Project Proposal Summary Sheet

Project Number	P_2025_04
Project name	ASAM QSQ – Quantifying Simulation Quality
Domain	Simulation
Impacted standard(s)	ASAM OpenX
Project type	Concept
Start date	Dec 01, 2025
End date	Apr 30, 2027
TSC Submission:	Oct 31, 2025
Proposer(s)	ASAM e.V.
ASAM Office Responsible (OR)	Diego Sánchez, Yash Shah
Initiating Companies	Hochschule Düsseldorf University of Applied Sciences Ansys Inc. Persival GmbH
ASAM funds	10k EUR (Concept Phase) and 25k EUR (Standard Phase)
Backwards Compatibility	*NA*

For more information on the ASAM project process and the proposal phase in particular, please refer to the ASAM Project Guide.

1 Executive Summary

The path to virtual homologation depends on the credibility of the virtual methods and tools involved. This refers to explicit credibility of the simulation process but also to credibility implicitly derived from the quality of the simulation models, frameworks etc.

Several initiatives (e.g., prostep IVIP, INTACS) and research projects (e.g., SetLEVEL, EVIDENT, HAIViSH) focus on credibility aspects as these seem to be of high importance in the discussion with authorities. The momentum in this area is rather strong.

As of today, no standardized metrics seem to exist for the assessment of the quality of simulation frameworks and simulation models, especially in the area of sensor simulation. An exception are vehicle dynamics models whose quality criteria have been described in various publications, e.g. in [1].

The parties involved in this proposal aim to begin with a concept phase that will initiate a new standardization project focused on defining a standardized data format for storing and exchanging quality metrics for simulation frameworks and models. They also aim for the definition of agreed, use-case specific weighting of the respective metrics in order to emphasize features according to their relevance for fulfilling specific tasks.

The project shall start with the identification or definition of a taxonomy, provide one generalized and two scope-specific metrics definitions (for simulation frameworks and models, respectively), exemplary KPIs for frameworks and models as starting point, and a Proof of Concept (PoC) that qualifies tools and models of participating parties along the defined criteria.

The results achieved within the PoC shall neither provide an assessment of market offerings nor indicate any preference thereof. It shall only prove the applicability of the results of the standardization effort.
Upon completing the data format definition, taxonomy, and metric development, and validating the concept through the PoC, the participating parties intend to transition directly to a formal standardization phase to continue development. The parties shall report to the TSC upon completion of the concept phase and propose a standardization phase within the remaining timeline of the project. This phase may not start without prior approval by the TSC.

The project shall help emphasize the role that compliance with existing standards of ASAM and other organizations (e.g. Modelica) plays for the quality of simulation frameworks and models. A key aspect is the exchangeability of data, libraries, and other assets between tools.

Based on experience from the concept project, and from previous research projects (e.g. SETLevel) and existing partnerships (e.g. with prostep ivip), a route shall be shown to come from a quality argumentation to a credibility assessment.

Users of the new standard may be tool vendors describing their offerings, tool users requesting specific properties in their RfQs, and authorities defining minimum criteria that tools and models have to fulfill in order to be acceptable for a given (homologation) task.

2 Overview / Goals

2.1 Motivation

Regulatory bodies are increasingly adopting virtual methods as standardized procedures for the validation and verification of automated driving systems (ADAS/AD). Ensuring the credibility of solutions that virtualize components of these systems is essential, particularly given the critical nature of the tasks involved. A central piece in the argumentation for credibility is the objective quantification of a solution's or individual components' quality through a harmonized set of quality criteria.

The initiators of this proposal seek to address this challenge by defining a machine-readable data format that formalizes the quality assessment of individual components within virtual solutions. Establishing a consensus across the industry, expressed through a common, standardized description, will be instrumental in introducing objectivity and, ultimately, credibility into the evaluation process of virtual artifacts.

2.2 Preparatory work

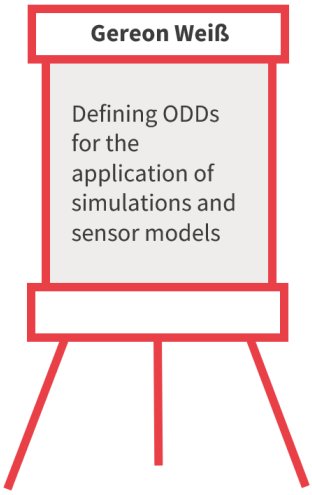
The co-initiating parties of this proposal, Hochschule Düsseldorf (HSD) and MD Engineering Services, held a first workshop "Quantifying Simulation Quality" in 2021 [2]. Preliminary solutions for assessing the quality of simulation artefacts were also developed and applied in the project SimCert [3] by MD Engineering Services.

HSD has been an active member in an IEEE initiative to quantify the quality of (real) image sensors in the automotive domain [4]. Several scientific publications proposed and demonstrated a novel approach "Simulating tests to test simulation" and its role in homologation [5, 6]

This new ASAM proposal was kicked off in an ideation workshop on Sep 10-11, 2024, jointly organized by HSD and ASAM. Under the title "Quantifying Simulation Quality" key players of the industry met to discuss aspects of simulation quality in general and way of describing it in a generalized, commonly applicable format [7].

Several workshops were held and questions were addressed:

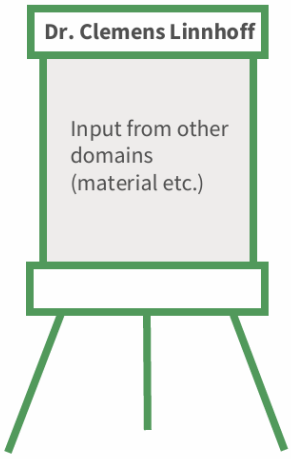
How it works



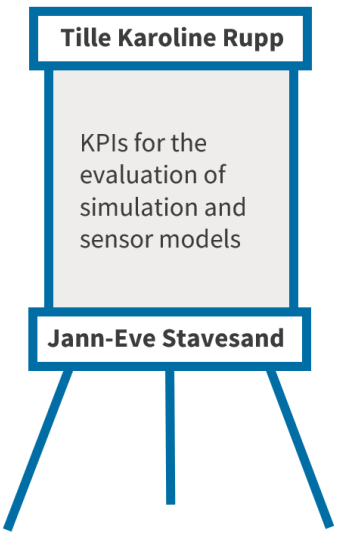
QUESTION	ISSUES	IDEAS & SOLUTIONS	KEY TAKE AWAYS
What is needed with respect to an ODD for simulations and sensor models?	- Detail of ODD - Is ODD same as validity domain? - Format/syntax is not enough - Who should define standard ODDs?	- New taxonomy for OD of sensors & sensor models - Standard / baseline for ODDs to be covered	- Machine-readable ODD for automation - Standardized ODDs - ODD taxonomy for sensors & sensor models
What should an ODD for simulations and sensor models describe?	- Needed content of ODDs is application / use case specific; e.g., ODD for vehicle vs. Function - Limited computing power / time	Levels of abstractions from system, over perception to physics - Consider coverage and transitions of ODDs	- ODD associated with item, e.g. sensor (model), defines if it can me used in a setting or not - ODD contracts/certificates
What could be usage scenarios of an ODD definition for simulations and sensor models?	- Testing and testing strategy - specification and requirements - ML training - simulations	- Detection of ODD boundaries - ODD-coverage testing - ODD-based simulation	- Digital twin ODD for predicting safe and unsafe states - ODD-driven development automation

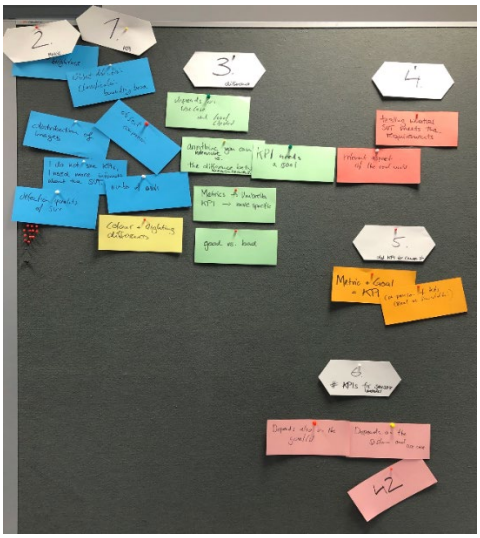
The workshops

How it works



QUESTION	ISSUES	IDEAS & SOLUTIONS	3 KEY TAKE AWAYS
What are the dependencies of simulation and sensor models along the data pipeline?	- Quality assessment of the downstream function is also hard - A lot of different functions (Object Detection/Tracking/Classification, Lane Keeping, Occupancy Grid, etc.)	- Take a specific downstream function to assess simulation quality - Deduce requirements for the sensor model from the sensitivity of the downstream function	1) Requirements on sensor simulation are very specific to the downstream function 2) Benefit for assessment of downstream function: Simulation always provides ground truth
Which quality criteria are applied for material and geometry descriptions?	- Different level of detail needed. - A lot of materials - Environment changes over PLC (e.g. new car models on the roads)	Look at Level of detail from sensor technology -> Camera needs different LODs than Radar	1) Materials are wavelength dependent 2) Different sensor modalities need different levels of detail in the environment 3) There are too many materials in the real world which cannot all be measured. There must be some simplification or interpolation,
What can we learn from the quality metrics of real sensors?	- Timing - Production tolerances - Interfaces - Lab performance vs. real world	Use sensor calibration metrics	1) There are a lot of requirements on real sensors that can be compared to the simulation 2) Calibration metrics for real sensors can be used to compare calibration results to simulation





2.3 Relations to Other Standards

2.3.1 ASAM standards

A key component of a solution's quality will be the compliance with standards, e.g. for data formats. Therefore, a strong reference to existing standards in the ASAM portfolio is expected. As those evolve so will also do the aspects of compliance with them.

The ASAM Quality Checker will be a key component to measure compliance of an artefact with ASAM standards. Successful completion of quality checks might become a metric in the prospective standard developed in this project.

The ASAM TestSpecification provides an overview of test means and test methods. Any categorization of the same and any identification of use cases within the proposed project shall align with the results of the ASAM TestSpecification project.

2.3.2 Other standards

A key aspect of simulation model quality will be the model's "correctness", i.e. the degree to which it reflects the simulated physical device or effect. IEEE has defined metrics for the quality assessment of cameras [4].

The project group will have to scan the scientific and industry landscape for similar standards for all sensor modalities (lidar, radar, ultrasonic etc.), as compliance with them may become an aspect of quality.

2.4 Relations to other activities

There are different activities on-going which are or might be relevant for this project. Once the project starts, one task will be to revise the state of the art of these activities which might have some impact on it. Just to mention some of them:

- EVIDENT (Enabling VIrtual valiDation and vErificatioN for ADAS and AD feaTures): This project will explore strategies which trades degree of fidelity (how realistic the testing environment is) in a quantifiable way, to move testing and validation from physical test track to virtual simulation. Which enables testing of higher complexity and the use of more parameters and greater variation in the parametrization of the scenarios. This would enable systematic testing of a software feature or complete vehicle before it is deployed.

- HAIViSH - Handling of AI-based Virtual Systems for Homologation of AD/ADAS, is a project initiated by prostep IVIP. In WP4, they address the aspect "Methodology for simulation model validation and credibility assessment".

3 Technical Content

3.1 Overview

The proposed concept project shall create a human- and machine-readable data format (e.g., based on XML) for the storage and exchange of information about the assessment of a simulation solution's fitness for a given use case and/or ODD. The definition of metrics to be stored in the data format shall be part of the project.

Use cases for simulation methods, as defined in the project ASAM TestSpecification and as extended within the proposed project shall be clearly identified and a weighting of metrics per use case shall become part of the standard (example: real-time capability of a solution is essential for HiL usage but not mandatory for SiL applications).

The metrics are expected to reflect measurable physical properties of simulation solutions (e.g., pixel resolution of a camera model) as well as quantifiable but hard-to-measure properties (e.g., structured user documentation). Recommendations or, better, clear instructions on how to acquire relevant information shall be part of the standard.

A comprehensive assessment of tools on the market shall not be part of the project, but exemplary assessments with selected tools and models shall be performed on a consensus basis in order to prove the applicability of the new standard.

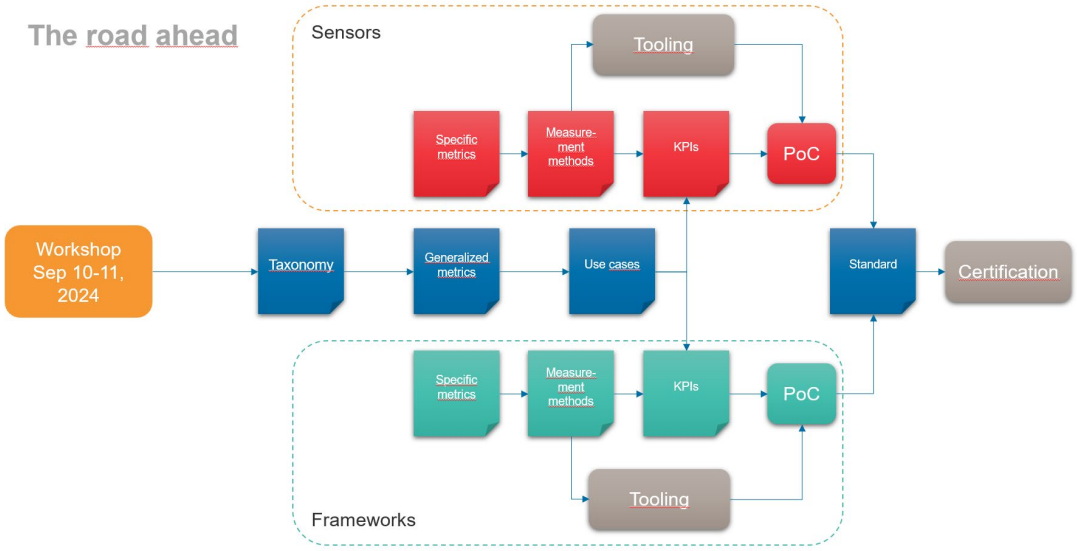
Existing ASAM standards, like ASAM OpenODD and ASAM OpenMATERIAL 3D shall be considered when defining metrics and their applicability to certain use cases.

The project group shall identify a method to describe the ODDs of simulation frameworks and sensor models by incorporating the results of the ASAM OpenODD project. The dependency of KPIs on uses case and ODDs shall be considered.

A set of checker rules for the ASAM Quality Checker framework shall be provided as one result of the project.

3.2 Organization

The project shall distinguish, e.g. in two sub-groups, between activities for sensor simulation and simulation frameworks:



Overarching topics for the entire project group shall be

- the definition of a taxonomy for “simulation quality”,
- the definition of generalized topics which apply to sensors and frameworks, and
- the definition of use cases.

Each sub-group shall work on topics specific to its scope, i.e.

- specific metrics
- definition of measurement methods
- definition of KPIs

All of these topics shall become part of the future standard.

The applicability of the standard shall be tested in PoCs. It is proposed to have sub-group-specific PoCs and to use tools provided or proposed by group members. No tools shall be developed within the scope of the project. For the sensor sub-group, a collaboration with IEEE P2020 [4] shall be considered for camera use cases. For the framework sub-group, the initiative SimCert [3] may provide an initial set of metrics.

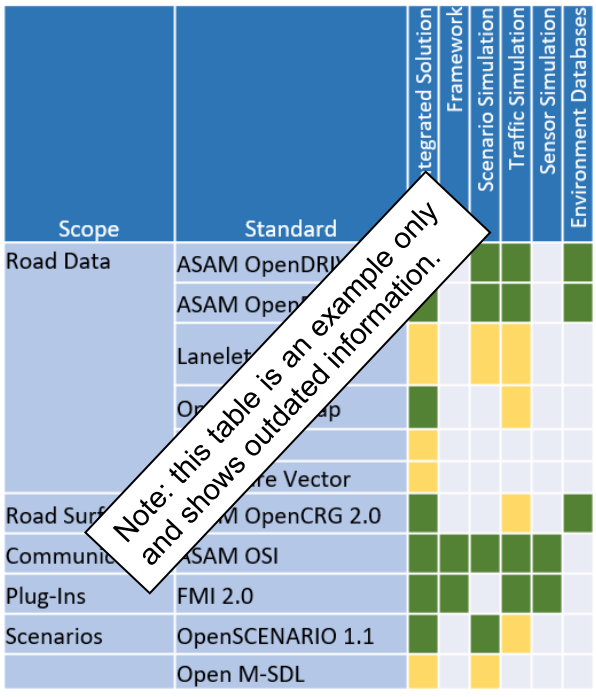
3.3 Examples from previous work

Under the SimCert initiative [3], criteria have been developed that give an idea of what topics might need to be covered by the project proposed here. The following is an excerpt of the rating criteria applied commonly across different tools and use cases:

> *Some criteria may be applied across various product categories. These are listed here. Be aware that the meaning of a criterion might need to be adapted slightly when applied to a specific category;*

if applicable, we will indicate any deviation from the common definition in the criteria list of the respective product category

- *Usability*
  - *writing a script in a text file is a powerful option for configuring a complex system; but hardly any user will want to start with this step*
  - *good usability means that the user is guided by an intuitive user interface throughout all stages of setting up and executing a simulation; the workflow has to be clear*
  - *a fragmentation of user interfaces or an explicit switch from one application to another for setting up a simulation is not considered a good option*
  - *the meaning and naming of components, parameters etc. has to be clear*
  - *data storage and retrieval has to be easy*
  - *installation and setup shall work out-of-the-box without any need for a support call*
- *Standards Compliance*
  - *the request for standards refers to data communication, interface and storage formats; others may be added in the future*
  - *a standard will typically be specific to the industry where it is applied; it may either be an offical standard issued by a recognized institution or a de-facto standard supported by a substantial fraction of players in the market*
  - *standards have been created for being used – at least this is what one hopes for; "standards compliance" does not so much rate how many standards are supported as how well relevant standards are supported and to which degree they are part of the product*
  - *the following standards are considered mandatory or "nice-to-have" (ie, relevant) for the different product categories:*



- *Documentation*
  - *a well documented product is of great value; the user will know what to expect and how to operate the system*
  - *a user manual is the bare minimum expected in terms of documentation*
  - *a tutorial, preferably interactive or by videos, is a big plus to a getting familiar with a complex product*

    - *meaningful documentation that is available up-front, i.e. while a potential user is still in the selection process, will increase a product's documentation rating*
    - *context-sensitive help within a product (tool-tips, links to documentation) are considered beneficial*
    - *release notes provide a quick overview of improvements and extensions a product may have undergone with its latest version*
- *Support*
    - *questions are meaningless without answers; and a (novice) user will typically have a lot of questions when it comes to using a product*
    - *the key factors for achieving a high rating for the support of a product are*
        - *availability*
        - *responsiveness*
        - *correctness of answers / information*
        - *follow-up on "hard" issues*
    - *it is quite common in industry that support questions have to be addressed within one business day*
    - *products are expected to come with a good basic level of support; an option for professional services will increase the rating since it may allow the user to better tailor the extent of support to his needs*
- *Licensing*
    - *not everything that comes for free is good; but it is good if something comes for free*
    - *a product that comes as open source or with little restrictions on use by a broad community has a strong point to make*
    - *hybrid products that come with some part of little restriction on use and "high-fidelity" extensions are a good compromise*
    - *fully commercial products may achieve good ratings by providing flexible licensing terms*
- *Completeness*
    - *more features do not necessarily mean that something is getting closer to completeness*
    - *we will rate this criterion according to the applications covered vs. the requirements of the market and how a product's feature set may help reducing the overall need for the user to perform an integration of their own or 3rd party tools.*
- *Openness*
    - *hardly any product will be used standalone, only communicating with itself*
    - *a high rating for the openness criterion can be achieved by making sure that the user has transparent access to all data flows and data formats (including documentation) that are used within the product*
- *Extensibility*
    - *no matter how highly integrated a product may be, there will always be something else that the user needs in addition or in exchange for a given component*
    - *it is crucial for a high rating of this criterion that a product provide a means to extend functionality either by recommended extensions of the solution provider or by well documented interfaces*
    - *bi-directional data communication interfaces are considered a plus in the rating of this criterion*
    - *a well-documented API for implementing custom functionality within the solution will lead to a higher rating*
    - *plug-in concepts are highly appreciated*
- *Scalability*
    - *one running simulation is good – many are better; if you want your system-under-test (SuT) to be "challenged" by as many test cases in as short a time as possible then you have two options: first, increase the number of tests that run in parallel, second, run each test faster (and "third", do both at the same time)*
    - *the higher the execution speed of a single simulation run that a solution can achieve without any decrease of accuracy, the higher its rating will be*

        - *in terms of increasing the number simulation runs, we will not only rate how many instances can be executed in parallel but, most of all, how good the management of the respective deployment and of the test cases is; this includes the selection or generation of test cases, the execution itself and the storage and retrieval of simulation results*
    - *Versatility*
        - *a product needs to adapt to the use case, not vice versa*
        - *development and testing of ADAS and AD systems by means of environment simulation solutions is performed in various operating environments:*
            - *Model / Software-in-the-Loop (MiL / SiL)*
            - *Driver-in-the-Loop (DiL)*
            - *Vehicle-in-the-Loop (ViL)*
            - *Hardware-in-the-Loop (HiL)*
            - *etc.*
        - *a high rating in versatility shall be achieved if the user is able to deploy the product to different operating environments (X-in-the-loop) with little configuration effort and with seamless exchange of data between these environments*
    - *Cloud Deployment*
        - *when it comes to managing hardware costs, cloud deployment is an attractive option (you pay what you use)*
        - *the rating of this criterion shall take into account whether the product can be operated on a cloud provider's hardware and what the effort is to get it onto the cloud*
        - *we shall give higher ratings to products that come as ready-to-go cloud deployments, either as simulation-as-a-service implementation or just by an easy way of getting started on virtual hardware*
        - *partnerships of a product with established cloud providers are considered a benefit*
    - *Determinism*
        - *the most annoying thing in assessing the system-under-test are "effects" that come-and-go with each new round of testing*
        - *determinism is a criterion which is important to achieve for a product since it means that with identical input (also in terms of timing) the results – even randomized parts – will be exactly the same.*
        - *the highest rating is reserved for products that provide determinism for the overall operation with all components being active*
    - *Real-time Capabilities*
        - *a product will typically claim that it can cover a broad range of use cases (see Versatility); one of the key uses of environment simulation software is the so-called hardware-in-the-loop (HiL) testing. But hardware hardly knows anything but real-time operation. Therefore, real-time capability is a strong point in the overall rating of a product (BTW: also SiL testing may require real-time operation)*
        - *real-time may be achieved in a couple of ways – by adapting speed or accuracy. Depending on whether a product is on the faster or slower side under normal operation, the strategy for real-time may be different and will be rated accordingly*
        - *how real-time is achieved, and whether accuracy has to be compromised in order to achieve this goal will have great influence on the rating*
        - *how "hard" the real-time is that a solution can achieve is also part of the rating; are there any frame drops over a given measurement period? Is there a mitigation concept for frame losses?*
        - *flexibility in triggering and synchronization when it comes to real-time operation will play a certain role in the overall rating*

The proposed project shall provide clear metrics for each of the aspects it deems relevant for the assessment of a solution's quality.

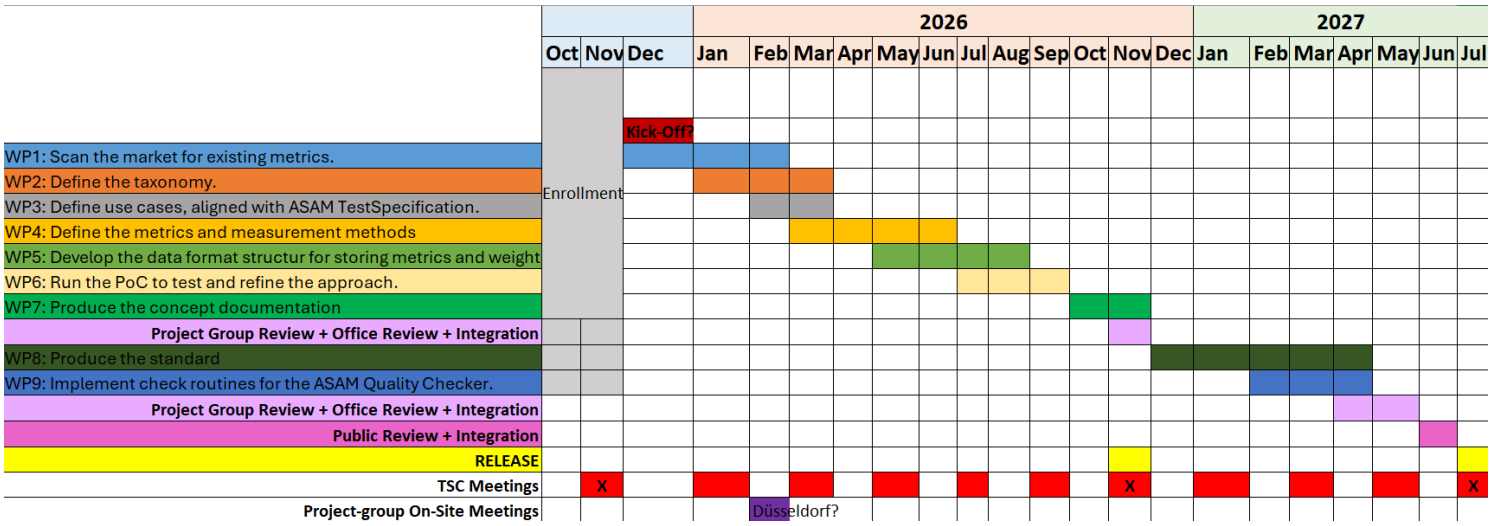
3.4 Work packages

The proposed project shall comprise the following work packages:

- **WP 1: Scanning of the market**: Identify applicable metrics for simulation solutions that have been defined in other projects.
- **WP 2: Taxonomy**: Define a taxonomy for the quality assessment, the components, and the use cases that are aligned, among others, with the ASAM TestSpecfication project.
- **WP 3: Define use cases:** In alignment with and extension of the findings of the ASAM TestSpecification project, define exemplary use cases for simulation methods that may require a distinction in the application of metrics (WP 4).
- **WP 4: Define metrics and instructions for their measurement:** This task may be implemented in three (or more) parallel groups, e.g., focusing on common components, simulation frameworks, and sensor models. Check for existing standards (e.g. P2020 [4]) for their applicability for the 'simulating tests to test simulation' approach.
- **WP 5: Define the data format:** Define a human- and machine-readable data format to store metrics; define, optionally, a format to store the weighting of metrics for identified use cases.
- **WP 6: PoC:** Apply the metrics (defined in WP 4) at a limited number of solutions, together with the definition of KPIs based on the metrics. Use the findings to reiterate with WP 4.
- **WP 7: Concept Documentation**: Write a concept paper document along the rules of the ASAM Editorial Guide.
- **WP 8: Standard Documentation:** Write a standard specification document along the rules of the ASAM Editorial Guide.
- **WP 9: Quality Checks:** Define and implement check routines for the ASAM Quality Checker to allow for testing of compliance with the data format (WP 5).

4 Project plan

The total project development (without the review phase) shall run for a total duration of 16 months. The concept paper release would be planned for November 2026 followed by final standard release by July 2027.



5 Deliverables

At the end of the project, the project group will hand over the following deliverables to ASAM:

Table 1 Concept Deliverables

Item No.	Description
1	Concept paper of the data format, taxonomy and metrics.
2	Documentation of application examples and results of the proof-of-concept stage
3	
4	

Table 2 Standard Deliverables

Item No.	Description
1	Specification of the data format for storing quality metrics.
2	Documentation of application examples and results of the proof-of-concept stage
3	Guidance and best practices, including methods for measuring specific properties of simulation solutions (i.e classified by metric type and use case)
4	Checker rules for the ASAM Quality Checker framework

5.1 Review Process

The process for deliverable review documented in the project guide is applicable to all projects (see here).

The ASAM OR will provide further details on quality criteria and tools used prior to the initiation of a review in a project.

Table 2 Selection of Review Type

Please indicate whether the project aims to perform an ASAM member review or a full public review. This is not required for maintenance projects.	Public Review

6 References

[1] Viehof, "Objektive Qualitätsbewertung von Fahrdynamiksimulationen durch statistische Validierung", 2018, https://tuprints.ulb.tu-darmstadt.de/id/eprint/7457

[2] Workshop "Quantifying Simulation Quality" at GSVF 2021, Quantifying Simulation Quality - GSVF 2021 - SimCert

[3] SimCert initiative by MD Engineering Services, Rating Simulation Solutions - SimCert - Certainty in Simulation.

[4] IEEE Standard for Automotive System Image Quality, P2020, published in March 2025, IEEE SA - IEEE 2020-2024.

[5] Patrick Mueller, Matthias Lehmann, Alexander Braun, "Simulating tests to test simulation" in *Proc. IS&T Int'l. Symp. on Electronic Imaging: Autonomous Vehicles and Machines*, 2020, pp 149-1 - 149-8, https://doi.org/10.2352/ISSN.2470-1173.2020.16.AVM-149

[6] Braun, Alexander. "Automotive mass production of camera systems: Linking image quality to AI performance" *tm - Technisches Messen*, vol. 90, no. 3, 2023, pp. 205-218. https://doi.org/10.1515/teme-2022-0029

[7] Workshop "Quantifying Simulation Quality" by HSD and ASAM e.V., https://www.asam.net/conferences-events/detail/quantifying-simulation-quality/