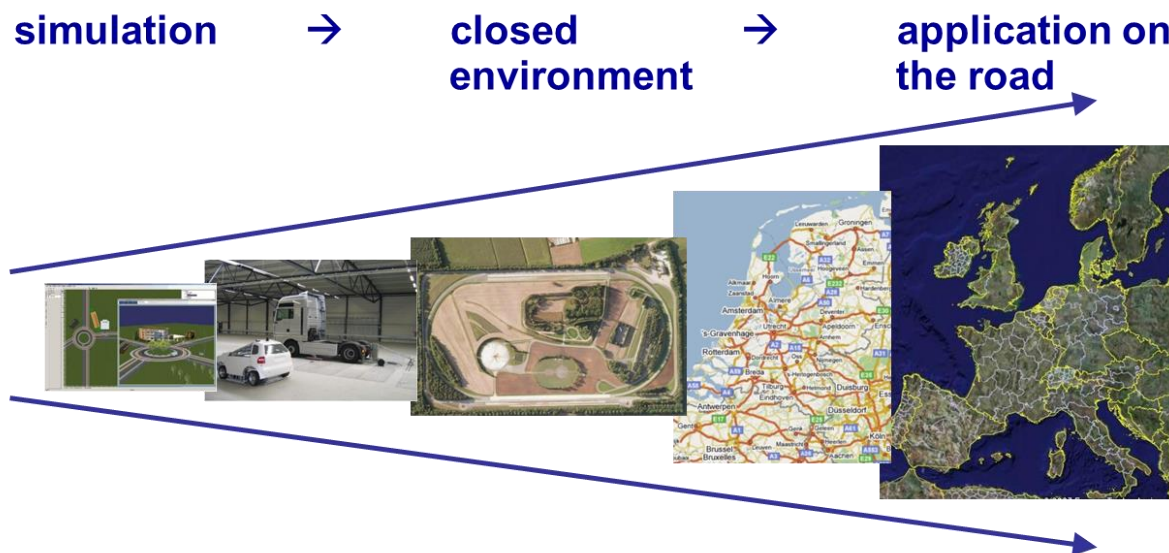


Figure 1: Testing and research environment for ITS

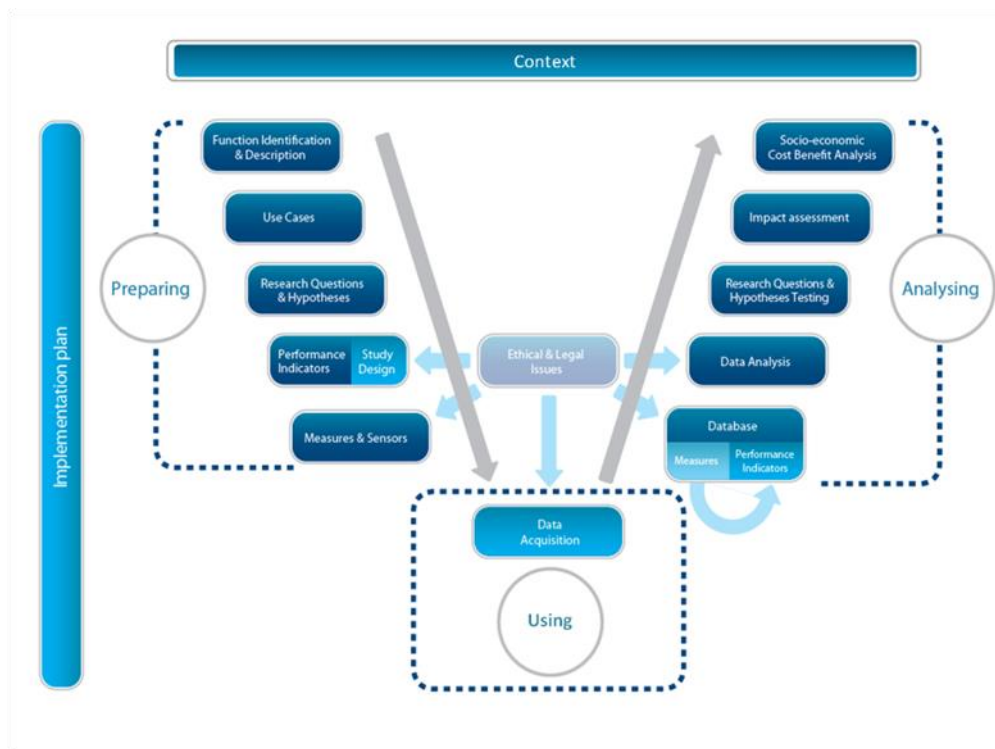


Figure 2: The FOT-chain for evaluation based on the FESTA methodology.

Since FESTA was developed, several FOTs have taken place at the European and national levels. These include euroFOT, TeleFOT, DRIVE C2X, FREILOT, simTD, SPITS and the PPA. Lessons learned from these large-scale efforts triggered the revisitation of FESTA.

The FOT-Net 2¹ Working Group on Impact Assessment and scaling up has collected feedback from already completed and running FOTs on evaluation issues [FOT-Net 2, 2014]. Some issues are specifically relevant for evaluation while doing. An important one is timing: impact assessment is at the end of a FOT, and it deals with all the problems and delays caused in other parts of the project. Often, the time available for data analysis and impact assessment is reduced compared to the original timetable, due to project delays. All FOT projects said they had too little time in the end to carry out the planned evaluation. An experience from already completed FOTs is that there is a time-lag between the completion of the FOT operations and the impact assessment, due to the data collection, processing, enriching and analysis steps. Experience also has taught us that if services were modified (slightly), a significantly larger positive impact could be achieved. These modifications come from insights or modelling in the areas of human behaviour, traffic science, environmental impacts, etc. This feedback would be much more valuable in the design stage of services, when it is less expensive and quicker to modify a design, rather than 2-3 years later after the completion of the FOT and evaluation. This was also an outcome of the FOT-Net 2 Working Group: feedback is very important. In the Working Group feedback from piloting was specifically mentioned, but also other feedback loops were discussed.

The increased use of apps and smartphone platforms as data sensors drive the data collection process and offer possibilities for evaluation. Enormous amounts of data are

¹ The EC projects FOT-Net and FOT-Net2 projects updated the FESTA methodology based on FOT experiences.

being collected. To select, filter and aggregate the data efficiently at (or close to) the source in order to learn, does require a more automated approach. The use of smartphones also enables research to reach out to the user to collect user feedback more quickly. These technologies can be used to solicit precise feedback on user acceptance and usefulness of the services based on specific events and their situational variables, in order to improve the service.

In the application development, start-ups and SME's try to implement innovations or test new ideas. These companies simply do not have a horizon of 3-4 years to look ahead. Every update is a new launch of their service and is evaluated by the customers as if it was a new product. The rate of improvements is more important than the perfection of the system. This is triggered by principles as software defined functionality, over-the-air-configuration, etc.

This paper begins by introducing connected and cooperative ITS, followed by the a discussion of the different stages in the FESTA-V that EWD is applied. The paper concludes with the implications of EWD and some reflections on its use.

2. Connected and cooperative ITS

This paper focusses on the application of evaluation while doing on connected and cooperative intelligent transport systems, called "C-ITS"², although the approach is in principle applicable to the evaluation of all types of ITS and related measures. The nature of C-ITS is that it requires different stakeholders to provide the C-ITS service. Examples include the ITS corridor services road works warning and probe vehicle data and the systems identified as Day-1 services by the Platform for Deployment of C-ITS, for example, road works warning, in-vehicle speed limits, intersection safety and traffic signal priority. In these examples, stakeholders directly involved in deployment include OEMs, telecommunication providers, road operators and cities. Evaluation provides valuable information for the service being developed, as well as data needed for deployment.

In evaluation, understanding how users use new C-ITS is crucial. There is some information available about how and under which circumstances users use C-ITS from European and national projects such as DRIVE C2X, eCoMove, simTD, etc., but this is limited. More feedback on end-user acceptance, trust and compliance and on the situations that determine this behaviour is crucial for the effectiveness of these systems as well as to commercial parties to understand better what users want. Because specific costs and benefits of C-ITS accrue to specific stakeholders, C-ITS stakeholder-specific costs and benefits are needed by the parties in order to develop their business cases.

² "Cooperative Intelligent Transport Systems" are Intelligent Transport Systems (ITS) that make use of communication to exchange information. Communication can take place between vehicles (V2V), vehicles and infrastructure (V2I) and between infrastructures (I2I). Sometimes there is V2D communication between a vehicle and a device, and V2IP communication between vehicles and the internet. Cooperative systems implies a certain level of cooperation or negotiation between vehicles or between vehicles and the infrastructure and/or a traffic management center or back-office. The American term "Connected vehicles" means the same as the European term "cooperative systems", except that connected can also mean that there is an internet connection in the vehicle. translation of definition in [TrafficQuest, 2014]

Ultimately, a choice of services and technology, plus scenario variables such as penetration rates, leads to benefits and costs. This information feeds choices that need to be made about organization and financing of the service and the stakeholders involved, who then need to examine their own business cases. Figure 3 shows evaluation in this context.

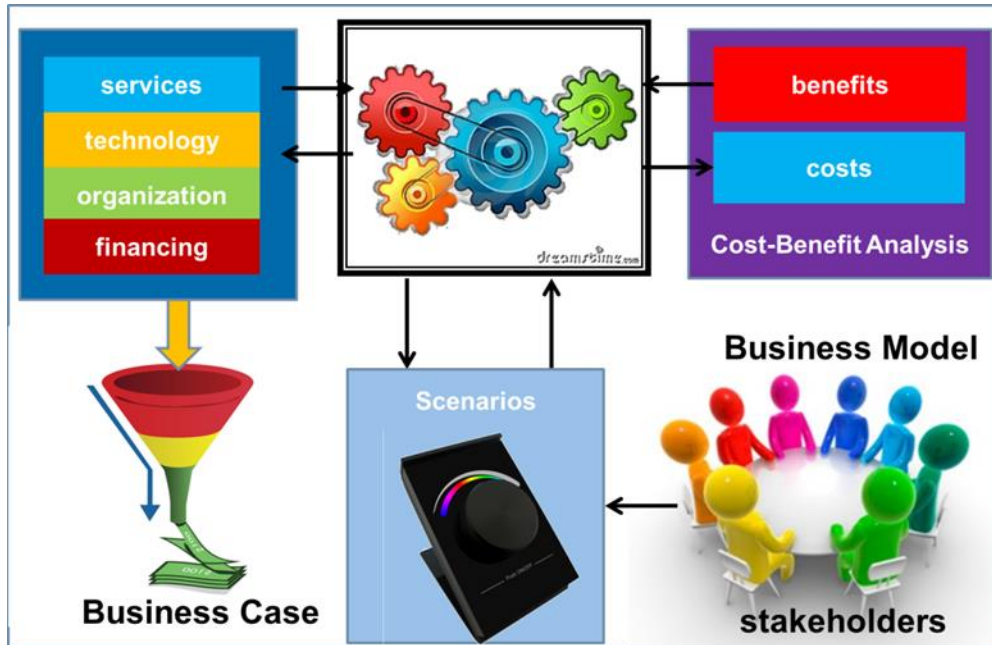


Figure 3 Context of Evaluation

3. The concept of Evaluation While Doing

Section 1.1 explains what we mean by “Evaluation while doing”. To summarize: a refined FESTA approach to evaluation, called “Evaluation while doing”, is needed to

- Increase C-ITS effectiveness and acceptance in the design, prototype and roll-out phases
- Accelerate the C-ITS development and deployment cycle
- Provide stakeholders with information they need for Business Case development

“Evaluation While Doing” (EWD) translates these needed improvements into the following proposed improvements in the design, prototype and roll-out phases:

- Carry out ex-ante impact assessment early in the FOT cycle: a better design and/or prototype with more impact developed early in the design and development process will produce a better product quicker and cheaper.
- Automation of the data collection, checking, processing, repair and enrichment process. Automating these processes will make the analysis of large FOTs and pilots more feasible. These processes are currently set up in an ad hoc manner, requiring manual data manipulation at some stages. For FOTs involving thousands of vehicles, such a labour-intensive process is not feasible. Automation of this process provides the secondary benefit of monitoring key aspects of the FOT while it is running, thereby measuring progress towards achieving the goals of the FOT.

- Use an iterative approach to design and deploy services, integrating feedback to produce improved versions of services.

The application of Evaluation While Doing (EWD) results in a modified FESTA-V diagram. The FESTA-V process consists of a single application of the series of steps from preparation on the left side of the "V" through to evaluation on the upper-right hand side of the "V". Evaluation while doing envisages a more interactive process where evaluation is carried out earlier in the process and in a more efficient way. Different 'types' of evaluation can be carried out, each with a different purpose and timing; however all evaluations give feedback on the ITS under investigation. The concept as written down in this report is flexible. It does not prescribe how many feedback moments there should be and what they are. The sections below describe the possible feedback moments and their outcomes in a logical order.

3.1 Feedback on estimated impacts

Feedback on estimated impacts is also called ex-ante impact assessment and is in fact a (quick) impact assessment at the beginning of a project. The circle labelled "1" in Figure 4 shows that this evaluation takes place high in the left-hand part of the "v"—early in the project. Through collaboration with the developer, the impacts of a system can be estimated in a modelling approach. Expert knowledge of human factors and traffic and environmental science, and tools available at TNO and other parties, such as the ITS Modellers, VERSIT+ and SCENIC, can be used to assess the traffic and environmental effects of the proposed system, for example based on a functional description. The goal of this step is to provide developers information on the estimated impact of their system (which they can for example use for further development or in search of investments or collaboration partners), and to provide decision makers or policy makers information on which they can base their decisions (e.g. to make a shortlist of solutions that can help them achieve their policy goals).

This evaluation step early in the project is related to feedback on design, as described in the next subsection.

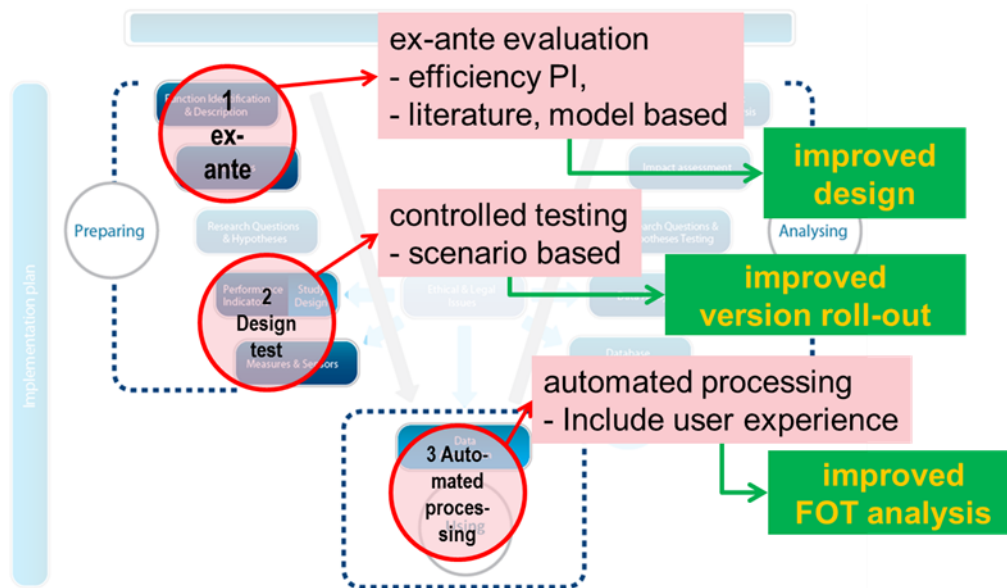


Figure 4: FESTA-V, revisited

3.2 Feedback on Design

As the previous step, feedback on design is given early in the project (in the phase where the system is designed). A (quick) impact assessment is done in collaboration with the service developer. The goal is to provide service developers with information on parameters that can be modified to achieve greater positive effects as well as higher acceptance of, usage of and compliance to the service. The intention of the timing of this step is to provide this feedback early enough in the process so that the service developer can make use of this evaluation results in order to modify the service efficiently.

Feedback on estimated impacts and feedback on design can in the same evaluation step, early in the project.

3.3 Feedback on prototype and data processes

Feedback on the prototype and data processes is aimed at the stage of the first prototype of the service(s). The circle labelled "2" in Figure 4 shows where this feedback takes place in the FESTA-V. Here, controlled testing (CT) aims to carry out technical testing and collect data for impact assessment in the field, although not necessarily on public roads. Pre-testing in a (low-cost) driving simulator could be used to test understandability, acceptance and to verify and fine-tune functionality in a controlled environment. The impact assessment provides feedback to service developers on parameters that can be modified at this stage of development, and possibly on the impact of the service (e.g. on traffic, safety and environment) – this depends on how extensive the evaluation of the prototype is. It also tests the automatic data collection, other data 'tools' (e.g. for checking, cleaning and analysing data) and impact assessment 'process'. Feedback is provided on these processes and tools as developed and this leads to an improved processes, automated as much as possible. Automation of a process is a costly operation and of itself. In cases of small field tests, automation may not be deemed necessary or feasible, unless the tools are readily available.

3.4 Deployment evaluation

The deployment evaluation or impact assessment is connected to the stage at which roll-out / deployment takes place. Impact assessment does not focus any more on providing feedback on functioning of the service (e.g. modification of parameters), although some fine-tuning might be involved, but at delivering realistic impacts, since they are based on tests in the real world. Use can be made of the fact that the (automated) data chain and impact assessment processes were already tested and improved, resulting in an accelerated evaluation process.

This last step can be repeated for every soft or hardware generation. The automated data processing quickens the evaluation process drastically and ensures timely results. Most likely, only the last development cycle following the FESTA-V results will be taken up to national or EU levels, and CBA/CEA calculations.

Evaluation without the advantages of efficient data collection and tools can still be carried out as described in the existing FESTA methodology, with the necessary modifications, under circumstances such as:

- Evaluators are engaged relatively late in the project;
- Data needs are not determined with input from the evaluators, resulting in problems such as the lack of a baseline or data inavailability;
- The field test is small.

The first two of these situations are not uncommon situations, based on experience with field tests.

3.5 Benefits of evaluation while doing

With evaluation while doing and using the feedback moments as described above, evaluation results can be gained quicker and in a more focused and efficient way. This in turn will help improve the system under investigation with respect to design, performance and impact.

Carrying out the different evaluation steps provides some benefits with regard to efficiency, because part of the work can be reused. If feedback on the data processes is provided, the deployment evaluation can make use of the improved (automated) data chain and impact assessment processes, which means a more efficient process. When feedback on estimated impacts is provided by carrying out ex-ante impact assessment, insight into the working of the system or service is improved, making later analyses more efficient.

In 'classic' field tests (with one impact assessment at the end: deployment evaluation) there are sometimes problems for the evaluation team to get the right data in the right format. It is difficult or impossible to influence choices made early in the project with respect to data being logged, test set-up, etc. Sometimes the evaluation team has not even started working in the project when these choices are being made. Another issue is that it can be difficult to predict fully what the impacts will be and what data are needed exactly (a request for data might come up during analysis). In the EWD approach the

evaluation team is involved in the work earlier and this has several advantages, one of them being that working with the data in an earlier stage enables the team to give better and earlier feedback on whether the data collected is what the evaluation team needs. It is even an option to – during the ex-ante impact assessment – wait and see what data are collected, and then give feedback and talk to the right people about what other data wishes there are. The FESTA approach does promote involvement of evaluation from early in the field test.

4. What does EWD imply?

EWD is a refinement of FESTA: the FESTA-V forms the basis of EWD. Extra activities take place in EWD at certain stages in the FESTA-V, described already in section 3. The sections below describe the process and some additional details of the application of EWD.

In the Feedback on estimated impacts and on design, the service designers and developers interact with the experts in human behaviour and impact assessment. The interaction can be based on discussions, research, and modelling, with as goal that the experts in design and technical aspects can make improvements to the original design at a relatively low cost. EWD refines FESTA in the process of application in this step by introducing a feedback loop to the design.

Feedback on prototype and data processes involves small-scale piloting of the prototype, data collection and data analysis. This step has as its goals technical testing, which is normal in the application of FESTA, and the collection of data for human behaviour and impact assessment analysis. The second goal aims to check whether the system is used as intended, the information and warnings are presented in a useful, satisfactory and understandable manner, and to check whether the (projected) impacts of the system as designed can be realized. Feedback on the prototype performance can be used to improve the subsequent version. EWD refines FESTA by introducing a feedback loop to the prototype design.

Deployment evaluation, as described here, aims to accelerate the impact assessment process and also monitor the field test by automating the data chain. Realization of the data chain automation requires methodological, algorithmic and tool development. The automation is foreseen to go up to and including some hypothesis testing. A conceptual model for the data collection and data flow has been developed, making use of tools developed in other projects. A re-usable framework for generic and field-test-specific data acquisition, monitoring and evaluation, FAME, was established and used in a demo. Ultimately, when clean, checking and repair of data and the production of Performance indicators can be automated, impact assessment and monitoring can be very efficiently carried out.

5. Additional thoughts

The foundation for EWD is the FESTA methodology. If EWD is to succeed, then the FESTA approach is followed. For many field tests, the FESTA methodology is applied to selected parts of the project. Often what is missing is the integration of the parts: the left and right sides of the "FESTA-V". Without the integrated application of FESTA, it is not worth applying EWD.

The FESTA methodology was developed at the beginning of the seventh European Union Framework program for DG-Information Society and Media (2008). At that time, several large field tests were about to be launched in later calls of the seventh European Union Framework program. FESTA provided a common approach for design, carrying out and evaluating Advanced Driver Assistance Systems (ADAS) and In-Vehicle Information Systems (IVIS) for vehicles. The FESTA methodology was applied in those field tests. A Dutch version of the FESTA methodology, "Leidraad evaluaties benutting (versie 2011)" [Wilmink et. al, 2011]. The FOT-Net and FOT-Net 2 projects updated the FESTA methodology based on lessons learned.

Since then, connected and cooperative technologies also been developed. Large scale pilots have taken place, while plans for roll-out are becoming reality. Navigation with real-time traffic information, Electronic Stability Control and other safety and eco-driving systems are being taken up by the market. However, these systems are still a small percentage of the total vehicle fleet (with the exception of navigation systems with real-time information). To assess the impacts of these systems on important Key Performance Indicators (KPIs) such as traffic efficiency, safety and environment, other sources of data need to be collected, in addition to the data from traditional sources such as loops. These other sources include in-vehicle logging, video data collection, road-side units, and data on weather conditions, traffic conditions, etc.

The *integrated* FESTA approach, which remains unchanged in EWD, is that evaluation needs need to be taken into account from the beginning of the project. Decisions about logging modules and data to be collected and its quality, are often decided at the beginning of the project by the party that designs and implements the service. Otherwise, it is too late. Thus, evaluators and developers need to communicate about data to be collected at the very beginning of the project. Experience shows that this very basic approach of FESTA, much less EWD, does not consistently take place. This can be due to the manner in which the project has been tendered (separate parcels for development and implementation and evaluation) or that there are no guidelines or requirements for the data to collect or the KPIs to be generated. Thus, good evaluation starts with an integrated FESTA application. Only then can EWD be of use.

Acknowledgments

This paper was made possible by SMO funding in het TNO Knowledge investment program Mobility and Logistics.

References

Beleidskader Benutten: Eén van de pijlers voor een betere bereikbaarheid, Ministerie van Verkeer en Waterstaat, 4 januari, 2008.

Beter Geïnformeerd op weg: Routekaart 2013-2023, Samenvatting,
<https://www.rijksoverheid.nl/documenten/rapporten/2013/11/04/beter-geinformeerd-op-weg-routekaart-2013-2023-samenvatting>.

FESTA, 2014,
http://fot-net.eu/wp-content/uploads/sites/7/2014/06/FESTA_HBK_rev5_ref3-2.pdf

FOT-Net 2, 2014, FOT-Net 2 D3.2: Working Group Additions to FOT Methodology.

ITS Action plan, "Action Plan for the Deployment of Intelligent Transport Systems in Europe", Communication from the Commission, COM(2008) 886 final,
http://ec.europa.eu/transport/themes/its/road/action_plan/, 16 December 2008.

TrafficQuest, 2014, Coöperatieve systemen & Automatisch rijden: State-of-the-Art Achtergronddocument, 12 November 2014.

Wilmink, I., K. Malone, E. Jonkers, R. Brouwer, R. de Lange, M. Keuken, A. Eisses, N. Rosmuller, J. Mak, "Leidraad evaluaties benutting (versie 2011)", 30 mei 2011.