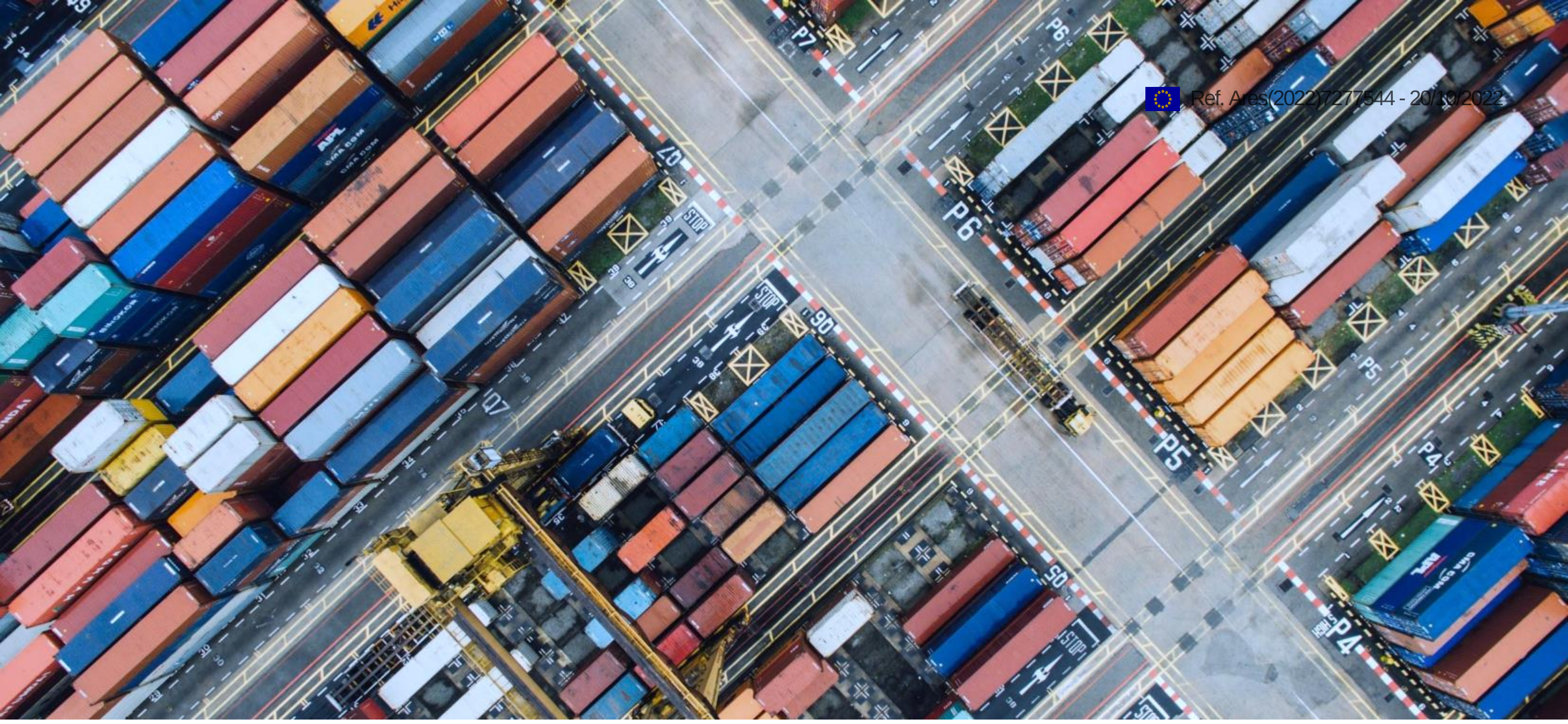

AWARD
Scaling autonomous logistics

D7.1
Test and Evaluation Plan

—

Lead: VTT

Date: 19/10/2022

Dissemination level: PU

The project has received funding from the European Union's Horizon 2020 research and innovation program under Grant Agreement No 101006817.

Document information

Project	
Project Acronym	AWARD
Project Full Title	All Weather Autonomous Real logistics operations and Demonstrations
Grant Agreement No.	101006817 - H2020-DT-ART-2020
Project Coordinator	EasyMile
Website	www.award-h2020.eu
Starting Date	January 1st, 2021
Duration	36 months

Deliverable	
Deliverable No. – Title	D7.1 Test and evaluation plan
Dissemination Level	Public
Deliverable Type	Report
Work Package No. – Title	WP7 Testing methodology and evaluation
Deliverable Leader	VTT
Responsible Author(s)	Sami Koskinen (VTT)
Responsible Co-Author(s)	Loha Hashimy (Enide), Guillaume Saint Pierre (Cerema), Matthias Neubauer (FH OOE), Peter Fröhlich (AIT), Arttu Lauhkonen (VTT), Sylvain Rheme (CertX), Hannes Watzinger (Digitrans), Mads Skovsgaard Rasmussen (DFDS), Steven Mouws (Dematic), Kolbjørn Njøten (Avinor), Dominik Schallauer (Austriatech)
Technical Peer Review	Mads Skovsgaard Rasmussen (DFDS)
Quality Peer Review	Anne Larribe (EasyMile)
Submission due date	31/12/2021 20/10/2022 (revised version)

Revision history

Revision Number	Date	Author	Company	Changes
0.1	29/03/2021	Sami Koskinen	VTT	Template for collecting information
0.2	12/11/2021	Sami Koskinen	VTT	Revised structure and input for all chapters
0.3	9/12/2021	Sami Koskinen	VTT	All test site plans collected
0.4	16/12/2021	Sami Koskinen	VTT	Version sent for peer review
0.5	20/12/2021	Mads Rasmussen	DFDS	Technical review
0.6	23/12/2021	Anne Larribe	EasyMile	Quality review
1.0	23/12/2021	Sami Koskinen	VTT	Version submitted to the EC
1.1	11/10/2022	Sami Koskinen	VTT	Updated version after project review: clarifies FESTA and WP6 relationships as well as data collection
2.0	20/10/2022	Magali Cottevieille	EasyMile	Quality check and final version for resubmission

ACKNOWLEDGMENT OF EU FUNDING

The project has received funding from the European Union's Horizon 2020 research and innovation program under Grant Agreement No 101006817.

CONTACT

Ms. Magali Cottevieille
Project Coordinator
EasyMile
21 Boulevard de la Marquette
31000 Toulouse
France

Email: magali.cottevieille@easymile.com
www.award-h2020.eu

Table of contents

List of figures

List of tables

List of acronyms

ADAS	Advanced Driver Assistance System
ADS	Automated Driving System
AGTS	Autonomous Ground Goods Transportation System
AGV	Automated Guided Vehicle
AV	Automated Vehicle
EDAS	Ecological Driving Assistance System
ESAW	European Statistics on Accidents at Work
FMS	Fleet Management System
GHG	Greenhouse Gas
GLMM	Generalized Multilinear Mixed Models
GPS	Global Positioning System
HCI	Human–Computer Interaction
HDV	Heavy-Duty Vehicles
HMI	Human–Machine Interface
ISA	Intelligent Speed Adaptation (system)
KPI	Key Performance Indicator
NACE	Nomenclature of Economic Activities
NRMM	Non-Road Mobile Machinery
PI	Performance Indicator
UI	User Interface
VR	Virtual Reality

Executive Summary

This report summarizes the preliminary test and evaluation plans for the operational vehicle testing phase of the AWARD project. The tests will use automated industrial trucks to transport goods on fixed routes. The project targets operations in any weather.

The test sites include two factory-like areas as well as an airport and a port. These industrial sites are partly access-controlled but some routes include public road segments and parking areas. One test vehicle will be operated at each test site.

The evaluation will cover several aspects, namely safety, environmental, efficiency, technical and user-related. The goal is to identify what changes with automation, when compared to earlier manual operations. Vehicles will be equipped with data collection systems to enable analysis of driving style, operational performance and different types of interesting events such as emergency stops from log data.

Users such as personnel directly involved with the industrial truck operations will be interviewed regarding their experiences. Other stakeholders such as drivers of other vehicles and operators of other systems will be interviewed, too. Since the automated vehicles play only a part of the operations at an industrial site, selected general performance indicators related to industrial operations will also be monitored.

This report compiles initial research questions and data needs at the end of the first project year. The data needs and focus are to be further discussed and finalized with the test sites during the second year of AWARD project, before operational tests are to commence. During the first year, the vehicles have been either in development or in pre-testing phase. During the second year, more details regarding the final setup of the tests will be compiled, as well as data logging and processing details. The test and evaluation plans will be updated and then published in D7.4 Final test and evaluation plan.

This report already includes a look in available statistics regarding the safety of similar operations. Also, test sites have outlined their current related operational key performance indicators. Such data will be later needed for scaling up the identified benefits from tests with only a few vehicles: what would change, if more of such vehicles would be automated. The main target for scaling up will be to consider the operations at the test sites, but secondly the project targets to review potential changes at EU level.

1. Introduction

The AWARD project is currently making detailed plans to test and evaluate automated industrial trucks in real-life operations. The project has four test sites: an airport, a port, and two factory areas. The sites are similar in the sense that the driving distances of the trucks are short and the areas are mostly occupied only by employees. One vehicle will be tested at each site. All operations involve automated transfer of cargo and certain human interaction regarding loading and unloading. The tests are to showcase and assess the feasibility of operations also in difficult weather conditions.

Adverse weather conditions are generally seen as one of the factors slowing down the take-up of automation. In the AWARD project, weather conditions are a research topic but the project will also demonstrate and evaluate reliability of new sensor technology in real operations.

The final operational tests in the AWARD project are not mere “demonstrations” in nature. They aim for deeper integration with other systems than what is common for first experiments. The project targets long-term use that would continue after the project. The tests aim to clarify, as is typical for “field operational tests”, the final issues that need to be addressed before wider commissioning. Regarding common terminology and the targeted scope of testing, the project can be best classified as a pilot project, falling between demonstrations and large-scale field operational tests.

In terms of testing and evaluation methodology, the project adopts the FESTA Handbook [1]. The Handbook was originally produced by FESTA support action in 2008. The methodology was to guide upcoming automotive field operational tests and a new wave of EU projects. Since then, the handbook has been repeatedly updated by follow-up networking projects, collecting lessons learned (FOT-Net, CARTRE and ARCADE). The FESTA mainly targets large-scale user tests, but in recent years it has been successfully applied in various smaller testing campaigns, as well. Use of a scientific testing and evaluation methodology supports scientific setup of tests and gives a proven structure for work. Scientific rigor can help to attain valid and comparable results.

The first year of evaluation preparations cover the left side of the FESTA V (figure 1), that describes the process and steps of carrying out a field operational test.

As in FESTA, the main topics in the beginning of a study are to scope research questions and work towards agreed focus and data collection. Tests and data collections have to be planned from the perspective of statistical evaluation – commonly this means collecting enough data both with and without the tested system in use.

The evaluation work package (WP7) has also contributed to other key areas in FESTA, although in the AWARD project, the results are documented elsewhere. These are the ethical and legal matters regarding handling of personal and measurement data in the project. The related deliverables are D1.1 EPQ-H-Requirement and D10.5 Data Management Plan.

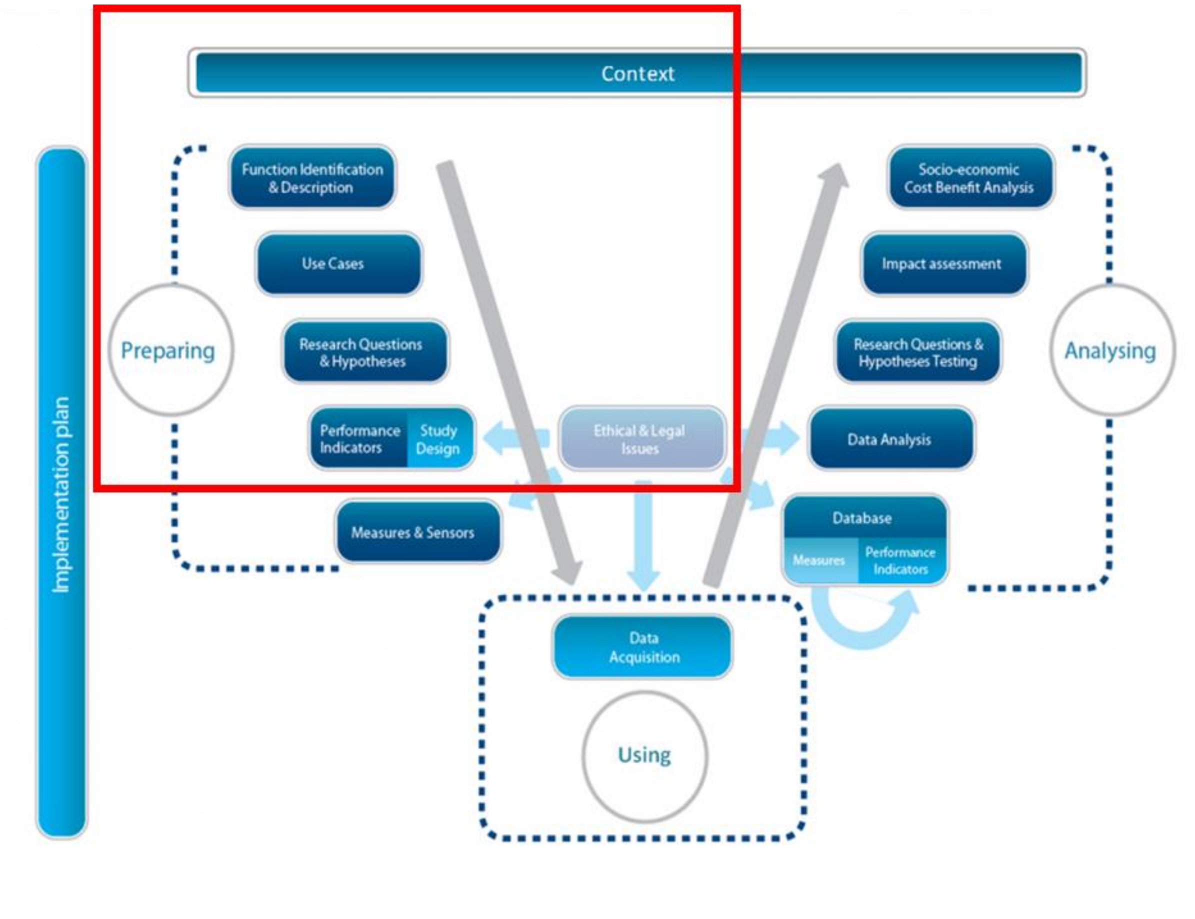

Figure 1. FESTA V and the focus of the preparations described in this deliverable marked with red

In addition to the operative tests in the final project year (the scope of this deliverable, WP6 and WP7), the AWARD project has earlier testing related to product development (WP3) and safety validation (WP4). Before the operative tests can begin, a certain amount of pre-testing and fine tuning will also be necessary, to ensure smooth performance. The results of these earlier development phases will be mainly discussed in other project reports.

The pre-testing period, however, must also include log data collection for checking correctness and quality. User-related aspects such as training and agreements are also necessary. The test plans in general are a joint product of both the test site teams in WP6 and the evaluation experts of WP7. Where WP6 considers the daily management, WP7 oversees the test setup.

The main goal of this evaluation work package is to assess the results of operative/commissioning tests: how the systems perform, after main development and initial setup have been completed. Operational performance will be analyzed statistically. For example, what was the downtime and what were the reasons for it, how many emergency stops did happen, how much cargo was transferred, what was the experience of different interviewed stakeholders etc.

The evaluation as well as the work package are divided into five main areas, the plans of which are summarized in this document:

1. User and stakeholder evaluation
2. Safety impact assessment
3. Process efficiency and quality evaluation

4. Environmental impact assessment
5. Technical evaluation.

This split is familiar from past FOT projects and also from the FESTA Handbook, when setting research questions. The exact focus naturally varies from project to project.

The user and stakeholder evaluation considers various factors in their data collection and interviews: for example, how much time has been spent doing tasks, how the work experience has changed, and have expectations been met. Safety impact assessment studies the changes in safety margins, number of near-misses and emergency braking events, and related likely changes in likelihood of different types of accidents. Process efficiency and quality consider the main performance numbers used at the test site, e.g., length of delays, timeliness and amount of cargo transferred. Environmental impact assessment focuses on energy use and related changes in emissions. Technical evaluation complements previous product development test results for selected focus areas. It will evaluate, for example, positioning accuracy and issues during long-term testing. That is, whether the machine got distracted from its route.

The evaluation teams of each area work rather independently but following a common structure and common log data processing. The main steps of evaluation are:

- Outlining possible research questions and related data needs.
- Deciding high-priority research questions and data logging together with test sites.
- Detailing test plans and practical arrangements with test site leaders.
- Analysis of sample log data and data quality, before operative tests begin. Adaptation and development of data analysis tools. Pre-testing of all phases that will come up during actual tests.
- Running the experiment, collecting data and interviewing workers and stakeholders.
- Finalizing, documenting and sharing the collected dataset for analysis.
- Calculating performance numbers from collected data and reviewing videos and interviews.
- Identifying and assessing benefits and drawbacks related to automated operations versus baseline human-driven vehicles. The evaluation is done by evaluation areas first separately. The results also provide input to cost–benefit calculations.
- Scaling up of the impact assessment results to consider wider possible benefits at factory level and further projections at industry/EU level, if more vehicles would be automated. Historical data on accidents, usage and performance are necessary to estimate the changes, e.g., a percentual change in workplace injuries or amount of human work necessary.

This report discusses the first three steps. Next year, the focus moves to steps 4–7. The final evaluation process will be presented in higher detail in the next deliverables, after it is seen, if the FESTA principles need to be adapted for industrial test cases. Each evaluation area will use their own detailed calculations.

Chapters 2 to 5 of this report introduce the current test plans at each test site. These are initial plans at the end of 2021, and they are used in internal project discussions to consider what needs to be arranged and decided next. The presentations are from evaluation perspective, whereas the build-up of test sites is tracked in more detail under WP6.

Chapter 6 introduces the five evaluation areas and their plans. Chapter 7 is about research priorities and Chapter 8 is about test data management. Chapter 9 considers future work, and Chapter 10 is about conclusions. Annex I gives an outlook on occupational accident statistics and Annex II discusses other statistical information available for scaling up the effects of automation to EU level.

2. Automated forklift

2.1. Test site introduction and routes

The automated forklift use case of the AWARD project will be tested at Linde Material Handling's headquarters in Aschaffenburg, Germany. Linde Material Handling is one of the world's leading manufacturers of forklift trucks and warehouse equipment. The tests address outdoor logistics within the factory. Empty racks will be moved between a transfer location and storage yard stacks.

2.1.1. Process description

The factory relies on a system of racks with which all parts are brought to the production lines. The racks are owned by Linde. They are shipped back and forth between Linde and their parts suppliers.

Full racks are brought to the production lines by manual forklift trucks. The products within them are consumed during the production process. Once a rack is empty, it is removed by a manual forklift driver and brought to the storage yard. At the storage yard, all empty racks are buffered and sorted per type.

Once a sufficient number of empty racks is reached, an automatic transport is organized to have them shipped to parts suppliers.

2.1.2. Automatic transport

The automated driving tests will focus on the empty racks returned to the storage yard. Components arrive on site in dedicated racks of several types. A single, frequently used rack type was selected for automated transport: the steering axle rack (see figure 2).

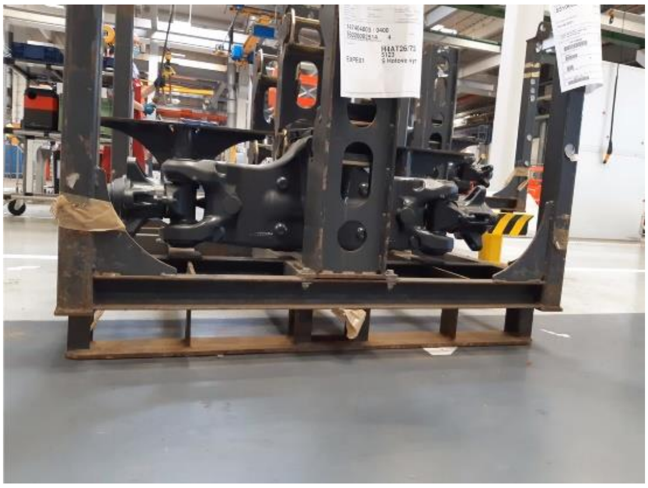

Figure 2. Steering axle rack. Two axles fit on one rack.

The retrieval of empty racks from the production lines will be handled by manual forklift drivers. They will collect empty racks and bring them to a transfer location (figure 3), just outside the factory hall.

Figure 3. Various racks prepared for transport to the yard

From the transfer location, the automated forklift truck will retrieve the racks and bring them to the storage yard where they will be stored in a block stack (figure 4, figure 5).

Figure 4. Storage yard

Figure 5. Block stack in storage yard

The outdoor section is best suitable for automation:

- It matches the project goals.
- The indoor environment is not (yet) suitable for automated load handling (lack of floor space).
- There is no system triggering the collection of empty racks. This is done by the forklift drivers, based on experience.
- The piloted operation doesn't affect main production and cannot slow it down considerably.
- Therefore, the value of automation is focused on the outdoor section (the long drive).

2.1.3. Additional demonstration plans

The test site has planned an additional demonstration about automatically loading and unloading a tugger train with a forklift. This demonstration would take place on a small test area, aside from the main tests. However, this demonstration will be rather a technical development activity and it is not to be evaluated using WP7 processes. There might be possibilities to consider the case in a future scenario work.

Currently, all factory logistics are being restructured and there are plans for wider use of tugger trains, in the future.

2.1.4. Route

Figure 6. Automated route: Source, Path and Destination

The test route (figure 6) includes some of busiest roads in the factory. Interaction with various types of traffic will be required:

- Pedestrians
- Cyclists
- Cars
- Vans
- Trucks
- Forklifts.

In addition to the challenging traffic situation, an 8 % ramp is included in the route segment towards the storage yard. The route is repeated 2–3 times per hour, currently driving in two shifts.

2.2. Description of automated vehicle functionalities

Figure 7. Current vehicle.

Figure 7 shows a current vehicle. The automated vehicle will be comparable but somewhat larger. There are both diesel and electric forklifts in use. The electric forklifts are normally recharged during the nights. There are plans to recharge the automated vehicle during operations, as the task is slow and there should be an opportunity to drive to the charging station. This is currently being investigated. No big delays in transport would be expected for the charging, but there would also be a short trip to the charging station.

In contrast to the manually driven forklifts, the automated forklift is set to reverse. This is to ensure braking behavior without dropping the load. The direction also allows for improved environmental sensing, without obstructions.

Human drivers do not always strictly follow lanes in the factory area, but the automated truck will do that. In intersections, the human drivers sometimes cut straight. Forklifts do not even need to follow traffic rules, e.g., right of way.

During the tests, the vehicle will always have a safety operator onboard.

2.3. Affected other operations

Truck drivers at the site are not only drivers but logistical operators. They know that certain parts will be soon needed. When empty racks are being moved manually, the truck drivers can focus on other tasks.

Full racks will be manually transported during the project, as there is no alternative system (enterprise resource planning system, fleet management) to count and manage the parts. Potential problems with moving empty racks are not likely to disturb the main production.

A fleet management system will order the automated truck to detect and pick up empty racks. The automated vehicle will continue to transport the racks to a dedicated row at the storage yard. The vehicle will report accurate rack coordinates back to the fleet management system.

2.4. Performance goals and pre-existing indicators/statistics

The main target is to save human hours. The factory operations heavily depend on forklifts. Secondly, the manufacturer has plans to manufacture automated forklifts – the pilot allows various developers to examine the possibilities, firsthand.

No ready baseline data exists. Production numbers will later be checked regarding how many parts have been used and how many movement operations have been needed.

Current vehicles have high uptime standards, and their failure intervals are tracked. Factory equipment is expected to be used effectively. The automated vehicle should reach similar levels to manual forklifts. However, automated vehicles will also face other than traditional system/technical issues: for example, an unknown box in front of an automated truck will stop it, correctly. The calculation of related performance indicators could be different, and it could be interesting to consider such aspects. AGVs (automated guided vehicle) commonly provide vehicle status and fault states to fleet management and cloud-based evaluation tools.

The performance of an automated vehicle, when comparing maximum number of transferred racks per hour to human-driven vehicles, is expected to be lower. However, the operation should be more consistent, creating a steady flow of parts. Finally, the throughput over a longer period should be comparable to that of human drivers. In this selected transport operation, a steady flow will suffice.

2.5. Data logging

2.5.1. Baseline data collection

A manually driven forklift is to be equipped with an aftermarket GPS and a video data logger. This data will be processed to analyze human driving behavior on the route. The data logging equipment has not yet been chosen, but data collection would be tested during 2022. The vehicle to instrument should be electric, as the automated vehicle will be. A human observer will make notes at least of battery level during the baseline data collection period. The driver ID should also be marked, to see how much variation is there in driving behavior.

2.5.2. Automated vehicle and treatment period data collection

The automated vehicle is not likely to arrive early enough in the project to allow long-time manual use of the same vehicle for comparison purposes. Still, it is to be driven e.g., a couple of days in manual mode. Due to the different driving direction in manual and automation modes, logged safety margin data might not be directly comparable. Such a short data collection phase might, anyhow, reveal comparable energy use data.

Automated driving software provider EasyMile is reviewing evaluation data requests and the target is to implement similar log file collection from each test site.

2.5.3. Access to log data

The test site leader is the controller of the dataset. The data will be made available to named persons carrying out evaluation, as confidential data. If any sample data is to be published, that is to be agreed separately.

2.6. Participants and stakeholders

2.6.1. Participant agreements, instructions and introductions

An introduction session about the automated driving experiment will held for a large group at the factory, focusing on safety aspects and possible changes in practices. Special training will be organized for the safety operators, selected from forklift drivers. They will be made aware of data collection. Further details are to be clarified.

2.6.2. Planned interviews

The foreseen list of persons to interview:

- Safety drivers (a couple)
- Other truck drivers regarding the interaction
- Supervisor for the drivers
- Health and safety department

- Other factory workers (training about automation will be for everyone)
- Logistics department (here: customer point of view)
- Optionally a manager with economical background.

2.7. Tentative timeline

Table 1. Timeline of UC1

Phase	Start month	End month
Pre-testing	32	32
First data sample for evaluation		During 2022
Baseline data collection	32	33
Operations and interviews	33	33 (36)
Dataset finalisation		34
Evaluation and reporting	34	36

2.8. Additional controlled technical tests and use of human observers

The test site offers a wide range of somewhat chaotic traffic situations (pedestrians, cyclists, other forklifts) and they could be of special focus. Regarding possible detection tests, detecting the forks of other forklifts could be challenging.

Open area positioning accuracy seems like a candidate topic for technical testing, as extra landmarks will possibly be installed. The accuracy of stacking racks on top of each other could be measured. If the stacks are not made accurately, the piles could even fall. Uneven terrain can place further requirements.

The area should be easily accessible for human observers to make notes of the automated vs manual operations. Safety operators will keep notes, as well.

2.9. Emergency procedures

The tests will follow general safety guidelines described in more detail in D5.2 and D1.1. Safety validation tests and risk assessment will be carried out, before tests are to begin. As a forklift manufacturer, there are existing practices for vehicle risk and site risk management.

The vehicle logic includes certain precautions against stopping in unsafe areas. For example, the vehicle should not remain long in specified zones such as fire exits and emergency vehicle access routes.

There will be introduction sessions held for the people affected and specific training for the safety operators of the automated vehicle.

3. Hub-to-hub shuttle service

3.1. Test site introduction and routes

The companies Rotax and Schenker are developing together with Digitrans an “Industrial hub-to-hub autonomous logistics on public road” use case. It is to improve the efficiency of operations from a human resources and infrastructure point of view.

The goal is to automate the existing empty goods milkrun between the engine production factory of BRP Rotax and the logistic center of DB Schenker in Gunskirchen, Upper Austria. The two sites are connected via public roads, including crossing areas and a main road (figure 8).

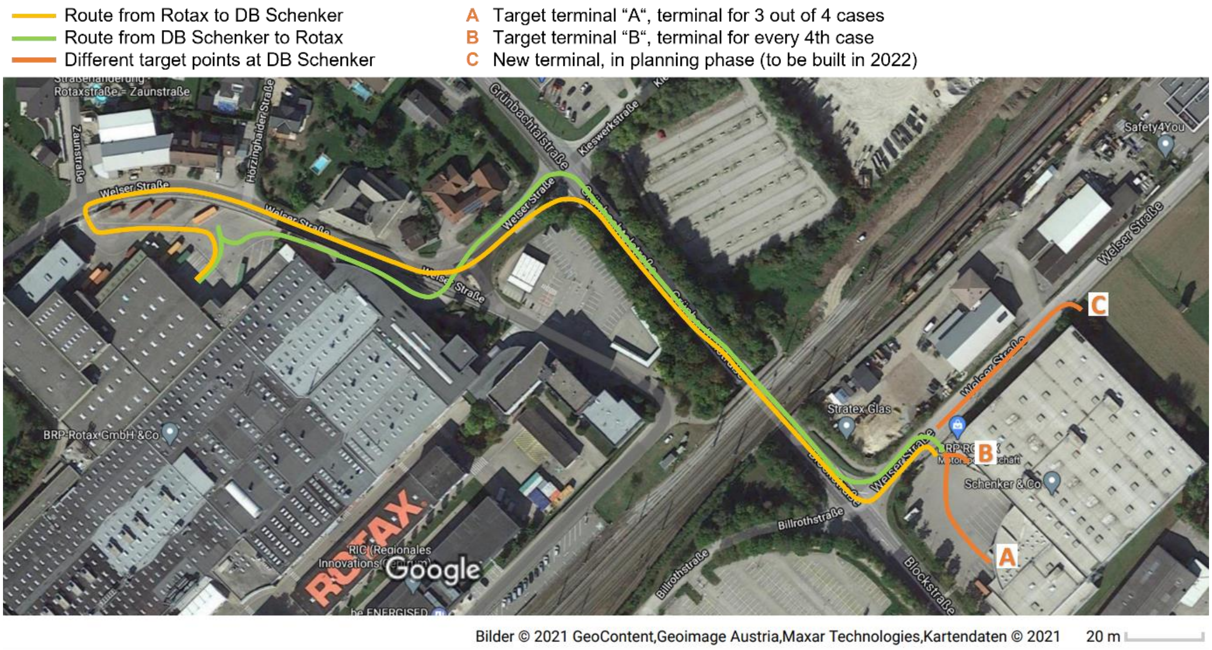

Figure 8. Route map

The complete process is currently performed by one person: loading the truck, driving the truck, unloading the truck. The work itself is not very satisfactory for the drivers and they have around 1/3 of his/her working time left, which cannot be put to use.

Year-round 24/7 operation would facilitate a working business plan for the automation use case stakeholders BRP-Rotax and DB Schenker. The targeted 24/7 operation would allow an optimization of the operation working time with respect to safety and security standards as well as a CO_2 emission reduction due to night delivery with low traffic. The goals are improved use of existing roads and infrastructure without obstructing new land regions. Thereby, the efficiency and safety of the whole industrial logistic value chain would be improved for all involved parties. The Industrial operator could increase the equipment ROI (Return on Investment), the employees could work in a safer and more efficient environment and the remaining road users and neighborhood residents would benefit from a reduced traffic volume and noise pollution.

Therefore, this use case focuses on the highly automated transport for an empty goods milkrun. which is here done by lattice boxes in an enclosed loading unit (here: swap body) with a truck between two hubs including public roads and restricted areas. The process for this use case is described and visualised in the current form (As-Is) and in the planned automated form (To-Be) in AWARD Deliverable 2.2 Chapter 6.2

The example application situation for the Austrian use case can be described as follows:

- Filled boxes (kanban system) are loaded onto the truck at the logistics centre.
- The boxes are carried to the production site via public road connection.
- The boxes are unloaded and available empty boxes are loaded onto the truck.
- The empty boxes are carried back to the logistics centre.

At the moment, this process/route is repeated every full hour from 6:00 am to max. 10:00 pm from Monday to Friday. The driving part takes some 5 minutes of an hour, the rest of it is unloading and loading. The automation plans are about the driving part. Otherwise, the process would remain the same as with a human-driven truck.

During the course of this project, the autonomous vehicle will be operated with a safety driver on board.

3.2. Description of the automated vehicle

Currently, the route is driven with a standard diesel truck (figure 9).

Figure 9. Current diesel truck transports

The instrumented vehicle will be a KAMAG PM (figure 10). There are not yet photos of the final vehicle. The vehicle will first undergo safety validation on the Digitrans's proving ground and further pre-tests at the test site, preparing for automated operations.

Figure 10. The swap body transporter KAMAG PM

With the new electric vehicle, there will be a route change (figure 8), even while driving in manual mode. The conventional diesel trucks are not allowed to drive the new route, but electric vehicles are. The new route will be a bit shorter and more direct. There are plans to install new traffic lights on the route, to provide a green wave.

With the electric vehicle, the process requires a new loading cycle. However, the vehicle will be charged overnight, so it should not affect testing. A charging station will be installed.

3.3. Affected other operations

Production depends on timely delivery of goods. Even a half an hour delay would be considered significant. The buffering options are to be clarified, later.

3.4. Performance goals and pre-existing indicators/statistics

As a future vision, the transportation could run 24/7, benefiting from silent night hours. That could affect noise, emissions etc. It would also mean a change in production hours. However, the extended 24/7 operations are not to be tested during the pilot phase.

The main goals during the pilot are about:

- reducing and optimizing human work;
- providing safety and efficiency for employees;
- providing a proof of concept, a starting point for further use cases and new legislation.

There are no pre-existing statistics regarding the operation.

3.5. Data logging

Most of the data collection will happen with the same instrumented vehicle when it is ready. It will be driven both in manual and automated modes. Baseline data will be collected by using a human to drive for selected periods. Additionally, transitions between human and automated driving will be visible from log data.

The current diesel truck is already recording road weather data using a Mobile Detector MD30 product by Vaisala. It is a spectroscopy-based device that detects water/ice/snow on the road. Data is also available from the closest weather station and a roadside weather station is to be installed. The data collection started in summer 2021.

A sample GPS (Global Positioning System) dataset of two weeks will be collected from the diesel truck in 2022, to enable development of evaluation calculation tools. Fuel consumption notes will be made as well as explanatory notes of possible interruptions in operation. A data sample could also be collected during different seasons and weather conditions.

The electric vehicle will include a data logging software by the navigation provider EasyMile.

Use of a human observer outside the vehicle is yet to be clarified, but the safety driver will be able to mark down important events according to WP7 instructions.

3.5.1. Access to log data

The test site leader is responsible for data collection and protection. According to good data protection practices described e.g., in FESTA, data will be shared with named evaluation organizations under confidentiality.

3.6. Participants and stakeholders

3.6.1. Participant agreements, instructions and introductions

The truck drivers participating in the study will receive a training session, introducing them to the study. They are to sign a participant agreement, so that they understand the nature of data collection and that their driving data can be used for analysis purposes.

3.6.2. Planned interviews

The initial list of persons to interview:

- Truck drivers (2)
- Schenker, who employ the truck drivers
- Rotax
- Management regarding future visions
- Safety responsible.

Probably there are no other workers regularly near the truck. Neighbors and such are not likely to be interviewed, either.

3.7. Tentative timeline

Pre-testing is to start end of 2022. Operations will begin with a safety driver onboard, Phase 1 (table 2). There will be winter months during the Phase 1. Phase 2 aims at demonstrations without the safety driver during the final months of the project (table 3).

Table 2. Timeline of UC2, phase 1

Phase	Start month	End month
Pre-testing	22	24
First data sample for evaluation		20
Baseline data collection	20	27
Operations and interviews	24	27 (36)
Dataset finalisation		27
Evaluation and reporting	28	36

Table 3. Timeline of UC2, phase 2

Phase	Start month	End month
Pre-testing	27	28
Operations and interviews	29	36
Dataset finalisation		33
Evaluation and reporting	33	36

3.7.1. Additional controlled technical tests and use of human observers

Positioning accuracy is seen as the best candidate for technical testing. There will be both ultrawideband and GPS in use. Reversing to the ramp requires the highest accuracy. There is also a narrow passage when entering Rotax site. Although, collision avoidance sensing will help to center the vehicle.

Communication between the truck and the new traffic lights could also render technical evaluation topics.

3.7.2. Emergency procedures

The tests will follow general safety guidelines described in more detail in D5.2 and D1.1. Safety validation tests and risk assessment will be carried out, before tests are to begin. In addition, fleet management will consider and document emergency procedures beforehand. There will be introduction sessions held for the people working close to the vehicle and specific training for the safety drivers of the automated vehicle.

4. Airport baggage tractor

4.1. Test site introduction and routes

Testing at Oslo airport (figure 11), Norway, addresses the use of automated bag transport. The test vehicle is an automated baggage tractor that is used to move bag carts.

Figure 11. Oslo airport

At the test site, testing of different missions/routes has already begun with the vehicle itself and with related safety validation. These early tests have formed the phases 0 and 1 of deployment efforts.

The vehicle will remain in test operations throughout the rest of the AWARD project. The complexity of missions will increase in phase 2, in April 2022. At that phase, the performance of the vehicle will begin to be compared against human-driven vehicles for evaluation purposes. The previous tests have tested functional features and safety validation. Prior to phase 2, ground handler training needs to be completed, risks related to the new route assessed and data collection issues sorted out.

The current plans for phase 2 missions are as described below. The mission/route 1 (figure 12) would have the following segments:

1. TractEasy waiting mission point
2. Drive to airplane stand
3. Pick-up filled carts and bring them to PMZ Arrival
4. Manual unloading of baggage
5. Return to intermediate storage
6. Drive back to another gate or to TractEasy waiting mission point.

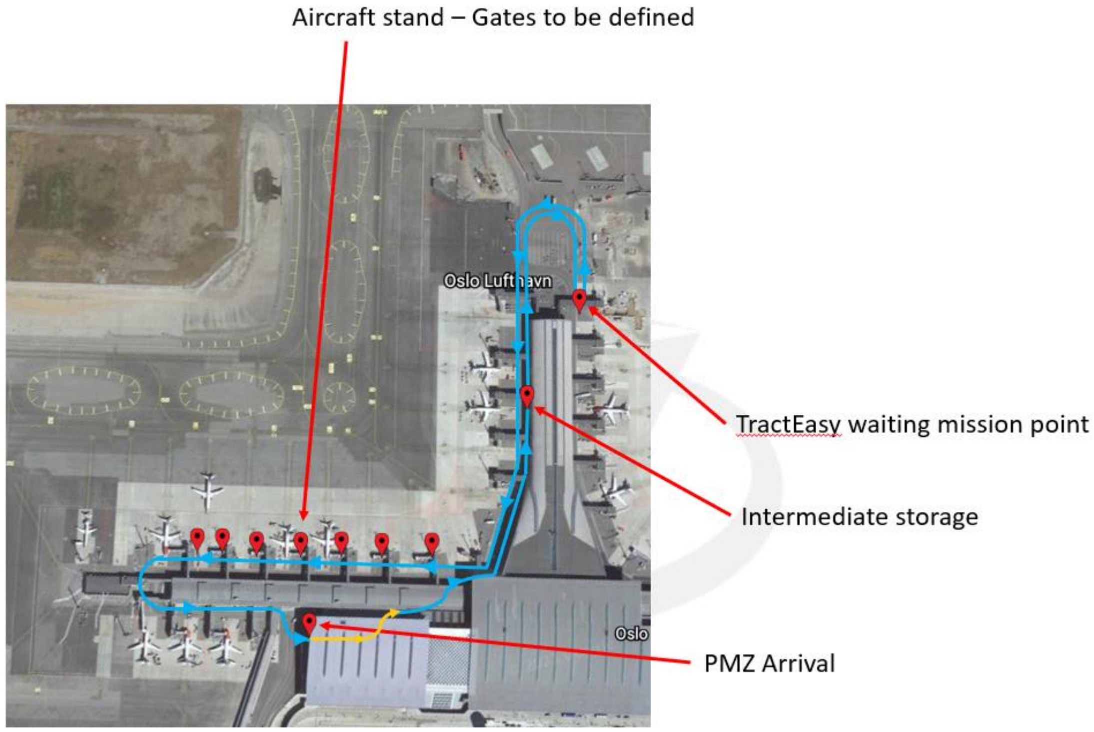

Figure 12. Phase 2, planned mission 1

The second route (figure 13) has the following segments:

7. TractEasy waiting mission point
1. Drive to carts' intermediate storage
2. Pick-up empty carts and bring them to carts' storage
3. Return to waiting mission point or to intermediate storage for another trip.

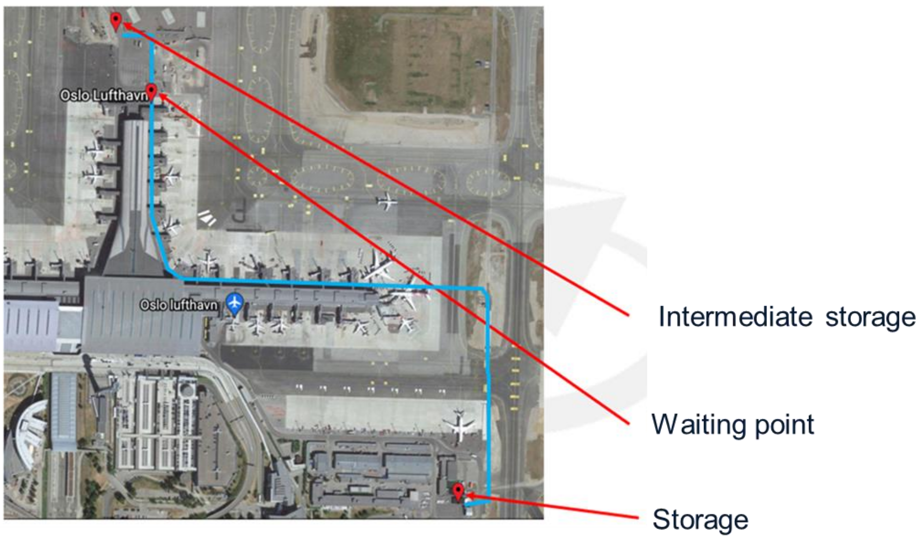

Figure 13. Phase 2, planned mission 2

Phase 3 and subsequent phases are to be designed in 2022, using the experiences from phase 2. The focus for phase 3 has not been fully discussed and planned, but will probably be shifted to driverless operations with an escort car, operations in more difficult weather and on more complex routes. The assignments can include night operations as well as difficult locations regarding positioning. These phases examine the broader feasibility of baggage tractor automation.

4.2. Description of automated vehicle functionalities

The tests use an instrumented TLD baggage tractor (figure 14). TLD is a leading airport equipment provider. The vehicle will be instrumented by project partners and use EasyMile's navigation software stack. The vehicle is electric. Even previously, all indoor and outdoor transportation of luggage at the Oslo Gardermoen airport is performed with electric vehicles.

Figure 14. Automated baggage tractor

The vehicle will drive automatically, while unhooking and hooking carts remains a manual operation during the project timeline. The fleet management will require signaling when such operations have been completed, for the vehicle to continue.

The maximum speed of the vehicle is 30 km/h. The targeted speed will be similar to human-driven vehicles. The automated vehicle is slightly wider than a normal tractor, approximately 40 cm, due to sensor instrumentation.

During the tests, there will always be a safety operator inside the vehicle.

4.3. Affected other operations

The automated vehicle will interact with (luggage) handlers, but also regularly with refueling, catering, cleaning operations and other actors participating on apron during the turnaround process.

There is no direct monitoring currently about baggage tractor related timings nor delays. Only if an airplane gets delayed, the main reason for that delay would get marked. In case the reason would be ground handling operations and especially a malfunction with the baggage tractor, such notes would exist.

If there was a problem with the automated vehicle, probably a manually driven vehicle would be ordered as a replacement. The arrival of one could cause a delay in the airplane leaving, considering turnaround times around 20–30 minutes which may cause penalties to be declared (according to operational agreements).

4.4. Performance goals and pre-existing indicators/statistics

The targeted long-term advantages of automating luggage tractors are:

- Reduction in number of drivers
- Safety improvements
- Better utilization of luggage tractor capacity
- Less driving, if automated vehicle trips are better planned and managed
- Less manual planning with improved fleet management
- Better utilization of cart and container capacity.

It is not currently seen that there would be big changes in driven routes, as the airport doesn't have many alternatives. Delays are not generally expected to be affected, either. As the planned automated driving speed is comparable to human drivers, no large differences are expected in luggage damages due to them dropping in corners.

Energy consumption and operational hours data is currently only in the bookkeeping of several companies performing ground handling operations. It will be easier to directly collect baseline data using a data logger.

Regarding other automated operations at the airport, vehicles with such features are already used to follow a lead car, when cleaning snow. Other operations have not been planned for the near future but the baggage transport.

Airport accident data is collected systematically and similar data is available from many countries.

4.5. Data logging

The automated vehicle will be used for all data collection. It will be driven in manual mode in selected periods, to collect baseline data.

The AV data collection is under development. It will be similar to other project test sites, using the automated driving software stack provided mainly by EasyMile. The software stack will be able to record vehicle status such as position and mode using a high sampling rate (2–5 Hz). Additionally, fleet management API can provide vehicle location, emergency stops, battery level and information on travelled distances. The vehicle has a front-facing camera for recording video.

Video recording must be stopped within certain coordinate areas. The related discussions are ongoing.

The dataset will be managed by the test site partners. It will be uploaded to an FTP server for named evaluation partners to access, under confidentiality.

4.6. Participants and stakeholders

A few selected persons will be trained to operate the automated tractor. Due to ambiguities related to insurance matters, the ground handlers can only be passengers in the vehicle while the safety driver is an employee of Avinor or Smart Airport Systems. The ground handlers will later be interviewed on how they see the automation, interaction with the automated vehicle and generally changes to their work.

4.6.1. Participant agreements, instructions and introductions

The test site will draft a consent form for the safety drivers whose driving behavior is logged in the project. An introduction session will be held for the affected workers, discussing safety and data collection.

4.6.2. Planned interviews

The proposed persons to interview:

- Safety operators/drivers
- Ground handlers.

In addition, some of the other actors on the airport could be interviewed based on possible interaction with the automated baggage tractor:

- Refueling/Catering/Cleaning
- Airport safety responsible
- Head of the baggage handling
- Handler team lead.

4.7. Tentative timeline

Phase 2 is currently being planned to run from May to June in 2022, with pre-testing preparations in April (table 4). One of the main topics to sort out is improved data collection from the automated vehicle. These tests will form the first complete dataset to be evaluated in the AWARD project.

Table 4. Timeline of Phase 2

Phase	Start month	End month
Pre-testing	16	16
Baseline data collection	16	18
Operations and interviews	17	18
Dataset finalisation		19
Evaluation and reporting	20	23

Upcoming Phase 3 and onwards will be 2-month periods testing new missions/routes, using the same setup as phase 2 (table 5). Phase 3 will be about driverless operations with an escort car, more complex missions/routes, harsh weather, operations in dark, and difficult spots for localization (to be discussed).

Table 5. Timeline of Phase 3

Phase	Start month	End month
Operations and interviews	Autumn 2022	

4.7.1. Additional controlled technical tests and use of human observers

Airports are not an easy place for additional scenario-like technical or observation tests due to being strictly closed areas. However, the test site also features some of the most difficult weather conditions in the AWARD project, testing in Nordic winter. The automated operations in difficult weather are a natural topic for more detailed evaluation.

Beside weather-related sensor data evaluation, some of the first ideas for technical focus would be positioning while driving tunnels (there is also considerable slope) and achieving required visibility in certain intersections with uneven ground that partly blocks lidar rays. More detailed technical evaluation plans are to be discussed in 2022.

4.7.2. Emergency procedures

The tests will follow general safety guidelines described in more detail in D5.2 and D1.1. Safety validation tests and risk assessment will be carried out, before new missions are to be run. Routes are analyzed section by section. Site risk assessment has been carried out.

There will be introduction sessions held for the people affected and specific training for the safety operators of the automated vehicle.

5. Port operations

5.1. Test site introduction and routes

The port demonstrations will take place at DFDS's Rotterdam (Vlaardingen) terminal in the Netherlands (figure 15). The tests will focus on an automated Terberg Tug that will move trailers in the terminal area. In addition to rearranging trailers, the planned routes include gate transits to and from public road and also loading and unloading of a ship.

Figure 15. Terminal area

The Rotterdam terminal is a busy Roll-on/Roll-off (RoRO) terminal with ferry routes to Immingham and Felixstowe (both UK). The terminal has 22 weekly departures, transporting more than 150.000 trailers from the Netherlands to UK every year. The terminal has three length spans for ships to moor at.

The port terminal has a total number of 32 tugs similar to the Terberg truck that will be demonstrated with autonomous capabilities. The tugs are being used for different purposes:

- 25 tugs are being used vessel operations
- 6 tugs for Technical Services Workshop
- 1 tug for Transhipment Area.

The operations include three missions for the tug:

1. Trailer move from drop off area to a holding area, ready for ship loading
2. Last mile/hub-to-hub transport from terminal to public road, including gate-in and -out process.
3. Vessel loading.

Mission 1: Trailer move from drop off area to holding area

The AV (Automated Vehicle) will pick up a trailer from the drop-off/pick-up area in parking slot area “F” (figure 15). The AV will drive along the “main road” in the terminal and drop it of at a newly adapted terminal area, parking slot area “R” ready for loading the ship. The parking area is being adapted with new “herringbone” pattern parking slots, which make the parking and placing of trailers more efficient. The trailer is unhooked by the safety operator and the AV leaves the parking area.

Mission 2: Public road access and gate-processes

The demonstration is about accessing public road and demonstrating gate transits, gate-in and gate-out. The AV starts with a trailer attached on the terminal. The AV will then transit the gate and enter a roundabout just outside the terminal. The AV will make a revolution in the roundabout just outside the terminal and go back in through the gate demonstrating gate-in including access validation and damage detection. The AV then drops the trailer at a terminal parking slot.

Mission 3: Loading of a trailer onto a ship

This mission will demonstrate the AV’s capability to drive onto the ship in loading operations. The ship starts with the trailer hooked and drives to the ship’s ramp. The AV drives onto the ship to the main deck, does a U-turn and exits the ship. The AV then drives back to the drop-off/pick-up area.

5.2. Performance goals and pre-existing indicators/statistics

Some of the main foreseen benefits of automating trailer tugs are:

- More accurate location information for parked trailers, as the automated vehicle transmits exact coordinates. Currently, human drivers use area-level information, and occasionally a trailer is not found directly.
- Autonomous rearrangement of trailers in off-peak hours would provide efficiency gains through better planning. Unloading and loading operations could become more efficient as the site has been prepared better. Further, a faster turnaround is a key performance indicator in ports.

Currently, the port tracks fuel consumption and operating hours of the tugs, partly through manual processes. Also, the hours for maintenance are tracked. It is of interest how automated electric tugs perform on these parameters compared to diesel-driven tugs.

How many of the current tug operations could be fully automated is still an open question and the demonstration should provide better insight into this. The interaction between manual and automated transfers will also be studied.

One aspect of automation is linked with training time needed for conventional tug drivers: how long are the introduction phases and how long does it take for a new driver to become as efficient as the experienced drivers. These times could be compared to take-up and training of new automated vehicles and new routes for them.

5.3. Description of automated vehicle functionalities

At this moment, there are no photos of the automated vehicle. The main difference compared to human-driven tugs is that the vehicle is not planned to be driving at reverse. Whereas human drivers reverse when picking up and occasionally otherwise, the current safety validation plans for the vehicle include driving forward, only. This is to ensure visibility. A more detailed analysis of the process changes will be carried out during the second project year.

5.4. Affected other operations

The automated Terberg Tug will have to perform its missions together with other tug drivers but also alongside external drivers who come into the port. Besides these drivers, the main effect will be for dispatch operators, planning of loading and unloading of ships. Maintenance operations will be affected especially in future operational scenarios.

5.5. Planned interviews

The main participants and stakeholders to interview are:

- Truck drivers and trained safety operators of the automated tug
- HSSE (health safety security environment) representatives
- Fleet managers
- General management.

5.6. Data logging

5.6.1. Baseline data collection

It is seen as possible to install an aftermarket GPS and video logging for a selected human-driven tug. Baseline data collection will be tested in 2022.

In addition, the test site collects data on tug fuel consumption, operating hours and the need for maintenance.

The automated vehicle will be used to collect a baseline data as a human-driven vehicle, but only for a short period. The late arrival of the automated vehicle in the project will prevent long time data baseline data collection with it.

5.6.2. AV data collection

The AV data will be collected using EasyMile navigation software and will mainly include signals collected also at the other sites. No test-site specific signals have not been defined, yet.

The main comparisons from the test site perspective are related to energy consumption, braking behaviour, and time to move trailers.

5.6.3. Access to log data

The test site leader is the controller and owner of the data. The data will be made available to named evaluation partners as a confidential dataset.

5.7. Tentative timeline

The baseline data collection will be tested during 2022, while the automated vehicle pre-test preparations are likely to start in March 2023 (table 6). The actual testing is scheduled to start around May 2023. The foreseen weather for March can include rain and fog.

Table 6. Timeline of UC4

Phase	Start month	End month
Pre-testing	27	28
First baseline data sample for evaluation		Summer 2022
Operations and interviews	29	31 (continues until 36)
Dataset finalisation		32
Evaluation and reporting	32	36

5.8. Additional technical evaluation tests

The main planned topics for technical evaluation and extra controlled testing are:

- positioning inside the ship
- compensation of the dynamic movement of the ship, when entering it.

5.9. Emergency procedures

The tests will follow general safety guidelines described in more detail in D5.2 and D1.1. Safety validation tests and risk assessment will be carried out, before tests are to begin. In addition, fleet management will consider and document emergency procedures beforehand.

The vehicle logic includes certain precautions against stopping in unsafe areas. For example, the vehicle should not remain long in specified zones such as fire exits and emergency vehicle access routes.

There will be introduction sessions held for the people affected and specific training for the safety operators of the automated vehicle.

6. Evaluation areas

This chapter introduces the five evaluation areas, their main goals and methods. The areas and aspects to be evaluated in AWARD project, within the WP7, are:

1. User and stakeholder evaluation
2. Safety impact assessment
3. Process efficiency and quality evaluation
4. Environmental impact assessment
5. Technical evaluation.

The descriptions here reflect the initial plans and they are to be seen as a status report at the end of 2021.

6.1. User and stakeholder evaluation

This work is related to T7.3 and aims at assessing users' acceptance and experiences with users and stakeholders. Three main contributions are provided: a broad identification of acceptance factors across user groups and use cases, iterative evaluations of HMI (human–machine interface) designs, and investigations within the specific context of the pilot. This task contributes to this D7.1 with methods and questionnaires. It reports results in D7.3, at the end of the project.

The work is motivated by the implications from the user and stakeholder requirements that were derived in Task 2.2 (D2.2). The research questions motivated by this work are described below.

6.1.1. Research questions

Evaluation of HMI design for AGTS fleet management

To compile lessons learnt from fleet management HMI solutions that have been developed in T5.3, laboratory studies will be conducted with a dedicative prototype environment. figure 16 shows that there are several phases of human operators' tasks for AGTS (Autonomous Ground Goods Transportation System) fleet management. While the vehicle is driving or being used by direct process participants (loading, unloading), the fleet manager can use the FMS (Fleet Management System) as a dashboard.

There are two possible reasons the fleet manager might want or need to interact with the system. The first is a regular check which may include the KPIs (Key Performance Indicator), vehicle positions, protocols, statistics, and other parameters. The second reason is an interruption by the FMS which calls for the fleet managers' input. This interruption needs to be forwarded to the fleet manager either on the dashboard as a popup, a mobile push notification or other HMIs.

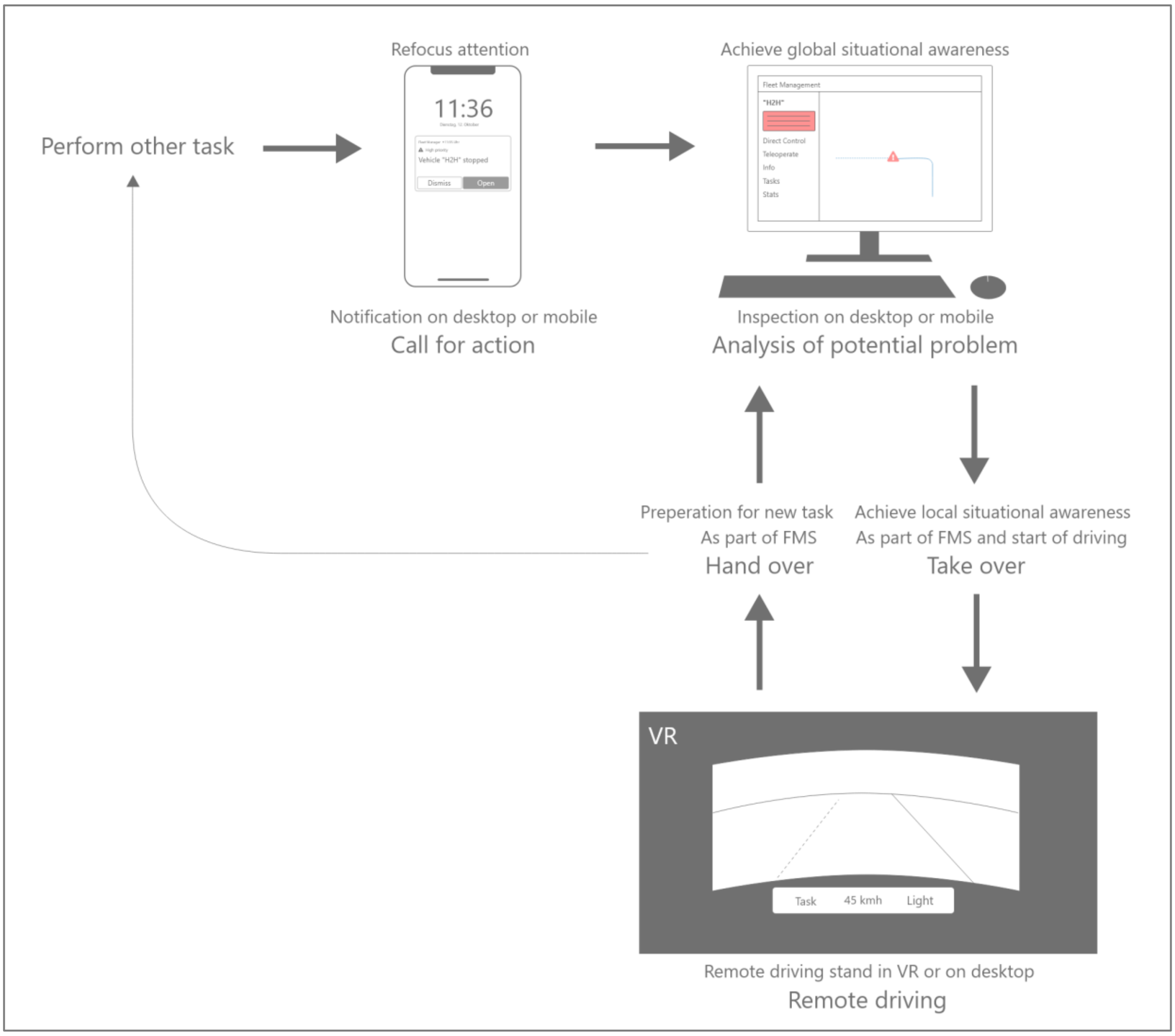

Figure 16. The cycle of human AGTS fleet operation

Regardless of the HMI type, the fleet manager must manage the interruption. The second step for both is checking the FMS for problems. If no problems occurred, which is most likely in the case of regular checks, the fleet manager can start her/his regular monitoring activity. Otherwise, if a problem occurred, which is the case if the FMS called for the fleet manager, he/she needs to analyze the problem and choose one of three processes. If the problem can be handled by the fleet manager directly, he/she can control the erroneous part directly from the FMS. This may include updating the mission, controlling the vehicles (no remote driving) or managing traffic lights.

For each of the different activities within the cycle of human AGTS fleet operation, a set of research questions for comparative user interface evaluations has been defined:

- Refocus attention:
  - Which HMI techniques should be used to gain attention of operators working in different contextual environments?
  - How should notifications be designed not to distract the operators from their respective main activities?
- Achieve global situation awareness:

    - Which information should be provided in a user interface, in order to make users quickly understand the overall situation of a remote vehicle?
    - How can decisions be supported on what to do next (e.g., whether to intervene immediately, postpone the intervention or hand over to another person)?
- Achieve local situation awareness:
    - Which user interface design elements support spatial situational awareness, i.e. indicate the position, heading and surroundings of the vehicle?
    - Which user interface design elements assist users best in assessing the state of the vehicle?
- Remote driving:
    - Which features must be displayed in the HMI for successful teleoperation, in scenarios with different levels of complexity?
    - Do information items only require temporary visualization? When are these needed by the user to successfully complete the task (before, during, after)?
    - Does immersive technology (e.g. virtual reality, VR, and moving bases) increase the quality of the teleoperation task?
    - Can a visual mission briefing at the beginning of the journey adequately prepare an unprepared operator for the via teleoperation?
- Preparation for new task:
    - Which features help to bring operators up to speed again to their previous task?

Contextual evaluation of AGTS acceptance at the pilot site

The final objective of the user and stakeholder evaluation will be the impact of the developed technology on the acceptance of the stakeholders within the specific context of the project pilot site. The acceptance factors described above will be used for this evaluation as well. This includes the following main questions:

- What is the influence of the designed AGTS on work processes?
- How is efficiency of work processes perceived by workers and managers?
- How are safety, security and reliability perceived?
- How does the fleet management interface impact situational awareness?

Investigation of overall user and stakeholder acceptance of AGTS

Achieving a comprehensive and balanced understanding of acceptance factors of AGTS – going beyond the four pilot contexts and getting a cross-stakeholder and across use cases – is a key goal of T7.3. In order to build and expand on the already achieved insights and methodologies of the AWARD requirements workpackage (WP2), the developed requirements framework will be further contextualized and detailed, by integrating user and stakeholder responses. Based on the content analysis and weightings from the qualitative responses from the requirements analysis, further research questions and items have been added.

The main research question in this regard is as follows: what are the main factors and factors contributing to acceptance of AGTS?

Further questions relating to intermittent variables are: In what respect do stakeholder groups differ with regard to acceptance; and in which regard do the different use cases differ with regard to acceptance?

6.1.2. Data needs

The most serious data needs are for the contextual evaluation at the pilot sites. As indicated in table 7, it is necessary that for each pilot, relevant stakeholder representatives should be approachable. This especially means that for each of the four pilot sites it should be possible to interview at least 10 employees handling the daily logistics od production processes, 10 other road users (i.e. persons nearby the cars), 2 managers planning the logistics operations, and 1 specialist in charge of policies and standards. Also, the organization shall support the AWARD ethics procedures, especially the handling for the consent forms. Also, T7.3 research should be supported by allowing for observations and documentations, under strict compliance with the AWARD ethics guidelines, as well as other overall standards and regulations applicable to the pilot site.

Table 7. Data needs for user and stakeholder evaluation

ID	Data need from tests	Purpose	Related RQs
1	Participant data gathering	Interview, observation and data logging shall be produced and made available, in compliance with the AWARD ethics guidelines. Test participants may not be members of the pilot sites, but they should be representative of the system target users.	Evaluation of HMI design for AGTS fleet management
2	Availability of user and stakeholder representatives for the pilot studies	In each pilot, relevant stakeholder representatives should be approachable. This especially means that, for each of the four pilot sites should be available – 10 employees handling the daily logistics od production processes – 10 other road users – 2 managers planning the logistics operations – 1 specialist in charge of policies and standards – Also, the organization shall support the AWARD ethics procedures, especially the handling for the consent forms.	Contextual evaluation of AGTS acceptance at the pilot site
3	Authorization of researchers to visit and observe the pilot test sites.	The researchers involved in T7.3 shall be allowed to visit the pilot test sites and collect observational data. The AWARD ethics and any company procedures will be followed.	Contextual evaluation of AGTS acceptance at the pilot site

ID	Data need from tests	Purpose	Related RQs
4		The teleoperation system shall be accessible for testing.	Evaluation of HMI design for AGTS fleet management
5	Wide distribution of acceptance survey.	The survey has to be distributed to various stakeholders and users.	Investigation of overall user and stakeholder acceptance of AGTS

6.1.3. Evaluation plans and methods

6.1.3.1. Evaluation of HMI design for AGTS fleet management

In order to address research questions, a laboratory-based evaluation environment by AIT within T5.3 will be used that allows for the user-centered demonstration and experimentation of HMI solutions to address the above challenges for human operation of remote management of vehicle fleets. The key principles of the TeleOperationStation are:

- A mixed-reality setup including VR, hardware interaction elements, and miniaturized sensing
- Miniature cars to enable for quick, safe and easy experimentation. Offering miniature cars (rather than graphically simulated cars) is expected to provide a more realistic control experience.
- Rapid HMI prototyping components, driven by previous HCI (human–computer interaction) theory and literature
- Coverage of all of the above interaction steps
- Interfaces for functional integration with automated fleet management systems
- Covering automated road transport logistics scenarios, facilitated by comprehensive requirements research
- Open to various further application contexts
- Useable by the community, through printable equipment, open software modules, scenario descriptions, and available UI (User Interface) Patterns.

Measures: For each study, a specific set of measures will be used. These will include a questionnaire on workload (NASA-TLX), attention analysis (eye tracking), an interview on subjective evaluation, behavioral measures on minimum distance to object threshold / nr. of collisions, as well as task completion and time to task completion.

Procedure: People are asked to fill in a questionnaire – this has nothing to do with the study and only counts as a distraction. While completing the questionnaire, they receive a notification on a phone indicating that a vehicle needs assistance. The subjects therefore have to go to the teleoperation booth, put on their VR goggles and take over the vehicle.

At the beginning, the task is described in VR so that the persons know what to do. Then the VR image changes to the camera images of the vehicle and the person can start steering. During the journey, several methods are used to measure the person's condition. Once the person has reached the destination, he or she can finish the task by pressing a button. A questionnaire follows. The ride is repeated six times (two conditions with 3 rides each). A questionnaire follows after each run. After each run, a final interview to ask for alternative design suggestions.

Scenarios: There will be different driving scenarios, which vary with regard to whether the destination is in sight and how many curves are in the route.

6.1.3.2. Contextual evaluation of AGTS acceptance at the pilot site

In order to evaluate the user and stakeholder acceptance of AGTS implementations within the application context, on-site investigations will be performed during functional investigations. These will be closely coupled to the scenarios defined for the four different pilots. The research will be performed in different phases:

- Introduction phase: The system is introduced to the staff and started up. Here, expectations can be raised.
- Pilot conduction phase: observation of staff interactions with the vehicles, while unloading, etc.
- Reflection phase: Inquiry about the experiences in the pilot phase.

6.1.3.3. Investigation of overall user and stakeholder acceptance of AGTS

Throughout the AWARD project, a technology acceptance framework (ARTLAM) is used, in order to structure the captured inputs of users[1] (figure 17).

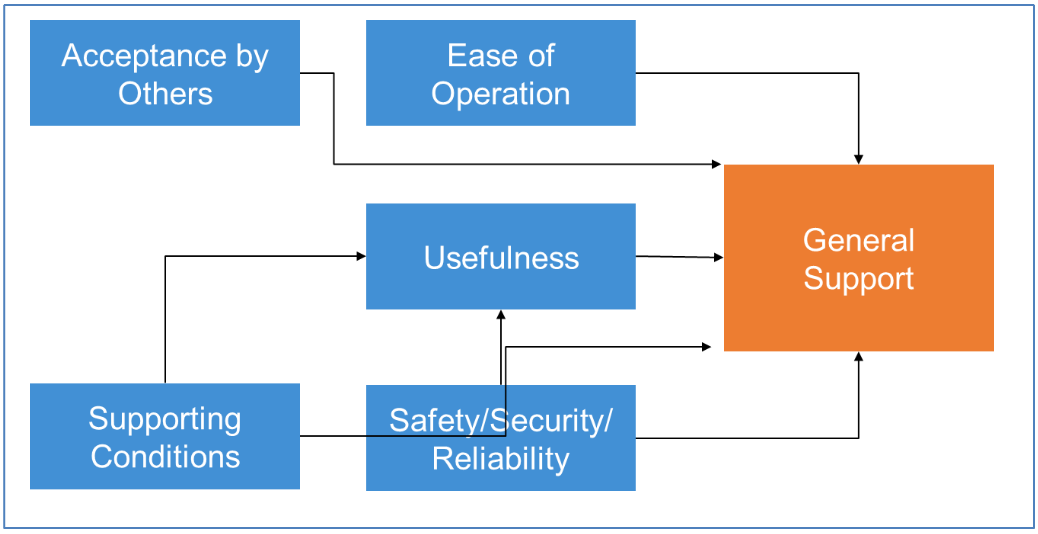

Figure 17. The Automated Road Transport Logistics Acceptance Model (ARTLAM) developed for the requirements analysis [2]

This framework was developed within WP2 (see D2.2) and a large set of qualitative data has been collected. Within WP7, more data will be captured, to increase the sample size and to be able to increase validity of and to refine the model.

The survey will be distributed widely, with the support of WP9 and the networking partners IRU, BizUP, CARA, CEREMA, and ENIDE. The wide distribution of this survey will contribute to achieving the KPIs foreseen in Objective 1 of the project.

[1] This description has been originally made by the project team in [2].

6.1.4. Timeline

The timeline for the three different activities within the user and stakeholder evaluation stream is shown in table 8.

Table 8. Timeline of User and stakeholder acceptance evaluation

Activity within the user and stakeholder evaluation stream	Start month	End month
1. Evaluation of HMI design for AGTS fleet management	13	24
2. Contextual evaluation of AGTS acceptance at the pilot site	25	36
Pilot 1: Automated forklift	33	36
Pilot 2: Hub-to-hub shuttle service	24	36
Pilot 3: Airport baggage tractor	17	36
Pilot 4: Port operations	29	31
Investigation of overall user and stakeholder acceptance of AGTS	13	33

6.2. Safety impact assessment

6.2.1. Introduction

Safety impact assessment will address the *potential of the tested automated systems to help avoid accidents and injuries that occur in similar manual operations*. As the operations and tests take place in industrial areas, the safety benefits of the systems will be reflected upon occupational safety statistics of moving work machines.

While certain route segments in the tests will include also public roads, AWARD's tests target a specific set of routes on and close to industrial sites, and how risks will be avoided there. There may be specific customizations in use, such as infra changes, communication or traffic control. The evaluation will not address generic automated vehicles driving on public roads, as there are other research projects studying that. AWARD will evaluate how risks are being avoided in selected scenarios, processes and locations, in an industrial setup.

The most dangerous accident types, according to current statistics, are a person getting hit by or being crushed by a vehicle: for example, a vehicle has tipped over, or the person got caught between two vehicles. Some more common accidents include hands, arms or clothing getting trapped and machine maintenance-related accidents.

Thanks to numerous existing safety protocols, workplace accidents and injuries are rare events. According to the ESAW (European statistics on accidents at work) database, in 2018 the accident rate of fatal accidents was 1.77 fatal accidents per 100 000 persons employed. In the transportation and storage sector there was ca. 4.6 fatal accidents and 2700 non-fatal accidents (at least 4-day absence from work) per 100 000 persons employed [3].

No severe accidents are expected to happen in short operational tests of new automated systems, if precautions have been successfully taken. While we are lucky to have reached this level of safety, that also means that safety improvements of new systems or methods on the number of yearly accidents or injuries cannot be measured directly in small-scale operations, involving but a few systems and workplaces. The safety potential of a new system or method has, therefore, to be assessed indirectly, using secondary factors and indicators that are known to either increase or decrease the likelihood or severity of accidents. Such secondary factors related to automated driving are, for example, changes in driving speed and safety margins (distances and times) that the vehicle keeps to other vehicles and pedestrians. Exposure is another main factor: how much of the time e.g., a human would be near an automated machine, how many events happen daily, how many kilometers are driven, etc. Further, certain types of accidents such as those where driver inattention plays a part, can be reasoned to happen less frequently with automated vehicles.

The link between accidents and such secondary factors is not usually direct. Many factors have to be considered in an assessment. In some cases, a statistical link has been established in past studies between a dangerous behavior and accidents. If such behavior can be minimized, safety is expected to improve. Advanced new simulations could also show that certain types of past accidents will be prevented from now on thanks to an improvement. More often than having such relatively direct tools, safety potential must be analysed and compiled piece by piece, scenario by scenario, building up from small effects and making an expert assessment on their relevance. The process of this stepwise expert assessment towards yearly fatalities and injuries must be kept very transparent, so that the readers can follow all steps and assumptions. In that way, if a certain assumption, effect or statistic is later found to be missing or misleading, as new information comes up, the impact assessment can be later elaborated in follow-up work.

While there are traditional methods both for risk assessment of industrial machines and safety impact assessment of new passenger car and truck safety systems, AWARD will have to combine methods, when initiating evaluation of new type of automated industrial trucks. The current industrial automated guided vehicles (AGVs) generally move indoors and use walking speeds. AWARD demonstrators are planned to use higher speeds and outdoor-capable sensing.

Safety impact assessment comes after development-phase steps about ensuring and validating safety. Especially the ongoing strong development of simulation methods to assess and validate the safety of automated features benefits safety impact assessment by giving new input to assess detailed benefits in specific dangerous scenarios. Detailed risk assessment and management during system take-up is another source of input: what is believed to be the remaining risk level, after precautions have been taken.

It is the role of earlier safety validation to act on the immediate issues such as sensor problems. Safety impact assessment, in turn, considers final effects and accident likelihoods statistically, after an introduction of a reasonably well-working system. Safety impacts are considered on the level of an industrial site and on similar sites. Further, the calculation can continue on industry-wide and societal levels, if statistics exist on accidents and number of similar vehicles in use.

Safety impact assessment addresses long-time operational use and changes there, and in final calculations it usually assumes that certain prototype glitches can be ironed out. It is the role of field operational tests to find out the final issues so that they can be fixed before large-scale introduction.

Safety impact assessment usually gives an estimate of maximal potential of a technology to reduce accidents, if all such operations and vehicles would be automated. Realistic negative effects of a final system are estimated based on the findings from the prototypes. For automated vehicles, such a negative effect might result e.g. from braking suddenly without no visible reason. The safety impact assessment commonly also addresses smaller, intermediate steps: what changes are foreseen in given timeline, e.g., after 5 and 10 years.

Scenario-based approach is the most common tool when assessing the value of new automation or driver support systems for vehicles: common scenarios/situations appearing in accident databases are brought up and then the new system's benefits are assessed against such groups of situations and their parameters. For example, a new warning system might show potential to prevent certain type of accidents in the dark but provide no significant improvements in daylight conditions. Similarly, the tested automation in AWARD addresses and is proven on certain type of routes and situations only. Therefore, when the results will be scaled up, it will not be about "if all ground vehicles at airports would be automated", but firstly, which percentage of operations would realistically be automated with the technology at hand.

The traditional methods to gather proof on safety benefits include operational/pilot tests of different lengths to collect driving data, human observation techniques for assessing e.g., driving style, and interviewing individuals and more generally stakeholders about changes. In AWARD, all these will be used to identify what starts to change in the currently human-driven operations with self-driving vehicles.

6.2.1.1. Manual operations and human drivers as a baseline

Automating driving brings generally faster-than-human reactions to sudden situations and tireless monitoring of nearby objects. Automated vehicles are usually able to keep their defined safety margins to avoid accidents. It is also common that these safety margins are larger than what human drivers might choose and the driving behavior more careful.

As the Annex I about accident statistics points out, automated guided vehicles (AGVs) have been proven to considerably improve safety over human drivers. However, the comparison against human drivers is not a simple one: not all freak accidents can be avoided even with automation (wrong maintenance, sabotage, drunken behavior and suicides), and even some new accident types may arise with the take-up of new systems. Further, the tasks and routes of a human driver are often much more complex than those of AGVs.

Probably the main shortcoming of any automated system, when comparing against experienced human workers, is that computers may have not been programmed to monitor or understand all complex situations. For example, the algorithms in use may or may not cover certain extreme weather conditions or rare obstacle types on the route. However, the emergency types that have been pre-programmed, e.g. a power outage, automation generally handles correctly and fast, competing well with human operators.

Therefore, when comparing the safety of self-driving vehicles against human drivers, in complex operations, automation does not necessarily bring just benefits – automation could theoretically also introduce new types of accidents that are currently rare in manual operations. Both the positive and negative safety effects have to be assessed to reach realistic impact estimates. Issues in interaction between humans and automated vehicles could result in such new accidents.

As a real example, a few years back, a worked got crushed in the U.S. right after removing an obstacle that prevented a driverless forklift from continuing [4]. From machine safety perspective, such maintenance-related accidents are actually a common type and it's often the maintenance man who has not shut down the machine correctly before maintenance. Automated driving systems with their multiple environmental sensors bring new possibilities to detect wrong use of a system, but incorrect use and maintenance will likely remain a source of accidents.

Complex and rare situations during operations are addressed in system design by safety strategies. Such strategies with self-driving vehicles define e.g. when to stop movement and safety margins to keep to other vehicles and humans. When the safety margins are chosen very conservatively, automated systems stop whenever humans come in the area, far before any danger. Such a careful approach would provide near-absolute safety. However, stopping very early can also reduce efficiency and limit fluent human-machine interaction. Fences could be designed around the machine and so forth. Therefore, new safety strategies target enabling certain human interaction.

Safety margin calculations can be based on walking speeds of humans and observing their limb movement with sensors, instead of using absolute maximum running speed in the formulas. For example, ISO 13855 and ANSI B11.19 define moving speeds of humans' upper limbs around 2 m/s for industrial safety systems. Such values actually leave room for human responsibility: it may be possible to cause an accident deliberately, for example, by running recklessly. Naturally this is the case with most systems, no matter how high fences have been built around a moving machine. The new targets for interaction call for better environmental sensing and more safety considerations and validation steps than in the fenced approach. While it may still be possible to deliberately crash, all common accidents and falling and tripping should be considered in safety calculations. Humans need to be confidently detected with monitoring systems.

Actually, the aforementioned industry standards rule out humans accidentally falling and tripping in their maximum speed values. Such topics may have to be revisited with the introduction of new automation – coming up with a rule set and methods for estimating human behavior and movement is a difficult task.

In automated driving, Mobileye has introduced the concept Responsibility-Sensitive Safety (RSS), where all road users would be assigned with reasonable maximum and minimum reaction times and acceleration values. For example, if a driver - approaching a vehicle that had to stop in front in an intersection – would not start to react within two seconds (a human normally reacts in one second) and carry out minimally a light braking, RSS-like calculations would point out that the driver who did not react in time – and not the stopped vehicle – caused the accident. However, such discussion on reasonable mathematical behavior has barely just started in the automotive industry and with legislators. Current traffic rules are not always easy to define mathematically: what does it mean to "keep a safe distance"? Still, the RSS seeks realistic extremes for human behavior, just like the industrial standards above.

The safety impact assessment in AWARD builds on WP4 safety validation work that evaluates correct driving behavior in different close-call scenarios; when driving close to other road users or workers. The RSS-like principles can be used to determine success criteria for tests and also help to assess approximate risk levels (based on safety margins) that the human drivers and automated vehicles take in logged driving. Safety margins can be calculated and defined for the automated vehicles to minimally keep in different scenarios, so that for example a human could not accidentally walk under the vehicle. Experiencing smaller than the defined values during tests would show a heightened risk of an accident and somewhat uncontrolled/unexpected situation. Such situations could be also manually reviewed.

The current best practice in road safety research is to detect crash-relevant events such as close calls by using certain search criteria such as hard braking events, then to classify these events manually using video. The frequency of such situations with an automated system can be compared against a human baseline. For more details on the technique, see FESTA [1]. This is also a leading principle in the AWARD project, to examine changes in driving style: changes in the number of crash-relevant events and levels of safety margins.

Human drivers rely on their interpretation and other capabilities and generally operate with smaller safety margins than today's automated systems. The small risks that they take will not usually result in accidents. The number of accidents per millions of hours of mobile work machine operations is small. However, automation is expected and required to do even better.

6.2.2. Research questions

Table 9. Safety RQs

ID	Research Question	Clarification	Priority
SA-1	What are the foreseen accident types in different operational modes of automated trucking on industrial grounds?	Review different accident types in different phases of piloted transport operations: e.g. loading, interactions with ground workers. Expert assessment and observation, comparison against current accident statistics.	High
SA-2	How many and which types of occupational accidents, injuries and diseases could be prevented through automated trucking?	To examine statistical safety potential in similar industrial operations and scenarios than in the tests	High
SA-3	What are the changes in material damages, when comparing manual operations with automated operations?	Collect information on material damages and small accidents that include no human injuries	High
SA-4	What is the frequency of safety-relevant events during automated vs manual transport operations?	Detect and analyse close call situations (low time-to-collision, maximal braking) and compare safety margins generally used while driving	High
SA-5	How reliable automated prototype vehicles prove to be during operational tests?	Analyse human take-over actions, unexpected stops, undetected objects, nuisance alarms and similar.	High
SA-6	How ground workers and maintenance men view safety, reliability and trustworthiness of the new vs old operations and interactions?	Conduct interviews, study how much concentration working safely with automation requires.	High
SA-7	How does the interaction between vehicles change in a mixed fleet?	Study the interactions between automated and manually driven vehicles, e.g. surprising situations, overtaking, queues. Interview drivers and measure changes also in manual driving.	High

ID	Research Question	Clarification	Priority
SA-8	How does the risk management process change at the test site, when taking self-driving vehicles into use?	Examine changes in risk assessment and mitigation methods, incident reporting processes.	Medium
SA-9	Assess the difference in safety related events and related prevention strategies on industrial premises vs public roads	Compare different road segments. Compare AV operations on different road segments. Finally, compare industrial processes and current traffic safety research.	Medium

6.2.3. Data needs

Table 10. Safety data needs

ID	Data need from tests	Purpose	Related RQs
1	Workplace operational statistics	Compare key operational indicators about safety (and efficiency, environment) before and after AV introduction	SA-2, SA-3, SA-6
2	GPS logs from baseline and test vehicles: 2 Hz time, coordinates, speed, heading and number of satellites	Trip count, analysis of speed, stops and harsh accelerations. Also of interest: how others drive near automated vehicles (overtaking and such)	SA-5
3	Longitudinal and lateral vehicle acceleration data in 2-10 Hz (optional)	To support GPS log analysis	SA-4
4	Timestamped video from the vehicle to the front	Manual analysis of stops, hard braking and low time-to-collision events	SA-4, SA-5, SA-7
5	AV: continuous data of the closest object in the front, near the vehicle trajectory: lateral and longitudinal distance, object type	Search of close-call and interaction situations with given proximity criteria	SA-5, SA-8
6	AV: Number of safety operator take-overs, minimum risk maneuvers or unexpected malfunctions	To analyze current reliability and the needs for support and operational changes. Vehicle logs including video and paper test diaries.	SA-5, SA-6
7	Diary of material damages during baseline (or site statistics) and automated driving operations	To analyse the potential differences regarding material damages that include no human injury	SA-3
8	AV: Operational mode or route segment	To separate different operations such as loading and straight driving in statistical analyses	SA-2, SA-7

6.2.4. Evaluation plans and methods

The safety evaluation team will review the changes in driving style and safety margins, when comparing log data from human-driven vehicles against data from automated vehicles. In addition to generally used distances and timings, the nature and number of possible near-miss and emergency stop events are of interest. These changes will be analyzed to establish likely percentual changes in different types of accidents. The percentual changes will be reviewed against accident statistics.

The project has been able to access confidential (but anonymous) airport incident reports to review them from automated driving perspectives. The traditional incident categories do not provide enough separation to consider potential safety improvements of different driving behavior in detail. Therefore, the project will consider new categories for the statistical data. The opportunity to use detailed new categories could help to better assess the safety potential of tested new systems.

In addition, the safety evaluation team works in collaboration with the safety validation team, having first identified critical route segments and different risks. After risk mitigation techniques have been selected and final vehicle behavior programmed, safety validation will ensure that the vehicles are able to carry out e.g. emergency braking actions and detect objects from specified minimum ranges. The safety evaluation can then continue to consider the changes in different risks, statistically, but using the comparison between humans and automation.

Some risks may have been mitigated by means of installing new infrastructure support such as a new gate, a traffic light or a fence. Such infrastructure could improve the safety for human drivers, as well. If the infrastructure changes have not been previously considered for human operations, such safety improvements could be seen as benefits of the automation. Such benefits obviously come with extra cost.

Finally, the safety evaluation team has started drafting and listing different impact mechanisms for the safety assessment of automated goods transport trucks. These mechanisms are to give structure for the assessment and calculations. The mechanisms systematically guide to take into account different aspects such as direct and indirect changes. The mechanisms will be loosely based on the widely used 9 mechanisms of road safety studies. The latest update regarding automated vehicles has been published in Trilateral Impact Assessment Framework for Automation in Road Transportation [5].

To give examples, such systematic assessment should cover areas such as:

- Modification of exposure (more or less driven kilometers) and route risk categories (safer routes)
- Modification of driving style of the vehicle when compared against a baseline
- Modification of accident consequences due to different vehicle design
- Changes in interaction with other vehicles and machine operators.

6.3. Process efficiency and quality evaluation

6.3.1. Introduction

In addition to the safety of automated ground transport systems (AGTS), the impact of AGTS on process efficiency and quality is a key evaluation aspect within the AWARD project.

Andrejić et al. [6] review related work on measuring transport efficiency and they distinguish between two basic aspects of measuring transport efficiency: (i) fleet efficiency and (ii) vehicle efficiency. These two aspects will also represent core evaluation aspects within the AWARD project. Therefore, certain research questions related to the two aspects as well as measures and means to support answering the questions will be outlined in the subsequent sections.

In addition to the two core aspects, the efficiency of handling of goods has been identified as relevant by the AWARD project partners. Automation may introduce changes in handling procedures, e.g. at hubs, and thus affect the efficiency or quality of handling procedures. For this reason, the AWARD project partners included this aspect within the initial evaluation design.

In the following section, research questions related to the evaluation aspects (figure 18) are:

- fleet efficiency
- vehicle efficiency
- efficiency of handling of goods.

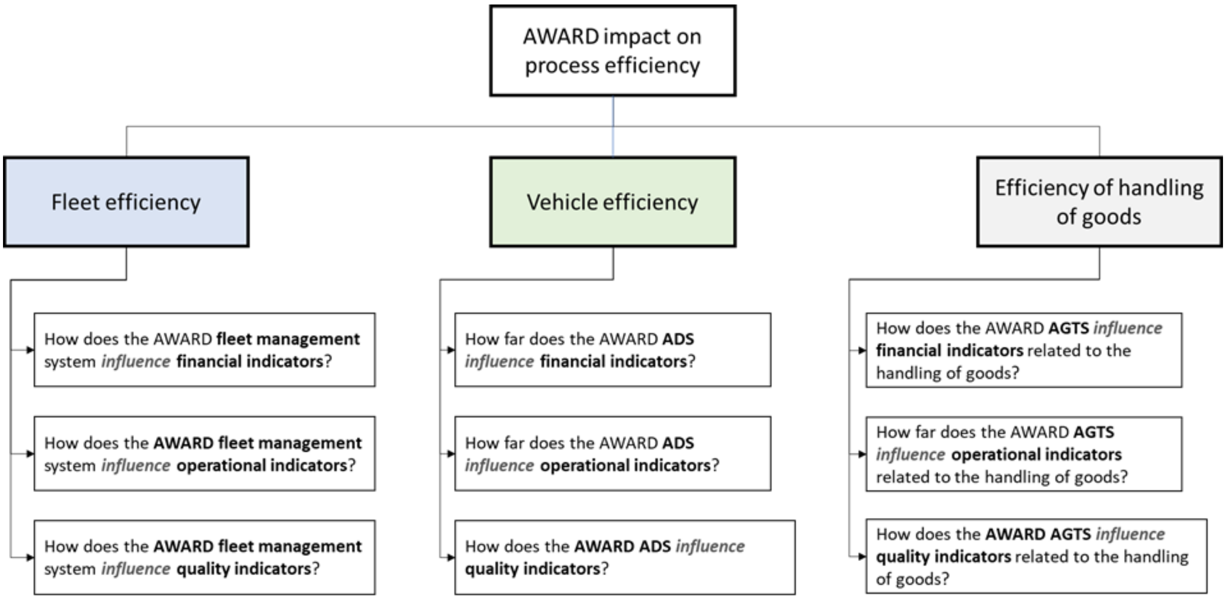

Figure 18. Overview - process efficiency evaluation design

6.3.2. Research questions and hypotheses

Following, the research questions sketched above are detailed in terms of hypotheses and their priority within the project. The subsequent prioritized list of hypotheses (table 11) represents the outcome of the initial evaluation design and serves as basis for detailed evaluation activities related to the AWARD use cases.

Table 11. Process efficiency research questions

ID	RQ	Hypothesis	Priority
Fleet Efficiency			
EF-1	How does the AWARD fleet management system influence financial indicators?	The FMS reduces fuel costs	High
EF-2		The FMS reduces total costs per kilometer	Medium
EF-3		The FMS reduces costs for spare parts	Low
EF-4		The FMS reduces labor costs	Medium
EF-5		The FMS reduces maintenance costs	Low
EF-6	How does the AWARD fleet management system influence operational indicators?	The FMS increases vehicle utilization	High
EF-7		The FMS increases the amount of shipped goods	Low
EF-8		The FMS minimizes the distance driven	Medium
EF-9		The FMS minimizes the number of vehicle breakdowns	Low

ID	RQ	Hypothesis	Priority
EF-10	How does the AWARD fleet management system influence quality indicators?	The FMS minimizes the average maintenance downtime	Low
Vehicle Efficiency			
EF-11	How does the AWARD ADS influence financial indicators?	The ADS supports reducing personnel costs	High
EF-12		The ADS increases purchase costs	Low
EF-13		The ADS decreases costs of vehicle operation	Medium
EF-14	How does the AWARD ADS influence operational indicators?	The ADS reduces net transfer time	High
EF-15		The ADS reduces net waiting time	Medium
EF-16		The ADS increases vehicle uptime	Medium
EF-17		The ADS decreases mean time between failures	Low
EF-18		The ADS decreases personnel time to support (AD) vehicle while driving	High
EF-19		The ADS decreases personnel time to support (AD) vehicle in unexpected situations (breakdown, accidents...)	Medium
EF-20		The ADS decreases personnel time to maintain (AD) vehicle	Low
EF-21		The ADS increases transport capacity	Low
EF-22		The ADS reduces fuel consumption	High
EF-23		The ADS increases vehicle range	Low
EF-24		The ADS decreases vehicle speed	High
EF-25		The ADS requires tighter maintenance intervals	Low
EF-26		The operational availability of the ADS (with respect to varying environmental conditions) is lower than the availability of a manually operated vehicle	High
EF-27	How does the AWARD ADS influence quality indicators in operations?	The ADS decreases the number of damages of transported goods	Low
EF-28		The ADS increases the timeliness of transport orders	High
EF-29		The ADS reduces transport time	Medium
EF-30		The ADS reduces transport costs	Medium
EF-31		The ADS increases the transport reliability	High
Goods handling Efficiency			
EF-32	How does the AWARD AGTS influence financial indicators related to the handling of goods?	The AWARD AGTS reduces personnel costs for handling of goods	Low
EF-33		The AWARD AGTS increases purchasing costs for supporting logistics systems	Low
EF-34		The AWARD AGTS increases costs for supporting logistics systems operation	Low
EF-35	How does the AWARD AGTS influence operational indicators related to the handling of goods?	The AWARD AGTS application reduces net waiting times for goods handling	Low
EF-36		The AWARD AGTS application decreases personnel time to support goods handling	Low

ID	RQ	Hypothesis	Priority
EF-37		The AWARD AGTS application decreases personnel time to support goods handling in unexpected situations	Low
EF-38		The AWARD AGTS application decreases personnel time to maintain goods handling (logistics support) systems	Low
EF-39		The AWARD AGTS application decreases inventory size	Low
EF-40	How does the AWARD AGTS influence quality indicators related to the handling of goods?	The AWARD AGTS application increases timeliness of handling of goods	Low
EF-41		The AWARD AGTS application reduces (un)loading time	Low
EF-42		The AWARD AGTS application reduces costs for (un)loading	Low

6.3.3. Data needs

Subsequently, an initial draft of the data to be collected to answer the research questions is given (table 12). For each data need, data within the baseline situation as well as within the AWARD AGTS application need to be collected. In table 12 this is referred to as "As-Is Situation" and "To-Be Situation". This table represents an initial list, which will be tailored to the different AWARD Use Cases depending on the accessibility of data as well as the possibility to collect data during the project duration. The color coding of the table refers to the evaluation aspects fleet efficiency (blue), vehicle efficiency (green), and efficiency of handling of goods (grey).

Table 12. Process efficiency – data needs

ID	Data Need Name	Description	Data needed	Mapping
EFID1	Vehicle-related personnel costs	This indicator refers to costs originating from vehicle drivers as well as from personnel that maintains the vehicles used. Cost rates are required that serve as basis to estimate costs for operating manual and automated vehicles.	Personnel cost related to use cases; statistics of national personnel costs in Europe	EF-11
EFID2	Costs of vehicle purchase	This indicator refers to estimated costs for automated ground transport vehicles as well as for ground transport vehicles that are currently in use.	Estimated costs of purchase by OEMs; current purchasing costs	EF-12
EFID3	Costs for vehicle operation	This indicator comprises cost categories such as: - fuel costs - insurance - vehicle wear - vehicle maintenance costs (e.g. parts) - unexpected repair costs due to accidents	As-Is operation costs within use cases, operation costs within demonstrations, projection of potential future operational costs	EF-13

ID	Data Need Name	Description	Data needed	Mapping
EFID4	Net transfer time	This indicator refers to the driving time of a transport vehicle from a given origin (O) to a given destination (D). The Net transfer time does NOT include time dedicated to handling of goods or administrative tasks. The indicator may serve to compare if an AV is able to drive from O to D within the same amount of time as a manually driven vehicle in the same environment.	As-Is transfer time of manually operated transport vehicle, transfer time of automated transport vehicle in given use cases	EF-14
EFID5	Net waiting time	This indicator refers to waiting times that may occur along a transport route, e.g. at gates, doors, waiting times due to broken vehicle that needs to be overtaken.	As-Is waiting times, waiting times of automated vehicle in given use cases	EF-15
EFID6	(AD) vehicle uptime	Uptime is a measure of system reliability, expressed as the percentage of time a machine has been working and available. In general, uptime is the opposite of downtime. In the trucking industry, uptime is defined as the vehicle being available to perform its intended function – hauling freight.	Uptime in As-Is Situation, uptime of automated vehicles in use cases	EF-16
EFID7	Mean Time between failures	Mean time between failures (MTBF) is the predicted elapsed time between inherent failures of a system, during normal system operation. MTBF can be calculated as the arithmetic mean time between failures of a system. The term is used for repairable systems, while mean time to failure (MTTF) denotes the expected time to failure for a non-repairable system [7].	Mean time between failures in As-Is Situation VS future mean time between failures	EF-17
EFID8	Personnel time to support (AD) vehicle while driving	This indicator refers to personnel time dedicated to operators who are in charge of monitoring, teleoperating or even manually driving a vehicle.	As-Is personnel time dedicated to driving VS time within automated situation	EF-18
EFID9	Personnel time to support (AD) vehicle in unexpected situations	This indicator refers to the amount of personnel time required to solve unexpected situations such as breakdowns or accidents.	As-Is personnel time dedicated to support vs To-Be time within automated situation	EF-19
EFID10	Personnel time to maintain (AD) vehicle	This indicator refers to the personnel time required for maintenance tasks.	As-Is personnel time dedicated to maintenance vs time within automated situation	EF-20

ID	Data Need Name	Description	Data needed	Mapping
EFID11	Transport capacity	Transport capacity refers to the weight or volume of the load a transport means can carry under certain conditions.	As-Is transport capacity VS transport capacity of automated vehicle	EF-21
EFID12	Fuel consumptio n	The amount of fuel consumed when driving a given distance. It may refer to different fuels such as diesel, gasoline, electricity, gas...	As-Is fuel consumption vs fuel consumption of automated vehicle	EF-22
EFID13	Vehicle range	The vehicle range refers to the distance a transport vehicle can travel before it needs to be re-fueled.	As-Is vehicle range VS range of automated vehicle	EF-23
EFID14	Vehicle speed	Vehicle speed shall be monitored and analyzed using multiple indicators such as maximum speed the vehicle may reach (theoretical), maximum speed a vehicle performed during operations, average speed, …	As-Is speed VS speed of automated vehicle	EF-24
EFID15	Mainte-nance intervals	The frequency at which a particular maintenance task is typically performed	As-Is VS future maintenance intervals	EF-25
EFID16	Operational availability of ADS	This indicator refers to the operational availability of the ADS with respect to varying environmental conditions. The availability is tightly coupled to the uptime of an ADS. However, this additional KPI shall take into account varying environmental conditions. Within the project proposal an availability higher than 99.25% is stated	As-Is VS future availability	EF-26
EFID17	Damages of trans-ported goods	This indicator refers to the ratio between goods transported and the number of damages	As-Is ratio VS expected future ratio VS ratio within demonstration	EF-27
EFID18	Timeliness of transport orders	This indicator refers to the number of delays per day/week/month related to number of transports (e.g., 1 transport per week out of 80 Transport operations is delayed in average for 20minutes)	As-Is timeliness VS timeliness in automated transport solution	EF-28
EFID19	Transport time	This indicator refers to the real duration of the transport of goods from a certain origin to a destination. Geographical constraints such as weather or technical limitations (e.g., operational speed) have a direct impact on transport time [8].	As-Is VS demo VS future estimation	EF-29
EFID20	Transport costs	Costs related to transport operations, e.g. per loading unit	As-Is VS demo VS future estimation	EF-30
EFID21	Reliability of transport	The certainty that a transport order may be conducted within the expected time frame (schedule).	As-Is VS demo VS future estimation	EF-31
EFID22	Fuel costs of transport fleet	The costs per day/week/month/year for fuel.	As-Is VS demo VS future estimation	EF-1

ID	Data Need Name	Description	Data needed	Mapping
EFID23	Total costs per kilometer	This indicator considers personnel costs, costs for transport vehicles, vehicle operations and maintenance.	As-Is VS demo VS future estimation	EF-2
EFID24	Parts and labor costs	This indicator considers costs for spare parts and labor costs related to maintain/repair transport vehicles. Furthermore, labor cost to operate the fleet management system shall be measured/estimated.	As-Is VS demo VS future estimation	EF-3, EF-4
EFID25	Costs for mainte-nance manage-ment and downtime prevention	This indicator refers to personnel costs that occur due to managing the maintenance of transport vehicles and preventing downtimes	As-Is VS demo VS future estimation	EF-5
EFID26	Total distance driven	This indicator refers to the total kilometers a transport fleet requires to perform certain transport tasks within a timeframe. Fleet Management optimization algorithms may reduce the distance driven and improve the usage of transport capacity.	As-Is VS demo VS future estimation	EF-8
EFID27	Shipped goods	This indicator informs about the amount of shipped goods, e.g. tons, number of gitter boxes…, within a certain time frame.	As-Is VS demo VS future estimation	EF-7
EFID28	Vehicle utilization	This indicator provides the ratio between the volume of goods and the vehicle cargo space	As-Is VS demo VS future estimation	EF-6
EFID29	Number of vehicle break-downs	This indicator provides the number of vehicle breakdowns within a timeframe and distance driven.	As-Is VS demo VS future estimation	EF-9
EFID30	Average mainte-nance downtime	This indicator informs about the average time a vehicle requires to recover after a breakdown.	As-Is VS demo VS future estimation	EF-10
EFID31	Personnel costs for goods handling	This indicator comprises costs for: - handling of goods - maintenance of supporting infrastructure/supporting logistics systems	Personnel cost related to use cases; statistics of national personnel costs in Europe	EF-32
EFID32	Costs for supporting logistics systems	This indicator comprises purchasing costs for supporting logistics systems.	use-case specific costs As-Is/To-Be	EF-33
EFID33	Costs for supporting logistics systems operation	This indicator comprises costs for: - fuel costs forklift for unloading/loading - insurance - supporting logistics wear - maintenance costs (e.g. parts) - unexpected repair costs	use-case specific costs As-Is/To-Be	EF-34

ID	Data Need Name	Description	Data needed	Mapping
EFID34	Net waiting time for goods handling	This indicators measures changes related to waiting times of transport vehicles at un/loading points. Doing so, Questions like "How does AV operation change waiting times at docks (e.g. may pro-active notification of workers help to plan/adapt individual task management and reduce waiting times?)" may be answered.	As-Is net time per use case, net-times during demo	EF-35
EFID35	Personnel time to support goods handling	This indicator refers to personnel time dedicated to workers who are in charge goods handling.	As-Is time per use case, times during demo	EF-36
EFID36	Personnel time to support goods handling in unexpected situations (break-down, accidents)	This indicator refers to the amount of personnel time required to solve unexpected situations such as breakdowns or accidents of logistics support systems.	As-Is time per use case, times during demo	EF-37
EFID37	Personnel time to maintain goods handling (logistics support) systems	This indicator refers to the personnel time required for maintenance tasks of logistics support systems.	As-Is time per use case, times during demo	EF-38
EFID38	Inventory size	This indicator shall measure if changes within the inventory size (e.g. of logistics hub, factory,…) might be expected due to automation	As-Is, expected changes due to automation	EF-39
EFID39	Timeliness of handling of goods	This indicator refers to the number of delays per day/week/month related to number of good handles (e.g. 1 handle per day out of 80 goods handling operations is delayed in average for 5minutes)	As-Is versus Demo versus To-Be	EF-40
EFID40	(Un)loading time - Velocity	This indicator refers to the real duration of goods handling procedures -> loading and unloading at hubs, factories and may be stated as average, min and max value.	As-Is time per use case, times during demo	EF-41
EFID41	Costs for (Un)loading	Costs related to (un)loading operations, e.g. per loading unit	Use-case specific costs As-Is/To-Be	EF-42

6.3.4. Evaluation plans, tools and methods

The process efficiency evaluation team will review the changes related to process efficiency and quality within the different AWARD use cases. Thereby, log data from human-driven vehicles against data from automated vehicles shall be compared. Furthermore, data will be collected:

- within interviews with use case partners
- from existing systems at use case partners (e.g. fleet management system, ERP-system, sensors in place)
- via observations (video, on-site observers).

As depicted in the previous section, these data serve to answer research questions especially related to:

- Fleet management efficiency
- Vehicle efficiency
- Efficiency of handling of goods.

Regarding the fleet management efficiency, simulations of the fleet management system (cf. WP5) will represent a vital element to assess fleet efficiency within different scenarios. Therefore, an alignment between the WP5 and the fleet impact assessment will take place to ensure data availability.

Given the comprehensive number of hypotheses in the initial evaluation design and limited project resources, the evaluation activities related to process efficiency will start with focusing on high prioritized hypotheses and continue with lower prioritized hypotheses in case remaining project resources are available.

6.4. Environmental impact assessment

6.4.1. Introduction

The environmental impact assessment will aim at examining and evaluating various dimensions of environmental impact within the bounds of the AWARD project. In recent years, there have been multiple analyses and studies examining the change in environmental impact of automation in transport, often mainly focusing on passenger vehicles ([32] [33]). However, AWARD is focusing solely on the logistics sector. Considering that environmental impact can be local (air pollution, noise), global (GHG, Greenhouse Gas, emissions) or indirect (congestions, land use), the scope of the environmental impact assessment in this project will be quite large and will address the findings on the local, but also to some extent scale up to the global level.

The dimensions of the assessment are divided into six sections, which include impacts on energy, health, greenhouse gas emissions, nuisances and vehicle behavior. This is a wide variety of factors, which will be analysed using data from the initial situation with conventional non-automated logistics as a baseline and the situation during the project implementation. In the context of AWARD, the assessment does not aim at comparing the environmental impact of petrol-driven and electric vehicles, but rather of conventional manually operated vehicles and automated and electrified vehicles, while solely analyzing the usage phase of vehicles.

At first sight, autonomous vehicles will help to decrease the impact of the usage phase through a better optimization of the driving operations and planning.

Even human drivers' efficiency is known to be improved by the usage of ADAS (Advanced Driver Assistance Systems), particularly with speed regulation systems such advanced cruise control (see [9]) or ISA systems (Intelligent speed adaptation, see [10] and [11]). Various studies also reported energy savings from EDAS usage (Ecological driving assistance systems, see [12] for the ecoDriver project results). Even simple speed regulation strategies can help saving fuel (see [13]), while more advanced fuel economy optimization systems can provide both AV's and human drivers with essential guidance about optimal deceleration/acceleration profiles ([14]).

Despite these facts, and considering the recent review from [15], the potential effects of autonomous vehicles (AVs) on greenhouse gas (GHG) emissions are still uncertain. Some factors are decreasing the emissions, while others are increasing it (table 13). According to their findings, "the result shows that eco-driving and platooning have the most significant contribution to reducing GHG emissions by 35%. On the other side, easier travel and faster travel significantly contribute to the increase of GHG emissions by 41.24%".

From the AWARD point of view, the situation is much simpler than the issue of the worldwide adoption of AVs. For example, platooning is not an issue for AWARD use cases, nor are questions linked to acceptance or modal shifts. Nevertheless, some high impact factors may also be relevant for the AWARD use cases. They are indicated in bold in table 13. The environmental impact assessment's task is to capture these factors through test sites experiments and relevant data logging.

Table 13. Potential factors decreasing or increasing GHG emissions for autonomous private cars [15]. Relevant potential factors within the AWARD scope in bold.

Potential factors decreasing GHG emissions	Potential factors decreasing GHG emissions
– Easy-parking – Eco-driving – Eco-traffic signal – Platooning – Vehicle right-sizing – Collision avoidance – Carpooling – Traffic law adherence – Congestion mitigation and efficient routing	– Easier travel – Faster travel – Increased travel by underserved populations – Modal shift from other modes – Increased empty miles traveled – Land-use change

The evaluation team will not go deep into the vehicle's software to know which systems are active or not. Vehicles will be considered as "complex systems" and compared to the human-driven vehicles of similar types. Their global impact assessment method will be the same as for manually operated vehicles.

Direct measures of vehicle's energy consumption through AWARD test sites will be done when feasible, but it does not help so much in measuring GHG or micro-particles emissions. For example, CO_2 emissions are known to be linearly related to fuel consumption, but it is not the case for the other pollutants. In addition, considering most of the vehicles will be electric ones, converting energy use into CO_2 emissions may depend on the local electric mix.

There are two ways to access detailed information on emissions during a field test:

- Install relevant sensors where needed
- Estimate the emissions through surrogate measures and/or literature models.

At the time of writing, both approaches are scheduled within AWARD. Particularly, the environmental impact assessment will rely on vehicle's behavior data logging to extract surrogate measures for safety (see previous sections) and eco-friendly driving behavior.

For clarity purposes, we organize the evaluation according to different dimensions (see also figure 19):

- Energy: This dimension captures the direct consumption of energy, taking into account the efficiency of operations.
- Health: This dimension represents the direct impact on human health through harmful emissions such micro or nano-particles (PM10 levels (ug/m^3); $PM_{2.5}$ levels; NO_x, SO_x, CO, O_3).
- Greenhouse gas emissions: This dimension represents the indirect impact of emissions on the climate change (CO_2, N_2O, CH_4), with moderate immediate impacts on human health.
- Nuisances: This dimension considers other types of environmental pollution, mainly noise, as the others like radio emissions are not measurable within this project.
- Vehicle behavior: This dimension is linked to the direct efficiency of the operations. It measures how smooth or eco-friendly the vehicle is moved.
- Other: This dimension represents the collateral or indirect effects of the operations, such the needs for land use or the impact on traffic jams.

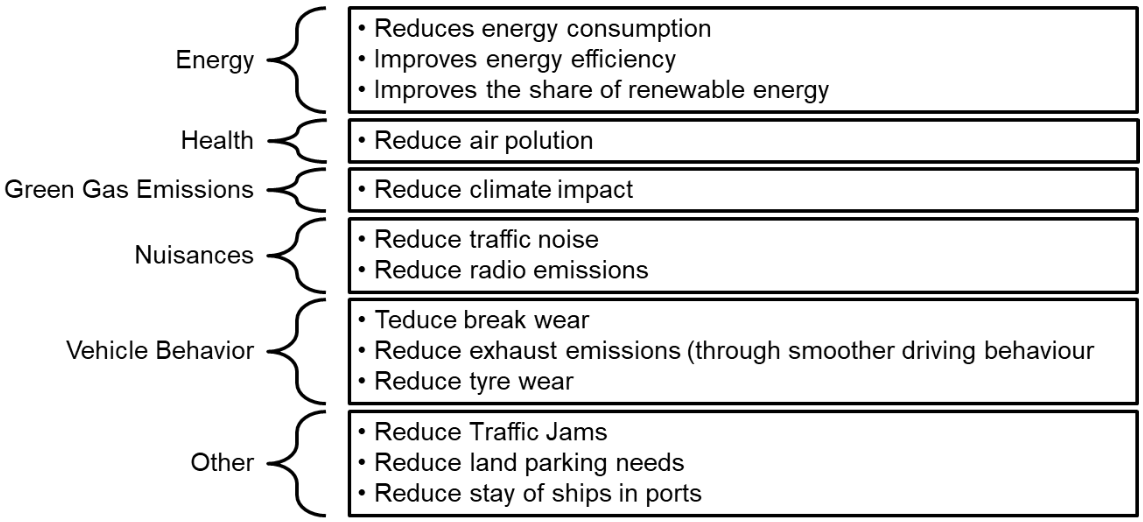

Figure 19. Dimensions of the environmental impact assessment.

6.4.2. Research Questions

The environmental impact assessment team wrote these dimensions into research questions, which in their turn were refined into more detailed research hypotheses. The process result together with the initial prioritization of questions and hypotheses are presented in table 14.

Table 14. Research questions and related refined hypotheses

ID	RQs	Refined hypothesis	Priority
EN-1	What is the impact of Autonomous & electric logistics on energy share & consumption?	Autonomous & electric logistics reduces energy consumption	High
EN-2		Autonomous & electric logistics improves energy efficiency	Medium
EN-3		Autonomous & electric logistics improves the share of renewable energy	Medium
EN-4	What is the direct impact of Autonomous & electric logistics on health?	Autonomous & electric logistics reduces air pollution	High
EN-5	What is the direct impact of Autonomous & electric logistics on greenhouse gases emissions?	Autonomous & electric logistics reduces the impact on climate change	High
EN-6	What is the direct impact of Autonomous & electric logistics on other nuisances?	Autonomous & electric logistics reduces traffic noise	Medium
EN-7	What is the indirect impact of Autonomous & electric logistics on environment?	Autonomous & electric logistics reduces traffic jams	High
EN-8		Autonomous & electric logistics reduces land parking needs	Medium
EN-9		Autonomous & electric logistics reduces the stay of ships in ports and related fuel consumption	Medium
EN-10	What is the impact of Autonomous & electric logistics on vehicle's behavior that could have an indirect effect on the environment & health?	Autonomous & electric logistics reduces brake wear	High
EN-11		Autonomous & electric logistics reduces exhaust emissions (through smoother driving behavior)	Medium
EN-12		Autonomous & electric logistics reduces tyre wear	Medium
EN-13		Local emissions versus global emissions (including construction phase, and recycling phase-life cycle)	Low

6.4.3. Data needs

These research hypotheses need to be tested through experimental set-ups together with a suitable data collection process. This so-called experimental plan needs the input of the desired data needs. This need can be derived from the hypotheses, conditionally to the fact that we can link them to a performance indicator (PI) able to capture the desired factor/behavior. The results of the environmental impact assessment teamwork are proposed in table 15. Each research hypothesis is linked to one or more PI, designed to reflect the desired aspect. PIs have been extracted from literature review or previous or ongoing EU research projects (AVENUE D8.11, MODALES D3.5). Detailed definitions of PIs will be produced later during the project as the exact experimental set-up will be defined.

Table 15. Link between research hypotheses and performance indicators.

ID	Hypothesis	Linked performance indicators
EN-1	Autonomous & electric logistics reduces energy consumption	– volume of fuel (or total energy) consumed per unit distance per unit mass of cargo transported; e.g., l/100 kg·km or MJ/t·km (To be computed from similar period).
EN-2	Autonomous & electric logistics improves energy efficiency	– Distance per vehicle per unit energy; e.g., miles per gallon equivalent (mpg-e).
EN-3	Autonomous & electric logistics improves the share of renewable energy	– Percentage of renewable energy sources (%)
EN-4	Autonomous & electric logistics reduces air pollution	– emissions of air pollutants (Tailpipe, brakes, tires): – PM 10 levels (ug/m3); PM2.5 levels; NOx, Sox, CO, O_3, emissions – In case of unavailability of the adequate sensors, use models instead (from the literature if any exist) or surrogate measures (driving behavior, see EN-11)
EN-5	Autonomous & electric logistics reduces the impact on climate change	– GHG emissions: CO_2, N_2O, CH_4
EN-6	Autonomous & electric logistics reduces traffic noise	– Average traffic noise (dB), noise level, number of people exposed to noise levels
EN-7	Autonomous & electric logistics reduces traffic jams	– Average Traffic queue length per day
EN-8	Autonomous & electric logistics reduces land parking needs	– Total land parking surface
EN-9	Autonomous & electric logistics reduces the stay of ships in ports and related fuel consumption	– Average duration of the parking time (duration of stay in port)
EN-10	Autonomous & electric logistics reduces brake wear	– Deceleration rate of braking (ms^{-2}) – Average deceleration rate of braking – Braking distance – Braking time – Initial speed when braking – Average initial speed when braking
EN-11	Autonomous & electric logistics reduces exhaust emissions (through smoother driving behavior)	– Aggressiveness (% of time in acceleration >0.9 ms^{-2}) – Average acceleration – % of time in speed interval of 20–50 km/h – Average speed – Average driving speed without stops – % of time in deceleration interval of -0.9 to 0 ms^{-2} – Average deceleration
EN-12	Autonomous & electric logistics reduces tyre wear	– Deceleration rate when right braking – Acceleration rate when right accelerating
EN-13	Local emissions versus global emissions (including	– Automation may have collateral positive effects (side effects) that may improve efficiency of the overall system

ID	Hypothesis	Linked performance indicators
	construction phase, and recycling phase-life cycle)	– Rebound effect, improvement of efficiency may increase the demand (like robots working day and night)

Knowing the desired PIs needed to test hypotheses, it is straightforward to derive the data needs (table 16).

Table 16. Environment data needs from the tests

ID	Data need from tests	Purpose	Related RQs
1	Fuel consumption per day per vehicle (or equivalently, energy in kW/h)	To evaluate the energy consumption during operations	EN-1, EN-2
2	Distance traveled	To evaluate the efficiency, that has an indirect impact on energy consumption	EN-1, EN-2
3	Percentage of renewable energy among the energy mix	To evaluate the impact on the use of renewable energy	EN-3
4	PM 10 levels (ug/m^3); PM2.5 levels; NOx, SOx, CO, O_3	To estimate the impact in local air pollution	EN-4
5	CO_2, N_2O, CH_4	To estimate the impact on climate change	EN-5
6	Traffic noise (dB)	To estimate the impact on comfort	EN-6
7	Traffic queue length observed continuously	To estimate the impact on traffic jams (may be relevant for some use cases only)	EN-7
8	Parking needs (m^2)	To estimate the impact on the needed total land parking surface	EN-8
9	Duration of stay in port for each ship	Relevant for the port use case only.	EN-9
10	GPS logs (lat, long, speed)	To compute behavior indicators, that can be used as a surrogate measure of energy consumptions and emissions	EN-10, EN-11, EN-12
11	Can Bus information (systems activated, speed, turning angle for every wheel, blink signal, brake pedal signal, wiper signal…)	To compute behavior indicators, that can be used as a surrogate measure of energy consumptions and emissions	EN-10, EN-11, EN-12

Additionally, some information is required about the operations context, because they can affect the energy consumption or have an impact on driver's & autonomous software behavior. It could be the case for adverse weather conditions, which affect software's behavior, or for the temperature that could induce less energy efficient conditions for engines operations. Such factors will be used as control factors only, and their impact will not be evaluated from the environmental point of view. The list of control factors appears at table 17.

Table 17. Contextual data needs

ID	Data need	Purpose	Related RQs
12	Maximum load capacity, current load	This is used to compute the load factor (%) Transported weight can impact engine energy consumption	All
13	Air temperature	Hot or cold weather may reduce battery efficiency	All
14	Weather	Adverse weather may reduce system efficiency	All

6.4.4. Evaluation plans, tools and methods

The environmental impact assessment plan will be similar to the safety assessment plan. It will adopt a double approach: The first one is to rely on objective measures of the phenomenon of interest (here, direct measure of energy consumed or emitted GHG emissions), the second being an alternative, relying on surrogate measures of this same phenomenon (here for example, quantification of eco-friendly driving behavior).

The evaluation team has made some requests to access internal vehicles energy consumption data. If available, such data will be analyzed under the scope of the FESTA methodology, which aim is to assess statistically the significance of performance indicators deviation from the baseline to the experimental condition (autonomy for AWARD). The observed change will be estimated using suitable statistical models (Generalized multilinear mixed models, GLMM). As autonomy's goal is not primary intended for energy efficiency, the expected effect size is small, but still it can be significant for reasons explained in the introduction.

The basic comparison situation (i.e. the main factor in the upcoming statistical analyses) will be from the baseline condition (human-driven vehicles) to the experimental condition (same vehicles autonomously driven). As usual, external confounding's factors will be controlled as much as possible by gathering information about the operations: weather conditions, load factor, travelled distance. The analysis will focus on test sites where conditions between baseline and experiment will be very similar. That means findings the situations for which autonomy is the only changed factor. Currently, the airport test site is the best candidate.

Even on the same test site with homogeneous external conditions, operations can be viewed as a sequence of scenarios. Usually, data is analyzed per trip, which can then be divided into shorter "chunks" of few minutes. The analysis team will study the operations to extract recurrent situations that could improve the generalization of the results. But due to the relatively small experiment size of AWARD experiments, there will be few possibilities for an inference to larger situations (country size impact assessment). Environmental impact assessment will therefore be done for different scale: the global test site, the trips, and the scenarios.

When considering the environmental impact assessment, energy consumption is only a part of the picture. Indeed, GHG emissions and other micro or nano particles are difficult to measure and are not linearly linked with the energy consumption as it could be the case for CO_2 (CO_2 emissions being linearly associated with fuel consumption). The project is studying the possibility of logging particles emissions while vehicles are operating, but this seems unlikely for practical reasons.

Nevertheless, it is still possible to estimate these emissions, or at least estimate the direction of the change, by relying on surrogate measures: Near-crash situation is a proxy for road safety, as eco-driving is a proxy for fuel/energy efficiency. Hypotheses will explore the change in driving behavior with full autonomy compared to the same operations performed by a human. A literature review and ongoing European projects (MODALES or AVENUE) will be refining the list of PIs best reflecting fuel-efficient driving, and emission-efficient driving (which could be different).

6.5. Technical evaluation

6.5.1. Introduction

Technical evaluation will assess the potential of the automated vehicle to perform defined tasks without or with minimal human assistance. The four vehicles running the use cases of AWARD project (described in the previous chapters) use a very similar set of sensors. The technical evaluation will assess the performance of these components and the decision-making algorithms.

The technical evaluation is planned at two levels (figure 20). The first will be at the level of the test sites. Indeed, the expected behavior of automated vehicles is known for the situations for which they have been designed (description of the operational design domains from WP 2). It is therefore possible to assess the ability of vehicles to react in accordance with what is required of them. Indicators such as false positive rate of object detection, cases requiring human support or needs of teleoperation support will be recorded during field tests in order to compare the actual behavior of the automated vehicle with the expected one.

The second level of performance evaluation is at the level of the vehicle automation systems. For example, the capabilities of a detection system can be evaluated by analyzing the detection range, the angle of resolution of the cameras etc.

The purpose of these assessments is to identify strengths as well as areas for further development. This analysis will be carried out in parallel and in addition to other verification and validation tasks of this AWARD project such as WP3 and WP4. Indeed, the AWARD project aims at the development of automated vehicles performing well in harsh weather conditions. The verification and validation scenarios related to functional safety and SOTIF activities are planned in WP4 and will focus on the ODD parameters. In this WP7 analysis, the focus will be on the performance of the vehicle and its components in long-running operational tests.

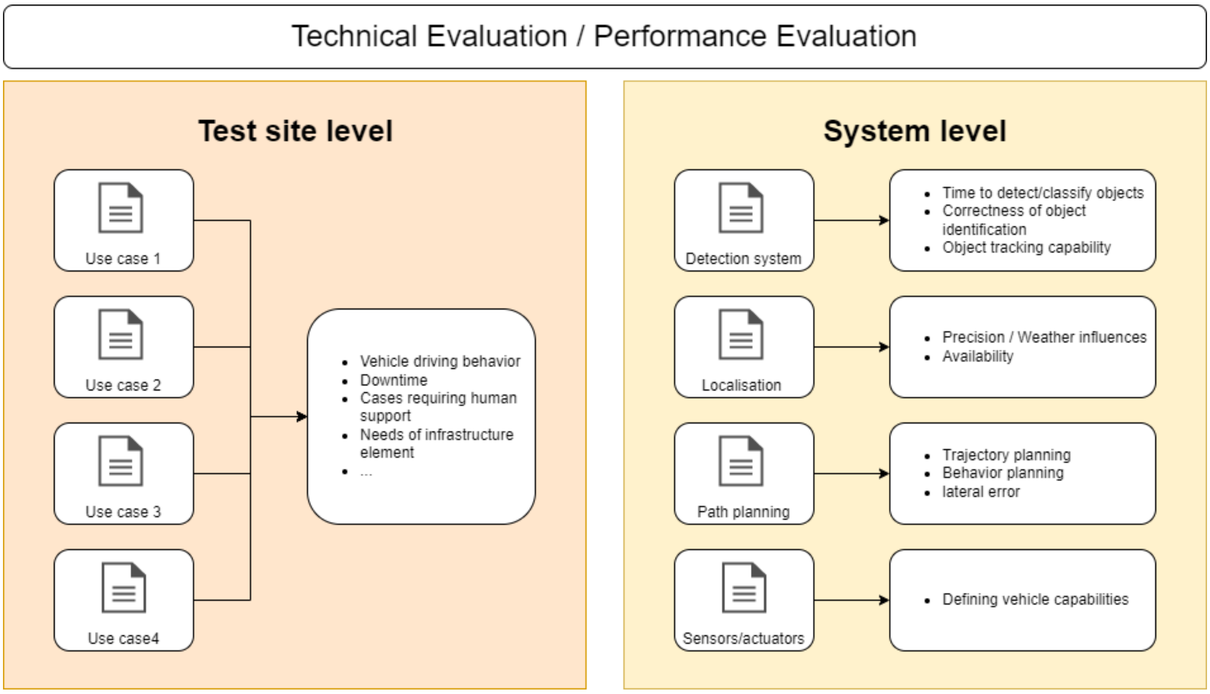

Figure 20. Two levels of evaluation

The following section defines the initial research questions that will guide the next steps in the technical assessment.

6.5.2. Research questions list

Table 18. RQ for technical evaluation

ID	Research Question	Clarification	Priority
Test site level			
TE-1	How often does the prototype vehicle need human intervention?	It is planned that a safety driver will be present during the on-site tests. This issue is related to the number of times the driver will have to take control of the vehicle.	High
TE-2	What is the latency time between the sending of a command by the teleoperation system or an element of the infrastructure and the action of the vehicle?	The vehicle receives data from the entire infrastructure. As the vehicle is in motion, it is important to know the time needed to process the information.	Medium
System Level			
TE-3	How long does it take the perception system to identify and classify an object?	The time needed to identify and classify an object is a parameter that can limit the speed of an automated vehicle.	High
TE-4	How do harsh weather conditions influence confidence in object detection and classification?	One of the main aims of the AWARD project is to evaluate the behaviour of vehicles in adverse weather conditions. This RQ is related to the performance of the perception system.	High
TE-5	How accurately is the automated vehicle able to follow the planned path?	A considerable lateral deviation of the vehicle from the intended path could even be dangerous in the intended use cases.	High

ID	Research Question	Clarification	Priority
TE-6	How accurately and how often is the actual position of the vehicle known?	Key parameter in the evaluation of the performance of a positioning system.	High
TE-7	What is the influence of weather conditions on the availability of the location?	As for question TE-5 but this time linked to the location system	Medium

6.5.3. Data needs

Table 19. Data needs for technical evaluation

ID	Data need from tests	Purpose	Related RQs
1	Activity logs of safety driver	Identify in which situations the safety driver takes over.	TE-1
2	Infrastructure to vehicle latency time	Glass-to-glass (end-to-end) test	TE-2
3	Logs of object classification with time and position	Performance evaluation of perception system	TE-3/4
4	Lateral deviation from expected position	Identify positioning accuracy and path planning capability	TE-5/6
5	Recording of weather conditions during the test phases	Identify weather influence	TE-1/5/7

6.5.4. Evaluation plans and methods

The technical evaluation is quite different from the other evaluation areas in this work package. Indeed, some parameters depend on the maturity of the vehicle development and the time available to carry out the tests. It is possible that some research questions cannot be answered directly as outlined in this document.

However, as described in the previous sections, the main purpose of this technical evaluation is to specify the performance of the vehicles for the following points:

- reliability and confidence in the automation systems;
- infrastructure and FMS capabilities;
- perception system;
- localization system;
- path planning.

These analyses will be carried out in parallel and in collaboration with the activities of WP3 and WP4.

7. Establishing final priority of research questions and data collection

After the evaluation areas had outlined their initial research questions and data needs, which were listed in Chapter 6, the data needs were compiled for starting detailed discussions with each test site. These discussions are currently ongoing.

Test sites comment the feasibility of collecting certain data and the research priorities from their perspective. Each test site has certain operational goals that they aim to reach with automation, and minimally such goals need to be reflected in data collection and evaluation.

Evaluation teams consider the broader picture and the project achieving its evaluation goals. The new technology should be reviewed from several aspects, using traditional topics of each evaluation area.

Priority of data collection and research questions will be established based on the following criteria:

- Operational goals and project plan
- Evaluation team's priorities
- Data collection efforts
- Size of the foreseen effect, considering operations and cumulative effects in the industry
- Relatively new and unknown aspects and contribution to public knowledge.

It is mainly the data collection efforts and project resources that will limit the scope of the work. The final list of research questions will be published in upcoming D7.4.

7.1. Minimum test and data collection durations

Another ongoing topic is estimation of data amounts to be collected and minimum test durations for answering different research questions. These will vary per research question.

Optimally, several research questions would benefit from a year-long data collection period, providing information on performance in winter and summer weather. The longer human operators, maintenance crew and stakeholders work with the systems, the clearer the long-term effects and usage as well as the nature of unintended effects. Additionally, longer data collection periods would allow better insight on rare situations and e.g., emergency stops.

Due to the project being rather a pilot project than being able to carry out year-long tests, the discussion on test durations becomes rather about how to best bring out the effects. For example, tests could take place e.g., in the spring, enabling data collection possibly during several weather types: winter, rainy and dry weather.

Short test periods may also provide a lot of knowledge, if the automation is deeply integrated with production systems. On the contrary, if automation is tested without integration, the normal operations and work practices are not affected.

If we consider the data needs from the perspectives of different evaluation areas, safety is rather about safety margins and irregular events. Large changes in safety margins, when compared against human drivers, can be determined based on rather short tests, since the vehicle behavior is coded and repetitive. Collecting data regarding handling of edge cases is another story: here even a year of driving might not yield a large number of events. Still, tests lasting only a few days could not provide much information on reliability aspects or safety-related traffic interaction events. Minimally a few weeks of data gathering will be needed for an initial assessment.

Regarding efficiency, it is about how much performance fluctuates in different situations and how, for example, weather might affect performance. To gather enough data of various production situations, again, a few days will not suffice. Still, the short trips can offer perspective into how much driving speed and travel times and waiting times change, when introducing automated driving. Even after a few days of operation, data might start to show how frequently safety operators are needed and for how long per stop.

Environmental data collection, focusing much on energy use, has similar data collection viewpoints than the efficiency.

User-related studies hardly even start in the first few days, when the users are learning the systems. Their first impressions might change considerably later on, after work practices and use of the systems come to a new balance. Here, the timeline would rather be weeks or months instead of days. Certainly, safety operator experiences can be collected even after a few days. The number of participants to interview should generally be above 20 to start to consider statistical aspects, and different stakeholder groups should be well represented, raising the required sample size.

The data needs to assess technical performance depend on the focus points. Detailed focus on e.g., positioning accuracy might only be possible on a proving ground. However, longer tests might bring out incidents, where the positioning was lost e.g., due to bad weather.

8. Test data management and processing

The AWARD project will collect vehicle log data along with human observer notes. In addition, the project will interview several users and stakeholders. All this data will be processed from statistical perspectives to generate summaries for evaluation and to assess changes in key performance indicators.

The vehicle log data will consist of time and location, vehicle status information and front view video. Besides gigabytes of video, the amount of other log data is expected to be in the order of megabytes per day, only. Automated vehicles are able to directly log situational information such as the distance to the closest obstacle on its path, and reasons for stopping. Their status such as waiting/mission is also possible to log directly. Logging such situational data instead of raw environmental sensor data should keep the log files rather compact.

In addition to vehicle logs, human safety operators/observers will be requested to mark down (Excel-like) activities such as refueling/recharging the vehicle, reasons for stop/downtime and other specific events that provide either further details or contextual information of tests.

The collected videos along with interviews will contain personal data. The confidentiality of these data types must be protected. No personal data will be published but only statistical averages of how the vehicles performed and anonymized findings on user experiences. Only selected short videos might be published, and regarding these videos, persons are to be blurred or publication permissions sought.

Further details on handling on personal data in the project are discussed in the D10.5 Data Management Plan and in the D1.1 EPQ-H-Requirement (ethical requirements), which also includes user consent form templates.

Figure 21 presents the vehicle data flow in the project. The gray parts are provided by the test site (or their fleet management) and generally include some company-confidential information and personal data. Whereas the blue parts are WP7 data processing components that generate statistical summaries from log data and a harmonized database for evaluation purposes. The database will consist of trip and event summaries in spreadsheet/database format. Minimally the evaluation database should be possible for all project partners to access, as the summaries should contain no sensitive information. Interview and survey data will be handled separately.

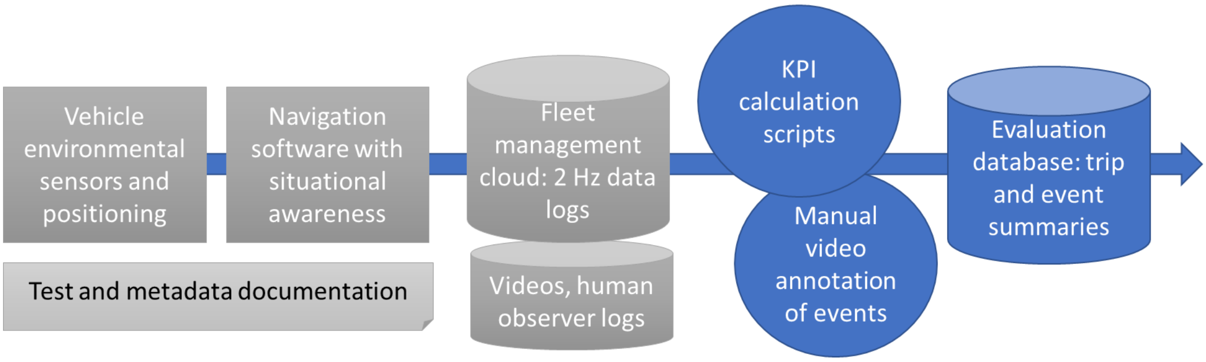

Figure 21. Data chain

8.1. Vehicle log file format

The project will use simple csv (comma-separated values) vehicle log files. Either the log files are provided directly from vehicles or, more commonly, extracted from fleet management telemetry. Each line of a text file will contain a timestamp, vehicle ID, coordinates, vehicle speed and a few key signals. Table 20 outlines the draft log file format as of September 2022.

Table 20. Log file fields, contents of one row

Field	Example
vehicle_id	T69444
vehicle_timestamp	2022-06-03 15:17:12.632
deployment_name	oslo-sas-avinor
localization_geopoint_latitude_wgs84	60.202499177233086
localization_geopoint_longitude_wgs84	11.104863496047077
localization_platform_theta_radian	1.036
localization_geopoint_sigma_latitude_meter	Assumed positioning accuracy, e.g. 0.013
localization_geopoint_sigma_longitude_meter	Assumed positioning accuracy
navigation_speed	0
platform_battery_level	60
navigation_mode	Manual
platform_front_traction_control_mileage	880
platform_blinker	Left
activity_mode (to be confirmed)	Mission 1
stop_reason (to be confirmed)	(navigation system stop reason)
free_distance (to be confirmed)	(free distance to the front)
leg_id (post-processed)	Trip/leg ID from post-processing
event_id (post-processed)	ID of a possible event-of-interest

Instead of collecting extensive amounts of raw sensor data such as lidar point clouds, the log files are to contain processed information, such as free distance or the reason for a stop from navigation software point of view.

The logic of automated vehicles covers status information such as if the vehicle is waiting for something or is on a mission. Selected information will be logged directly to support analysis of high-priority research questions.

Vehicle-based data will indicate moments of interest, or it can even contain errors. Video logs can be used to manually check and classify such moments. For example, cases where the vehicle has stopped abruptly, or objects move within safety margins. The main source of reasons for stopping will be the notes made by safety operators. The planned method to collect these notes is to use a mobile phone user interface, connected to the fleet management.

In some cases, it will be difficult to collect comparable data from human-driven vehicles. The project is unable to instrument older vehicles with more than a GPS logger and a dashcam. This will require video annotation work to compare driving behavior. Such manual annotation requires a lot of work and even so, results can be difficult to compare accurately. Therefore, the general idea is to use the automated vehicle also in manual driving mode, to collect baseline data. Using an automated vehicle to collect data enables easy comparisons and richer data. However, the availability of the automated vehicle is likely to set restrictions for how many days baseline data can be collected.

Test sites are free to collect video/dashcam logs in any standard format. It is recommended to use very low resolution to enable easy handling, small file sizes and partial anonymization, as persons and license plates become difficult to recognize. None of the research questions require high-resolution footage. The video should be timestamped by printing time on the video footage itself to enable easy matching with GPS time.

The logging format will be nearly the same from all test sites, as the data will be collected via fleet management telemetry. Datasets will be supported with manual data collection of selected aspects such as the test supervisor or safety operator marking down a few extra aspects. These notes will be in harmonized csv/Excel/database format as far as possible.

8.2. Documentation about data

Supportive documentation regarding collected data, often referred to as metadata, will be necessary for analysts to correctly process and understand the logs. Signals need to be explained along with their measurement process and values for “not available”.

The analysts will also require test documentation such as the general plans and dates outlined in this document. However, instead of plans, documentation will be needed on final execution: a test diary (table 21) containing accurate dates of different phases of the study. The test site leaders will be responsible over this documentation. This documentation will contain extra notes from the test leader and safety operators. The test diary will clarify situations such as “maintenance on Tuesday” or that “software was updated”.

Table 21. Test diary template

Date	Hour range	Test focus or route	Manual/ Automated	Weather	Road conditions (dry/wet/snow/icy	Notes (for example, stopped early, broken systems, accidents)
1.1. 2023	10–14	Mission 2	Automated	15 degrees, rainy and foggy	Wet	Braking problems, stopped tests early

The intended functionality of the vehicle needs to be clarified/documented as well, along with emergency stopping strategies. This documentation is expected from the vehicle and software makers.

8.3. Data storage and protection

The test site leaders/partners are the owners and managers of the collected data. They will control the access to the data and safeguard personal data. Access to parts of this information will be granted for evaluation purposes. The exact data protection concepts are currently under discussion and will be revisited in upcoming deliverables. It is foreseen that named analysts will get temporary access to the raw data, using a file sharing server. The datasets are to be treated as confidential material under consortium agreement.

Generally, all log data will be processed using common calculation scripts. These scripts are open for all project partners. The results of the calculation – the processed summaries – will no longer contain personal data.

When video clips will be used in project presentations, persons visible will be either blurred or their permission must be sought to publicly use the clip.

8.4. Data processing for evaluation

The project uses a Java-based software to extract trips, events and their attributes from vehicle log data. The attributes of a single trip will include, e.g., trip duration, average speed, number of stops, total time spent waiting (e.g. at an intersection) and many similar coordinate and speed-based values. Some of these attributes will serve as performance indicators to be used in the evaluation. The analysis will check, how the performance indicators differ between human and automated driving.

Safety operator logbooks and manual annotation of collected video will provide further information regarding e.g. emergency stop reasons. Only short periods of video data will be annotated. The initial plan is to clarify interesting events that have been picked up from positioning and environmental sensor data. Mostly these events will be situations where the vehicle has stopped for a longer time than usually, or that other road users are within a certain safety region, while the vehicle is still moving.

Post-processing will benefit from using the vehicle’s status information such as whether it is loading, waiting, or doing something else. Similarly, the route segment is known based on vehicle coordinates. The driving style data can then be segmented accordingly. These will be the operational phases and route segments.

9. Plans for scaling up the findings

The term scaling up is about forecasting impacts of a new technology, if it would be gradually taken into broad use. The initial findings from real tests, such as certain safety or efficiency improvements, are considered against different levels of statistics:

- How would the operations of a single industrial site change, if more/most of their similar vehicles would become automated?
- How might the industry and society change?

Scaling up can consider various scenarios, such as automating 30 % or 100 % of the vehicle fleet. Such scenario estimates can also be repeated and calculated for e.g., 5 and 10 years in the future. As future will bring also other variables and unknowns, rather often the effects are calculated for the present day: how would the current operations change, if the vehicles would suddenly become automated.

When analyzing a single industrial site, there are usually a few key performance numbers such as delays and damages, driven kilometers, and vehicle uptime, through which operational efficiency and safety are being tracked and where the changes could become visible. On European level, there are occupational safety statistics and e.g., numbers of vehicles that are being sold yearly, hours of operation and estimates on emissions.

Traditionally in transport studies, scaling up mainly refers to potential changes in travelling habits, accident statistics, environmental emissions and traffic flow. These changes are reviewed at EU level. The case of automated industrial trucks is somewhat different, as Annex I about safety statistics shows: The history of collecting data and e.g., the accident categories differ greatly. Impact assessment methods from transport studies do not translate directly, as accident and other statistics use different categories. For example, if the research would indicate that turning left at an intersection at an airport has now become safer thanks to automation, it can be difficult to scale up the effects of such a change at EU level, if the closest accident category would be something more general like a “struck or run over”. Several accident types would fall in the same category. Accident statistics from industrial operations do not generally detail driving maneuvers. The AWARD project is looking into options to review more detailed incident descriptions to analyze the safety potential of automated driving.

Scaling up may be easier, when certain accident types are not likely to happen at all with automated operations. Such could be the case with forklifts tipping over due to excessive speed or stability issues, as long as the automated vehicles strictly keep to the planned path. The number of accidents caused by sleepy drivers should drop. Also, certain stressful work tasks might disappear, fully. Such deductive analysis can give some rough estimates of future benefits, if all or more vehicles would become automated.

Now, automated vehicle related statistical data from different industries is scarcely available. The AWARD project must begin its work by focusing on single industrial sites. The potential changes at the level of that factory, port or an airport need to be considered. Scenario-based estimates will be worked on based on collected test data.

Scaling up involves clarifying, how many similar cases and vehicles are currently in use – what is the number of vehicles that could realistically be automated. The number of exactly similar sites and operations in Europe will be a difficult question. First, it must be established, under which conditions automation can be foreseen to happen and which percentages of vehicles are likely to be automated in the next years.

9.1. Future scenarios for airport baggage tractors

As an example of first thoughts into the scenarios for scaling up, we discuss airport baggage tractors. Some aspects are similar for factory areas and ports.

Based on interviews with Norwegian and Finnish stakeholders, one of the main foreseen benefits from having a large number of automated baggage tractors could be that there would be little need to wait for one to arrive. The number of human drivers is obviously limited and occasionally there can be waiting delay involved – although, such delays can be mostly mitigated with good coordination and fleet management. On some airports, delays related to baggage tractors are rare. Such delays related to baggage handling are only reported in case they are large enough to be the main reason for delaying an airplane for leaving. Still, one benefit to consider in scenario calculations could be minimizing waiting delays.

Stakeholders also noted that automation could generally improve fleet management overall and reduce even kilometers driven.

There might be changes in amount of luggage dropping from carts due to speedy driving, if the automated vehicles would drive more carefully and slower.

Instead of using the current routes to transfer baggage, automation might give extra options. The automated tractors could in some cases use longer and more remote routes, as human work time is no longer a factor. The time to deliver baggage must stay within desired window, but other options may open up. However, stakeholders commented also that no alternative routes are available at their airport.

It is mainly the reduction of human working hours and drivers that automation should bring beside better fleet management.

The main operational risk that was noted about an AV getting stuck. Before a human driver would be able to arrive, there could be e.g., a 20-minute delay in aircraft leaving.

10. Conclusion

This deliverable outlined the plans for upcoming operational tests of automated industrial trucks in the AWARD project. Log data will be collected to enable evaluation of operational safety, efficiency and environmental aspects. In addition, technical evaluation will highlight a few select aspects, to complement earlier development-related performance measurements in the project and establish performance in real-world operations.

Even automation tests are not just technical, but a number of user-related research questions will be studied, as well. A series of interviews, surveys and laboratory tests will be run to understand how workers and other stakeholders experience the changes that automated vehicles will bring. Current drivers become safety operators, and in the future, one operator might handle several vehicles.

This deliverable focused on scoping the research questions per evaluation aspect: technical, safety, efficiency, environment, and users. This exercise first was carried out from the perspective of what all research questions the project might be able to answer. Starting from the scoping and initial lists of data needs, the work continues to discuss the priority of the research questions with the test sites.

The final priority of the research questions will be affected by several aspects:

- What goals the different stakeholders have regarding automating operations – that is, what are the numbers and topics where they expect change?
- What does the evaluation team together see as the research questions of highest priority, based on the project plan and targeted operations?
- How much resources are required to collect enough for data to answer a specific research question?
- Size of the foreseen effect, considering operations and cumulative effects in the industry
- Relatively new and unknown aspects and thereby contribution to public knowledge.

Focus is important to be able to answer selected research questions well. In trying to cover too much ground there is the risk of not reaching many answers. Planning and data collection consume a lot of resources but without such scientific rigor, results might remain vague and would only be related to how a few persons experienced the automation experiment and that there is still “work” to do. Whereas operational tests should actually offer the first experimental proof on benefits of automated driving and try to answer clearly, what are the final steps before bringing the technology on the market.

During the next project year and with the update of these plans (as a new deliverable, D7.4), the project will set the final test plans and detail the data collection. Evaluation plans will be extended regarding scaling up: what are the plans and methods to consider what would change, if the industrial areas would be able to update all their trucks. Further, how would the gradual uptake affect similar sites in the Europe.

References

[1] ARCADE project. FESTA Handbook, version 8. 2021. https://www.connectedautomateddriving.eu/methodology/festa/ (accessed 16.12.2021)

[2] Fröhlich, P., Gafert, M., Diamond, L., Reinthaler, M., Neubauer, M., Hammer, F., Koskinen, S. (2021). Towards a Comprehensive Understanding of Stakeholder Requirements for Automated Road Transport Logistics. CEUR Workshop Proceedings, Vol-2905 (2021), ISSN: 1613-0073; Paper-Nr. 3, 7 S.

[3] European Statistics on Accidents at Work (ESAW). 2021. Database, Health and safety at work (hsw). https://ec.europa.eu/eurostat/web/health/data/database

[4] https://lni.wa.gov/dA/0a9c5cdac4/WorkerCrushedByLGVForks.pdfh (accessed 10.12.2021)

[5] Trilateral Impact Assessment Framework for Automation in Road Transportation : Version 2.0. / Innamaa, Satu; Smith, Scott; Barnard, Yvonne; Rainville, Lydia; Rakoff, Hannah; Horiguchi, Ryota; Gellerman, Helena. 2018. 42 p.

[6] Andrejić, M., Bojović, N., & Kilibarda, M. (2016). A framework for measuring transport efficiency in distribution centers. Transport Policy, 45, 99-106.

[7] https://en.wikipedia.org/wiki/Mean_time_between_failures (Accessed 13.12.2021)

[8] https://transportgeography.org/contents/chapter3/transport-costs/transport-time-components/ (Accessed 13.12.2021)

[9] Faber F., et al. 2012. Final results: Impacts on traffic efficiency and environment. Deliverable D6.5 of project euroFOT (FP7)

[10] Carsten, O., Lai, F., Chorlton, K., Goodman, P., Carslaw, D. & Hess, S. 2008. Speed Limit Adherence and its Effect on Road Safety and Climate Change. Report for CfIT and the Motorists' Forum.

[11] Saint Pierre G., Ehrlich J. 2008. Impact of Intelligent Speed Adaptation systems on fuel consumption and driver behaviour. 15th World Congress on Intelligent Transport Systems.

[12] Saint Pierre G. et al. 2016. Eco driving in the real-world: behavioural, environmental and safety impacts. Deliverable 43.1 of the ecoDriver project, EU FP7.

[13] Freitas Salgueiredo C., Orfila O., Saint Pierre G., Doublet P., Glaser S., Doncieux S., Billat V. L. 2017. Experimental testing and simulations of speed variations impact on fuel consumption of conventional gasoline passenger cars. Transportation Research Part D: Transport and Environment, Volume 57, pp 336-349, DOI:10.1016/j.trd.2017.09.029.

[14] Wu, C.; Zhao, G.; Ou, B. A Fuel Economy Optimization System with Applications in Vehicles with Human Drivers and Autonomous Vehicles. Transp. Res. Part D Transp. Environ. 2011, 16, 515–524.

[15] Massar, M.; Reza, I.;Rahman, S.M.; Abdullah, S.M.H.; Jamal, A.; Al-Ismail, F.S. Impacts of Autonomous Vehicles on Greenhouse Gas Emissions—Positive or Negative?. Int. J. Environ. Res. Public Health 2021, 18, 5567. https://doi.org/10.3390/ijerph18115567

[16] OSHA Safety Manual. Forklift fatalities. https://www.safetymanualosha.com/forklift-fatalities/ (accessed 22.12.2021).

[17] Bostelman, R. 2009. Towards improved forklift safety: white paper. In Proceedings of the 9th Workshop on Performance Metrics for Intelligent Systems (PerMIS '09). Association for Computing Machinery, New York, NY, USA, 297–302. DOI:https://doi.org/10.1145/1865909.1865968.

[18] TTK, 2021. Turvallinen trukkityö. Työturvallisuuskeskus. [The Centre for Occupational Safety: Safe forklift work]. Accessed 29.10.2021 https://ttk.fi/files/6428/Turvallinen_trukkityo_27102.pdf

[19] Heikkilä, A.M., Rantanen E., Hämäläinen, V. 2003. Trukkikoulutuksen arviointi ja kehittäminen. (Developement and evaluation of forklift training process.) VTT, Tampere.

[20] Belke, J.C., 2001. Recurring causes of recent chemical accidents. Loss Prevention Bulletin 159, 6–10.

[21] TOT-raportit vuosilta 1985-1998. Kuljetus. Raporttien analyysi ja johtopäätökset. Torjuntastrategia. Toimenpiteet vastaavien työtapaturmien estämiseksi. (‘TOT-reports, 1985-1998. 2000. Investigation of fatal accidents reports. Transportation. Analysis and conclusions. Prevention strategies. Measures to prevent similar accidents at work.’) Finnish Workers' Compensation Center. Helsinki.

[22] Safetyinnumbers. Forklift accident statistics. https://www.safetyinnumbers.ca/wp-content/uploads/2018/04/Forklift-Accident-Statistics-2018.pdf (accessed 22.12.2021).

[23] Occupational Safety and Health Administration. US Department of Labor. Accident report details. https://www.osha.gov/pls/imis/AccidentSearch.search (accessed 10.12.2021).

[24] Stasista. Automated guided vehicle shipments globally in 2018, by type. https://www.statista.com/statistics/882696/global-agv-market-volume/ (accessed 10.12.2021).

[25] GlobeNewswire. Global and China Forklift Industry Report (2020 to 2026). https://www.globenewswire.com/news-release/2020/05/22/2037524/0/en/global-and-china-forklift-industry-report-2020-to-2026-forklift-sales-are-expected-to-drop-as-a-result-of-covid-19.html (accessed 10.12.2021).

[26] European Union Aviation Safety Agency - Annual Safety review 2020. https://www.easa.europa.eu/document-library/general-publications/annual-safety-review-2020

[27] Balk, A.D. 2008. Safety of ground handling. NLR-CR-2007-961. National Aerospace Laboratory NLR. https://www.easa.europa.eu/document-library/general-publications/nlr-report-safety-ground-handling-nlr-cr-2007-961

[28] VTT’s TYKO. Calculation model for non-road mobile machinery. http://lipasto.vtt.fi/en/tyko/index.htm (accessed 10.12.2021)

[29] Standard A, Sabisch M, Kishan S, Fulper C. Measurement and Analysis of the Operations of Drayage Trucks in the Houston Area in Terms of Activities and Exhaust

Emissions. SAE Int J Commer Veh. 2018;11(2):77–92. doi:10.4271/02-11-02-0007. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6390959/

[30] Eurostat. 2021. Transportation and storage statistics - NACE Rev. 2 https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Transportation_and_storage_statistics_-_NACE_Rev._2

[31] EASA European Union Aviation Safety Agency. 2021. Overview of Aviation Sector - Figures and Tables. Figure 1.3. Total flights to or from EU28+EFTA airports (millions). Excel data extracted 5.8.2021. https://www.easa.europa.eu/eaer/topics/overview-aviation-sector/figures-and-tables

[32] Williams E. et al. 2020. Assessing the Sustainability Implications of Autonomous Vehicles: Recommendations for Research Community Practice. Sustainability 2020, 12(5), 1902.

[33] Mazur C. et al. 2018. Comparing the Effects of Vehicle Automation, Policy-Making and Changed User Preferences on the Uptake of Electric Cars and Emissions from Transport. Sustainability 2018, 10(3), 676.

Annex I. Accident statistics related to similar transport operations

This annex gives an outlook into available accident statistics regarding the piloted operations. The AWARD project is looking to gather and analyse further accident data in 2022. The data will be used for analysing the potential future impact that automation could have.

Available statistics

To form an overview on the safety level of transportation sector, European statistics on accidents at work (ESAW [3]) by Eurostat were under review. This database provides annual national statistics on accidents at work with more than 3 days of absence as well as fatal work accidents. Information is collected according the EU-level regulation on statistics on accidents at work, adopted in April 2011. The ESAW dataset is based on administrative sources in the EU-27 Member States. These Eurostat's accident statistics are generally divided according to NACE (Nomenclature of Economic Activities), which is the European statistical classification of economic activities [30].

In general, two types of notification procedures can be identified: Insurance-based schemes, in which the employer and the employee have a financial incentive to report an accident at work, and those based on a statutory obligation. Detailed country comparisons should be avoided due to the differences in reporting, but publicly accessible ESAW statistics are sufficient to better understand the work-related safety issues in the transportation sector:

- According to the database, there were 3.1 million non-fatal accidents that resulted in at least four days of absence from work and 3 332 fatal accidents in the EU-27 in 2018. The accident rate of fatal accidents was 1.77 fatal accidents per 100 000 persons employed in 2018.
- In 2018, one fifth (20.5%) of all fatal accidents at work in the EU-27 took place within the construction sector, while the transportation and storage sector (16.7%) had the next highest share.
- Non-fatal accidents were relatively common within manufacturing sector (19.1 % of the total in the EU-27 in 2018), wholesale and retail trade (12.1%), construction (11.6 %), and human health and social work activities (10.8%);
- Transportation and storage had the fifth highest share of non-fatal accidents: 9.97 %.

Figure 22 below shows the five NACE sectors with the highest risk levels for accidents at work in the EU-27. Agriculture, forestry and fishing as well as transportation and storage sectors had the smallest reductions in work-related fatalities. Non-fatal accidents reported in the framework are accidents causing at least four days absence from work.

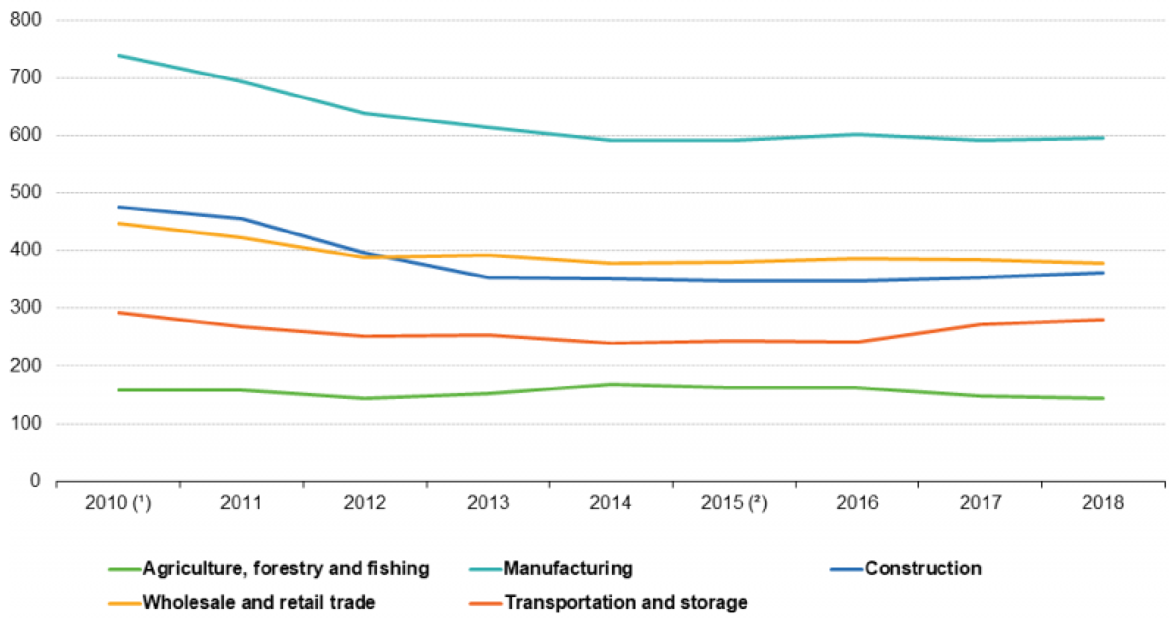

Figure 22. Development of non-fatal accidents at work for the five NACE sections with the highest risk levels, EU-27, 2010-2018 (thousands). Source: Eurostat

Taking into account the test site environments and project objectives, three key NACE sections were identified in the ESAW statistics:

- Transportation and storage
- Wholesale and retail trade; repair of motor vehicles and motorcycles
- Manufacturing.

It can be assumed that the logistics operations in the industrial areas and logistics concentrations, relevant to the AWARD's focus, are widely covered by these three NACE sectors.

The land transportation related safety levels in these sectors can be reviewed through ESAW statistics from different angles. One is the prevalence of accidents where land vehicles or conveying, transport and storage systems contributed to the accident. ESAW data includes more detailed classifications for variables related to the working person, their activity before the accident, working environment etc. than what is publicly accessible via the database (figure 23).

In total, land vehicles or conveying, transport and storage systems contributed to 363 334 accidents leading to more than 3 absences from work in EU-27 in 2018. From these accidents, 215 729 (59%) took place in the previously mentioned three NACE sectors (transport and storage, wholesale/retail etc., and manufacturing).

Another approach is reviewing the accident numbers by deviation classes and by causes of injuries. Figure 24 presents accident numbers caused by different abnormal events leading to the accidents (deviations).

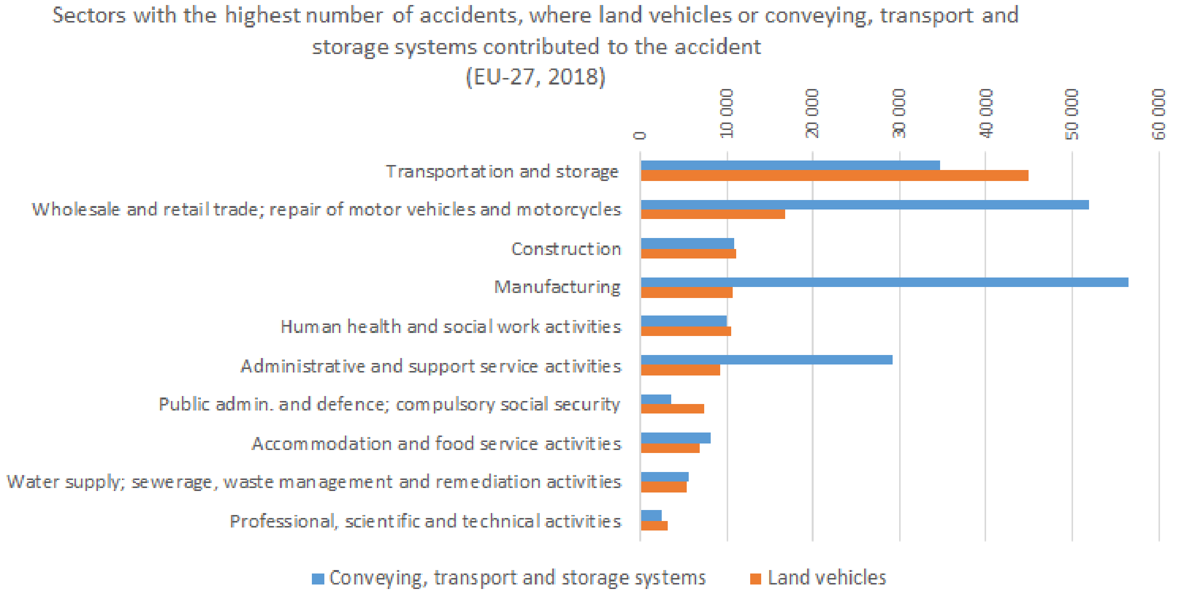

Figure 23. Sectors with the highest number of accidents, in which land vehicles or conveying, transport and storage systems contributed to the accident (EU-27, 2018). Source: Eurostat

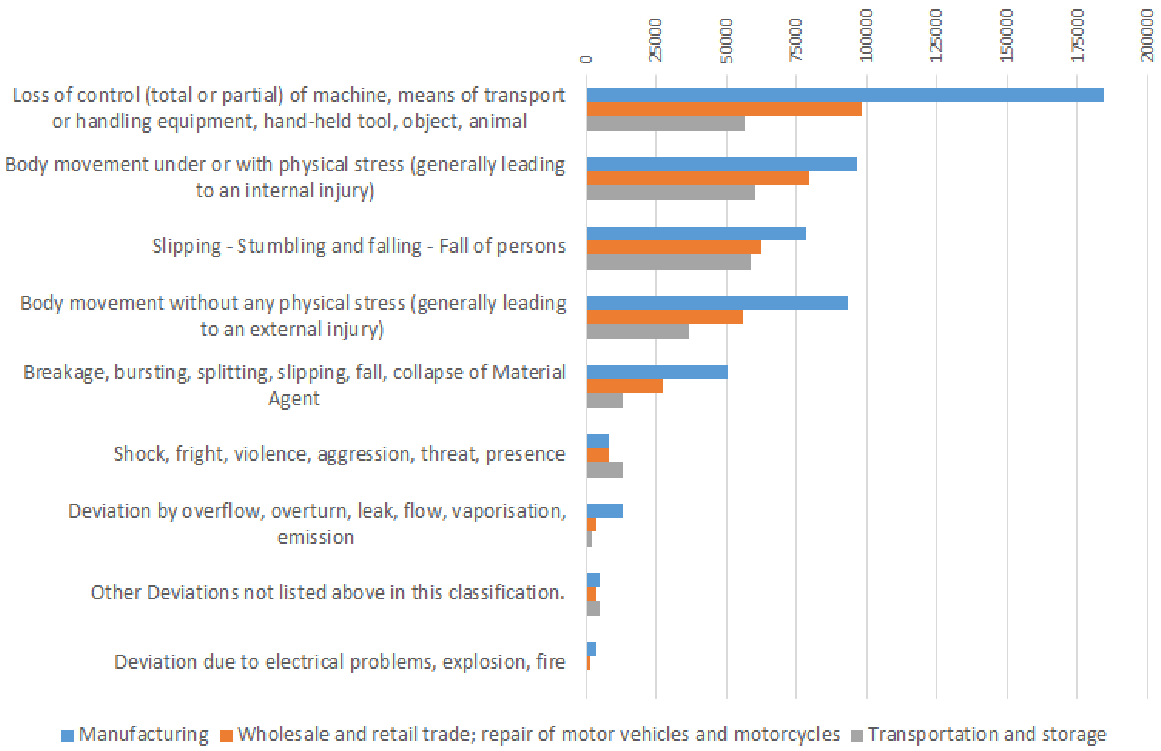

Figure 24. Non-fatal accidents in the scope NACE sectors by deviation (EU-27, 2018). Source: Eurostat

The most common deviations in the NACE sectors in focus were:

- Loss of control (total or partial) of machine, means of transport or handling equipment, hand-held tool, object, animal: 339 133 non-fatal accidents in 2018
- Body movement under or with physical stress (generally leading to an internal injury): 236 732 non-fatal accidents

– Slipping, stumbling and falling, fall of persons: 199 716 non-fatal accidents.

The most common causes of injury are presented in figure 25.

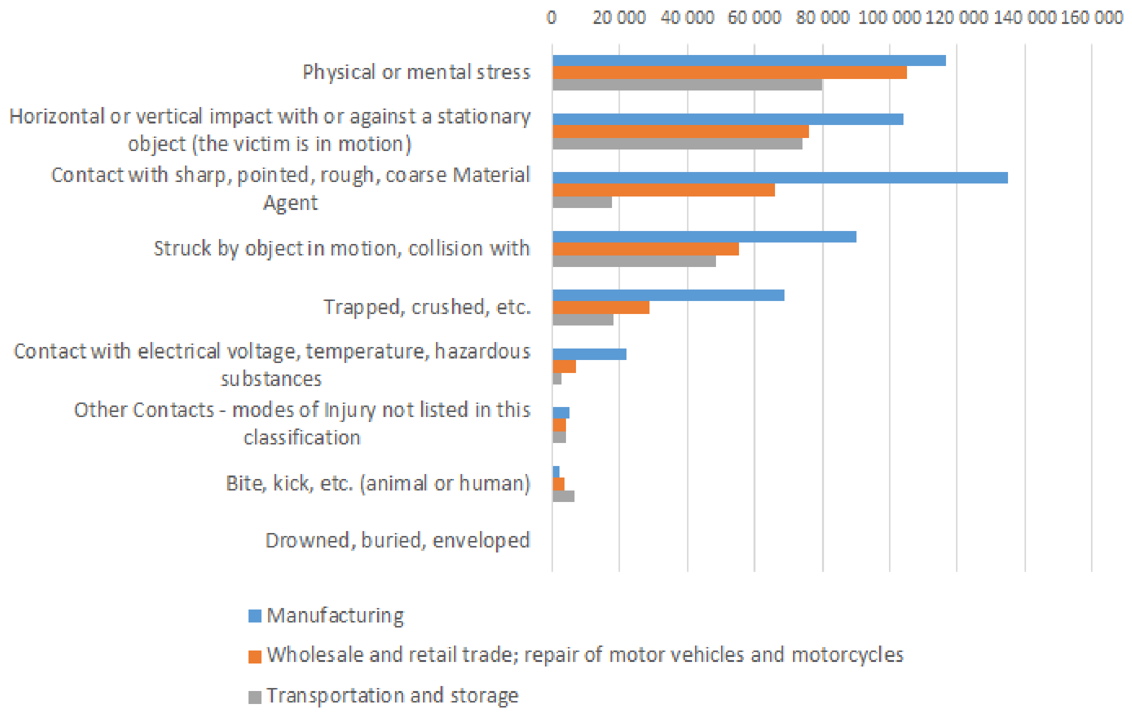

Figure 25. Non-fatal accidents in the scope NACE sectors by contact mode of injury (EU-27, 2018). Source: Eurostat

In addition to Eurostat's open data, different national statistics are being reviewed in the AWARD project. Finnish workplace accident statistics were received from Finnish Workers' Compensation Center (TVK), which is the official authority for statistics on occupational accidents and diseases in Finland.

The first reviewed dataset from TVK included workplace accidents that took place during 2016–2019. The dataset was limited to accidents that were caused by different mobile machines and devices (15 sub-groups) and took place in the following working environments:

- Production facility, factory, workshop
- Service area, repair shop
- Storage, warehouse, loading and unloading area
- Public working environment
- Other industrial facility.

When the transport equipment was causing the accident, the most common deviations that lead to accidents were:

- Collision, falling, sliding, breakage of the causative agent.: 757 accidents per year in 2016–2019
- Stepping on a sharp object, clinging to something, kneeling, sitting down: 597 accidents/a (year)
- Loss of control of equipment, working machine or animal: 536 accidents/a
- Falling down, jumping, falling, slipping: 516 accidents/a
- Sudden physical exertion: 488 accidents/a.

The most common contact modes of injury in accidents caused by transport equipment were:

- impact against a solid surface: 914 accidents per year in 2016–2019
- a hit or collision of a moving agent: 897 accidents
- compression, being crushed, bruising: 631 accidents
- sudden physical or mental strain: 482 accidents
- damage caused by a cutting, sharp or rough object: 218 accidents.

Figure 26 presents the most common contact modes of injury by transport equipment categories. The figure also presents four vehicle categories relevant to the project scope.

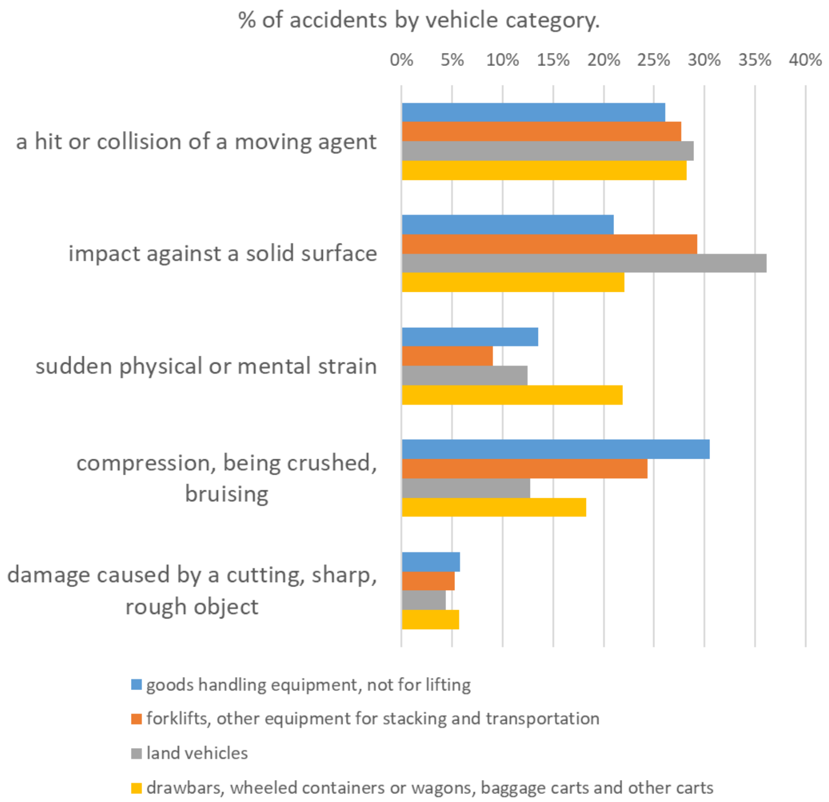

Figure 26. Most common contact modes of injury in accidents caused by transport equipment in Finland in 2016–2019. Source: TVK

Forklift accident statistics

Forklifts are a necessary piece of material handling equipment especially in warehousing and logistics and manufacturing. According to industry statistics in the United States, there's a 90% probability of a forklift being involved in a serious injury or fatality accident over its lifetime [16]. This is a significant finding due to forklifts' largely irreplaceable role in many industries.

In US, the most common fatal accident types are being crushed by vehicle tipping over and being crushed between vehicle and surface [16]. In 2009 around 80% of forklift accidents involved a pedestrian and over 18% of forklift accidents occurred when a forklift strikes a pedestrian [17].

Table 22. The most common types of fatal forklift accidents in U.S. [16]

Fatal Accident Type	%
Crushed by vehicle tipping over	42%
Crushed between vehicle and a surface	25%
Crushed between two vehicles	11%
Struck or run over by a forklift	10%
Struck by falling material	8%
Fall from platform on the forks	4%

A glance on the European level forklift safety statistics indicates that there is not comprehensive data publicly available. However, it can be assumed that the working environments and conditions of forklift operations are somewhat similar in United States and in most EU-27 countries. To form a better picture on the topic, some national statistics were searched and reviewed. According to Finnish national statistics, the most common deviations leading to injuries in accidents caused by forklifts in 2016–2019 are:

- Sudden physical exertion (27 % of forklift-caused accidents)
- Collision, falling, sliding, breakage of the vehicle (19 %)
- Loss of control of equipment or working machine (18 %).

The most common contact modes of injury are (figure 25):

- Impact against a solid surface (29 % of forklift-caused accidents)
- A hit or collision of a moving agent (28 %)
- Compression, being crushed, bruising (24 %).

In fatal and serious forklift accidents, the victim has been typically pinched between or under the forklift or the load transported. About 40 % of the fatalities were forklift drivers, 25 % assisted in forklift work and 35 % were bystanders [18].

The most common causes of forklift accidents are dangerous or new working methods, lack of co-operation, poor compliance with instructions and technical faults and deficiencies. Studies show that significant portion of forklift accidents could be prevented by better training. Also lack of supervision and job guidance plays a major role in many forklift accidents. Dangerous situations are also caused by deficiencies in work planning or deviations from the plans [16] [19] [20] [21] [22].

AGV safety

Forklift accidents cause roughly 35 thousand serious injuries in the U.S. every year. The closest automated counterpart, today, are the automated guided vehicles (AGVs). The AGVs mostly move indoors and use walking speeds. Due to the low moving speeds, fatalities and serious injuries are rare. In 2018, AGVs were involved in accidents resulting two fractured legs in the U.S and one laceration, all requiring hospital treatment [23]. The last fatality in the U.S. happened in 2012. A scale factor is that while there were 110 thousand AGVs sold globally in 2018, while forklift sales were almost 15 times higher [24] [25].

Airport ground handling accident statistics

Ground handling operations are generally operated by private companies and safety issues are not widely reported. Accidents and clear safety issues affecting the aircraft are covered more comprehensively in national and international statistics, but for example collisions between land vehicles that cause only material damage, often stay in the companies' own knowledge.

A key source for ground handling safety statistics is the Annual Safety Review provided by EASA (European Union Aviation Safety Agency). According to EASA there were 7 fatal accidents, 448 non-fatal accidents and 104 serious incidents during 2009–2018 in aerodrome and ground handling operations at EASA member state airports [26]. Using these accident numbers and total amount of flights in 2008–2017 provided by EASA, there was approximately 0.08 fatal and 5 non-fatal accidents per million flights in EASA member state airports in 2009–2018.

A Dutch study [27] by Balk et al., published by EASA and developed by the Dutch National Aerospace Laboratory (NLR), addresses ground handling safety of large airports worldwide. The study comprised data from several airlines and main airports in the United States, Canada, Europe, Australia, Far East and Africa. Data covered 14 million flights and 2841 incidents (before 2008). According to the study, there is ca. one ground handling incident with resulting aircraft damage per 5000 flights (200 / 1 million flights). Most incidents (61%) occurred when the aircraft is stationary and when interfaces are established between the aircraft and ground handling equipment.

The study provides detailed information on the baggage transport vehicles' accidents:

- From all ground incidents 12 (0.422%) were caused by aircraft colliding to baggage truck/cart
- 141 incidents (4.96%) were caused by baggage truck/cart colliding to aircraft.

Table 23 comprises results based on the above-mentioned publications and presents estimated annual accident rates on EASA airports. Data collection criteria for EASA airports (C1) and for worldwide airports in the Balk's report (C2) were:

- C1: Aerodrome and ground handling accidents collected by under Regulation (EU) 996/2010 on accident and serious incident investigation and Regulation (EU) 376/2014 on occurrence reporting, and through the active search of those events from other official sources.
- C2: Ground handling incidents resulting in aircraft damage and taking place in Taxi-in, Docking, Standing, Pushback, Towing or Taxi to runway.

Table 23. Estimations on annual numbers of ground handling accidents and incidents in EASA airports. Source: VTT

	Total accidents/incidents in	Time frame	Per 1 million flights*	Airport locations	Dataset	Criteria for data		Estimated numbers of accidents in EASA airports Accidents/year **	
Fatal accidents	7	2009–2018	0,08	EASA area	EASA 2020	C1	→	0,7	Fatal accidents
Non-fatal accidents	448		5					45	Non-fatal accidents
Serious incidents	104		1					10	Serious incidents
Ground handling incidents with resulting aircraft damage	2841	Before 2008	200	Worldwide	Balk 2008	C2	→	2800	Ground handling incidents with resulting aircraft damage
Aircraft causes collision to baggage truck/cart	12		0,8					8***	Aircraft causes collision to baggage truck/cart
Aircraft collisions caused by baggage truck/cart	141		10					90***	Aircraft collisions caused by baggage truck/cart

*For fatal accidents, non-fatal accidents and serious incidents in EASA airports, the accident rates were calculated using total numbers of accidents and average number of flights from a 10-year period. For ground handling incidents in global airports the frequencies are based on the Balk's report [27].

**Preliminary estimations that are calculated by applying presented accident rates to EASA flight statistics (Avg: 9 095 146 flights / year) [31].

***Worldwide incident frequencies are applied to European airports, which might have higher overall safety level.

Annex II. Other statistics for scaling up effects of automation

In order to scale up the effects of the automation of industrial trucks, that is, to project the test results to relevant industries and to the EU, the AWARD project needs to clarify e.g. the number of similar vehicles sold each year. A certain number of such vehicles would be seen to become automated in the future.

This Annex II discusses some of the related statistics available for the scaling up calculations. Beside the sales numbers and accident statistics, scaling up the environmental impacts would benefit from data on the energy use, average operation hours and emissions of the targeted vehicle categories.

Where the WP7 discusses mainly just the benefits and drawbacks, WP8 in AWARD continues by weighing costs versus benefits.

Emissions of non-road mobile machinery

Since non-road mobile machinery is not widely registered in Europe, the knowledge of the emissions of those machines is mostly based on estimates and calculation models. One example is VTT's TYKO-model [28], which includes an estimation of the NRMM (non-road mobile machinery) stock in Finland, including major machine categories. The input data is partially based on real data (e.g. sales). However, there are major inconsistencies included, since no organization actually records all NRMM sales and only minor share are registered. Also, the average power output and annual usage of NRMM is not well known and only based on estimates and narrow research knowledge. Also, the electrification of working machines is speeding up, and many new categories should be added to the model.

TYKO model presents estimates on forklifts, but baggage transport equipment does not have a category of its own. The model consists of 51 machine categories. From project scope, the most relevant categories and their key numbers for year 2019 are presented in table 1.

Table 24. TYKO data for 2019 [28]

	Forklift, gasoline	Forklift, gas	Forklifts, diesel	Other diesel-powered lift trucks	Other driveable machines, diesel	Other driveable machines, gasoline
Volume [pc.]	159	2 488	5 045	1 996	2 277	5 548
Sales [pc.]	15	235	230	115	130	800
Nominal power avg. [kW]	30	30	88	33	89	10
CO2 [t/a]	606	9 720	172 030	7 007	31 853	5 099
Fuel consumption [t/a]	217	3 387	54 447	2 217	10 080	1 822
Energy [GWh/a]	3	43	653	27	121	21

A significant knowledge cap is the annual usage of different working machines. This challenge applies to both safety and environment aspects of the project. Data should be collected or estimates formed to supplement the current knowledge and to further develop better understanding on usage-based emissions and consumption. Usage-based estimates on emissions are needed to create better baseline values used in impact evaluation.

Land vehicle emissions in ports

The US Environmental Protection Agency has committed a study on the exhaust emissions of heavy-duty diesel-powered tractor-trailer trucks that operate in drayage service. Drayage service involves the moving of shipping containers to or from port terminals. The study involved the use of portable emission measurement during test trucks' normal drayage service and the measurements were supplemented with terminal entry and exit logging [29].

Among other results, the study presented measurement-based average emissions per second and per mile during drayage operations during in-port phase and non-port phase for different age categories of trucks (table 25).

Table 25. Average emissions during drayage operations [29]

	Model Yr	HC	CO	CO2	NOx	PM	unit
In-Port	2004 +	0.002	0.036	2.881	0.031	0.0012	g/s
Non-Port	2004 +	0.002	0.036	13.367	0.143	0.0034	g/s
In-Port	2004 +	3.03	52.39	4 208.7	45.62	1.81	g/mi
Non-Port	2004 +	0.33	4.86	1 819.8	19.48	0.46	g/mi

The presented unit emission results must be viewed critically and cannot be applied to impact evaluation baselines directly. A highly beneficial result, however, is the difference of the average emission levels between in-port and non-port driving conditions. Frequent changes in driving speed and the need to stop the vehicle more often has a negative effect on the produced emissions. Automating driving may have significant effect on this if the average vehicle speeds during operations are more constant and if the total driving distances are similar to baseline situations with a human driver.