Ref. Ares(2026)2275295 - 02/03/2026

Efficient and affordable Zero Emission logistics through **NEXT** generation **E**lectric **TRUCK**s

HORIZON Innovation Actions ¦ Project Number: 101056740

D7.2
Data Collection and Management Plan

Co-funded by the European Union

NextETRUCK has received funding from the European Union Horizon Innovation Actions programme under grant agreement No. 101056740.

Dissemination level	Public (PU) - fully open
Work package	WP7: Use cases and demonstrations
Deliverable number	D7.2. Data collection and management plan
Version	V1.1
Submission date	02/03/2026

Quality Control

	Name	Organisation	Date
Peer review 1	Jean-Charles Pandazis	ERTICO	19/12/2025
Peer review 2	Leo Xenakis	AVL-AT	19/12/2025

Version history

Version	Date	Author	Summary of changes
0.1	16/12/2025	Victor Lejona, Sophie Naylor, Tom Allerton (Cenex UK)	Initial draft for peer review
1.0	23/12/2025	Victor Lejona, Sophie Naylor, Tom Allerton (Cenex UK)	Review comments issued
1.1	18/02/2026	Victor Lejona, Sophie Naylor, Tom Allerton (Cenex UK)	Updates based on latest communications with OEMs

Legal disclaimer

TABLE OF CONTENTS

ABBREVIATIONS AND ACRONYMS

Abbreviation	Meaning
BEV	Battery Electric Vehicle
CAN	Controller Area Network
CAPEX	Capital Expenditure
CPMS	Chargepoint Management System
DALI	Driving Activity Load Index
GDPR	General Data Protection Regulation
GPS	Global Positioning System
HMI	Human-Machine Interface
HVAC	Heating, Ventilation and Air Conditioning
ICE	Internal Combustion Engine
KPI	Key Performance Indicator
MCS	Megawatt Charging System
OCPP	Open Charge Point Protocol
OEM	Original Equipment Manufacturer
OPEX	Operational Expenditure
SUS	System Usability Scale
TCO	Total Cost of Ownership
WBG	Wide Bandgap
ZEV	Zero Emission Vehicle

EXECUTIVE SUMMARY

NextETRUCK aims to address different optimization challenges regarding tomorrow's urban and suburban logistics for medium-duty vehicles into a systems approach that is reliable, strongly integrated, affordable, and flexible enough to be re-applied to different applications via dedicated tools/methods.

This report defines the data collection, management, and processing framework supporting the validation of the NextETRUCK project demonstrators. Its primary purpose is to ensure that robust, high-quality data is gathered, processed, and analysed to evaluate key performance indicators (KPIs) and to verify whether project innovations deliver measurable improvements over established baseline technologies.

Objectives and Scope
The central objective of data collection is to enable a rigorous KPI evaluation of three electric truck demonstrators across different operational use cases. The report focuses specifically on KPIs related to demonstrator performance, including vehicle efficiency, range, charging performance, thermal efficiency, total cost of ownership (TCO), operational capability, and user acceptance. While the project defines 26 KPIs overall, this document concentrates on those directly linked to demonstrator validation within WP7, with references to other deliverables where broader KPIs are addressed.

KPI Framework and Baselines
KPIs are grouped into four categories: Vehicle & Charge Point, Digital Tools, Market/TCO, and End User. This report primarily addresses vehicle, charge point, TCO-relevant, and end-user KPIs. Performance improvements are assessed against two baselines: (i) comparable internal combustion engine (ICE) trucks, using public literature data, and (ii) medium-sized battery electric trucks representative of the state of the art at project launch in 2022, using a combination of telemetry and literature data. These baselines are critical to ensuring credible, comparable performance assessments.

Data Sources and Collection Methods
Data is collected through multiple complementary channels:

- Vehicle telemetry from original equipment manufacturer (OEM) systems (Irizar, Ford, and DAF), capturing energy consumption, distance, payload, HVAC usage, operating hours, and other parameters required for KPI calculations.
- Chargepoint telemetry, primarily via open charge point protocol (OCPP)-compliant systems, enabling analysis of charging power and efficiency.
- Questionnaires targeting drivers, fleet managers, and human–machine interface (HMI) usability to support end-user KPIs.
- Public literature and datasets, particularly for cost and TCO analysis where commercially sensitive data cannot be shared.

The report details use-case-specific arrangements for each OEM, acknowledging differences in telemetry systems, data completeness, and deployment timelines.

Data Management, Processing, and Quality Assurance
All raw data is securely transferred to Cenex and stored in its original form to preserve flexibility for reprocessing. A modular, Python-based data processing pipeline is used to clean, validate, harmonise, and standardise heterogeneous datasets across OEMs and sources. Quality assurance procedures detect missing data, faulty loggers, anomalous values, and inconsistencies, ensuring only reliable data feeds into KPI evaluation. Harmonisation includes timestamp alignment, frequency alignment, variable naming standardisation, and format conversion.

Security, Privacy, and Governance
Data handling complies with General Data Protection Regulation (GDPR) and industry best practice. Sensitive information, particularly global positioning system (GPS) data and vehicle identifiers, is protected through secure storage, controlled access, aggregation, and anonymisation where appropriate. Cenex maintains robust data resilience through multiple backups and recovery mechanisms and holds Cyber Essentials Plus certification.

Data Sharing and Timelines
Processed, standardised datasets will be shared with TNO via Microsoft SharePoint, typically on a weekly basis, to support KPI calculations in WP8. Full data transfers to TNO are expected to commence in late February or early March 2026. A tentative data cut-off of either end of April or May 2026 was set to stop accepting data receipts (by Cenex from project partners) and allow sufficient time for KPI evaluation prior to project completion in September 2026.

Overall, the report establishes a comprehensive and reliable data framework that underpins the technical, economic, and user-focused validation of NextETRUCK demonstrators and supports confident assessment of project outcomes.

1 INTRODUCTION

1.1 About NextETRUCK

NextETRUCK is a Horizon Europe project that develops zero emission vehicle (ZEV) concepts tailored for regional medium freight haulage, running from 1 July 2022 until 30 September 2026.

The project aims at playing a pioneering role in the decarbonisation of vehicle fleets, demonstrating next-generation e-mobility concepts. It also contributes to the development of zero-emission vehicles and ecosystems that are holistic, innovative, affordable, competitive, and synergetic.

NextETRUCK is expected to build concepts tailored for regional medium freight haulage with at least a 10% increase in energy efficiency compared to existing highest-end benchmark electric vehicles. In addition, it shall prepare concept and infrastructure demonstrators for fast charging and offer new business models to increase end-user acceptance and foster the market uptake of the project solutions.

The project's consortium consists of 19 partners from 8 countries: The Netherlands, Belgium, Germany, Spain, Greece, Austria, Turkey, United Kingdom[1]. The project's coordinator is TNO (Netherlands Organization for Applied Scientific Research).

NextETRUCK shall conduct demonstrations in Northwest Turkey, Barcelona (Spain), and Northwest England.

1.2 Purpose of the deliverable

This is a public report describing which data is collected in the project demonstrators and use cases to evaluate and validate the demonstrator KPIs defined in T8.1, as well as the procedures to be followed during data collection from their source, transfer within project partners, storage and processing.

1.3 Intended audience

This deliverable is available to the public, but it is especially intended for any original equipment manufacturers (OEMs), technology developers, fleet operators and research agencies who want to plan an electric truck trial in the future and seek guidance.

[1] The UK participants in this project are co-funded by the UK.

1.4 Structure of the deliverable and its link to other work packages/deliverables

The report starts with a brief overview of the data flow between project partners and their roles in section 2, followed by a description of the motivation for collecting data in this project in section 3. Section 4 then forms the core of the deliverable as it describes which data is collected and how depending for each use case. The last three sections (5, 6, 7) cover the inwards data transfer to Cenex, its safe storage, the processing and handling of this data, and finally the release to other project partners to fulfil the project evaluation objectives.

This deliverable is closely related to the following tasks:

- T7.1: Demonstration plan and processes, which describes the latest definitions of the use cases, vehicle baseline definition, time-plan for each use case, and the demonstration assessment questionnaires for drivers and fleet managers. This report uses key information from its deliverable D7.1.
- T7.2 to 7.4: Demonstration in the three pilot sites, which deliver the vehicle improvements and trial the upgraded demonstrator trucks on real-world roads.
- T8.1: Evaluation plan, which outlines the approach and process for validation and impact assessment of the project innovations and defines the 26 project key performance indicators (KPIs). The data collection process described in this report is guided by the KPI descriptions, assessment methods, data requirements, risks and mitigations.
- T8.2: Validation, which evaluates the technical and operational performance, reliability, usability and effectiveness of the technology tested in the three use cases, by calculating the KPIs defined in T8.1. This report describes how the data collection and management process ensures that the outputs are in the right format and have the necessary quality to enable the KPI calculation in T8.2.
- T8.3 Scaling up and impact assessment, which uses digital twins and simulations to extrapolate the innovations demonstrated in this project to country-wide fleets. This report ensures that the data collection is in the right format to inform this extrapolation exercise.

1.5 Main changes from previous version

The main changes of the present version of the deliverable compared with the previously delivered version are:

- V0.1: Initial draft for peer review.
- V1.0: Minor typographical edits and responses to review comments.
- V1.1: Updates based on latest communications with OEMs.

2 OVERVIEW OF DATA COLLECTION AND MANAGEMENT

2.1 Data flow

Figure 1 below shows an overview of the data elements to be collected and analysed during the project. The way to read this diagram is:

1. Why do we need to measure parameters? First level: KPIs to validate project innovations.
2. What do we need to measure? Second level: described in detail in this report with a focus on vehicle/chargepoint and end user items.
3. Who will provide the data? Fourth level: both project partners and external sources.
4. How will it be collected and analysed? Third level: described in detail in this report.

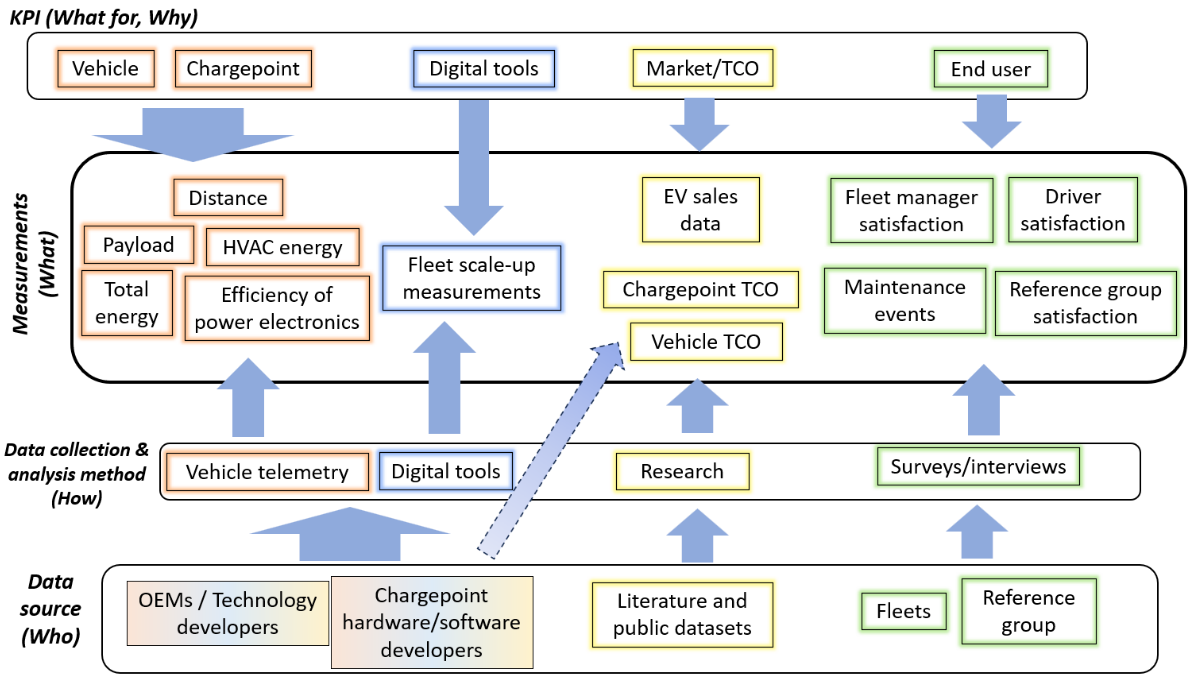

Figure 1: NextETRUCK data flow process. OEMs and technology developers are given three colours because they provide info across several KPI categories.

The colour codes are aligned with the four KPI groups:

- In orange, the 'Vehicle & Charge point' related KPIs. These want to capture the performance of the design innovations compared to the baseline technologies. They are primarily measured via telemetry systems and are the focus of this report.
- In blue, the 'Digital tools' related KPIs. They want to capture the simulated performance of a large fleet of electric vehicles across a variety of missions, locations and ambient conditions, by scaling up the behaviour of the three pilot vehicles and using their data to recalibrate themselves and increase their accuracy. They are not covered in this

report as they are not directly related to the demonstrators. Please refer to WP4 deliverables for more information.

- In yellow, the ‘Market / total cost of ownership (TCO)’ related KPIs. These capture the business case improvements to a fleet user due to the innovations adopted within the project. They will be measured using a combination of literature and/or public datasets, plus first-hand data from the OEMs and technology developers participating in the project. They are covered in this report with a focus on the data collected during the demonstrators that is relevant to TCO calculations (e.g. energy consumption).
- In green, the ‘End user’ related KPIs. They will capture the satisfaction of various users of the innovations developed within the project. They are covered in this report but are described in detail in D8.1.

2.2 Roles

During the data collection process, different project partners have different roles and responsibilities:

- Data providers/custodians/owners: OEMs, chargepoint manufacturers, and technology developers provide their data from their hardware and software and control to a large extent who can access the data.
- Data collectors/processors/analysts: Cenex and TNO collect data from data providers and process it to ensure it meets minimum quality standard requirements so that it can feed into the validation during T8.2.
- Insight producers: TNO with support from Cenex, CERTH and all pilot sites during T8.2.
- Data repository managers: This depends on the data source. For example, for data being made available in the project Sharepoint, that would be VUB. For data being made available by OEMs, e.g. telemetry data uploaded to Irizar to their company Sharepoint, it would be the OEMs themselves.

3 MOTIVATION FOR DATA COLLECTION

This section explains the main reasons to collect project data in order to motivate data providers and instil confidence.

3.1 Evaluation of demonstrators

The collection of data from the project demonstrators across the three use cases is essential for enabling their KPI evaluation in T8.2. This KPI evaluation will confirm if the trucks meet the defined specifications and comply with the project’s performance requirements. The process entails a structured review of the collected data to detect any deviations and verify adherence to the established criteria.

As part of the NextETRUCK validation activities, key parameters such as driving range, acceleration, payload capability, energy and charging efficiency, user acceptance, and safety systems will be assessed against the defined KPIs. Effective execution of this process supports confidence in the electric truck's operational performance within the project.

3.2 KPIs related to demonstrators

This report focuses on the data needed to evaluate the KPIs related to the demonstrators and not all 26 project KPIs (refer to D8.1 for the full list of KPIs). The KPIs that this report focuses on are indicated in Table 1.

Table 1: KPIs related to WP7 (demonstrators)

KPI id	Statement / Goal
KPI-1a	Achieving 10% longer range per energy consumption in comparison with the state-of-the-art similar EVs …
KPI-1b	... and for at least 90% loaded case in comparison to the internal combustion engine (ICE) trucks for similar missions
KPI-2	For all use cases, achieving at least 95% of the ZEV truck operation range while heating is used, compared to no heating
KPI-6a	20% TCO reduction of ZEV is achieved when compared with a modern e-truck
KPI-6b	TCO parity with ICE 2020 trucks.
KPI-7	Up to 4% more average charging efficiency increase is achieved at wide bandgap (WBG)-based power modules.
KPI-8	Vehicle thermal efficiency is improved by 15%.
KPI-11	Double shift operation for the fleet operator (thus, improving the TCO by minimizing the investment in rolling stock and the charging time) is made possible
KPI-13a	Goods distribution e-truck in Istanbul with FORD as OEM is successfully in operation for a daily range of more than 200km, for at least 6 months
KPI-13b	
KPI-14a	ZEV refuse truck in Barcelona with IRIZAR as OEM is successfully in operation for a daily range of more than 200km, for at least 6 months.
KPI-14b	
KPI-15a	Back to base logistics e-truck for express transport in the UK with DAF as OEM is successfully in operation for a daily range of more than 200km, for at least 6 months.
KPI-15b	
KPI-20	NextETRUCK demonstrates a reduction in operational CO2 emission relative to battery electric vehicle (BEV) and ICE baseline vehicles.
KPI-21	The operating hours of NextETRUCK are similar to ICE truck
KPI-22	The charging rate of the NextETRUCK will be as close as possible to the predicted by the megawatt charging system (MCS) in the range from 20% to 80% state of charge.

3.3 Baselines

As well as collecting data from project innovations, having a baseline is essential to compare those innovations against. This is especially important in the cases when KPIs are measured as an improvement against a baseline. For some other KPIs, the baseline is zero or no baseline is required.

Baseline data consists of the initial measurements and observations gathered to establish a reference for assessing and comparing the performance of technologies introduced within the project. It provides a benchmark against which later measurements and evaluations are conducted to identify performance enhancements or deviations for each vehicle.

The accuracy and reliability of baseline data are essential to the credibility and effectiveness of subsequent validation activities, as this data underpins all performance assessments. Within the project, two baseline datasets will be applied:

- One representing the performance of comparable ICE trucks. Due to the nature of the KPIs that require a comparison against ICE, no telemetry data is needed and public literature data about comparable ICE trucks will be used.
- Another one reflecting the performance of medium-sized battery electric trucks at the time when the project was launched (2022). Both detailed telemetry data and public literature data will be used to obtain suitable baseline BEV comparisons.

To know more about the vehicles being used as a baseline and their specifications, please refer to D7.1.

As several KPIs are intended to demonstrate measurable improvements in performance and efficiency achieved by the demonstration vehicles relative to the baseline, the baseline datasets must include signals that are directly comparable to those collected during the demonstration phase. This is covered in the following sections.

4 DATA COLLECTION PROCESSES

4.1 List of data items to be collected

Depending on the KPI being evaluated / validated, different data items need to be collected and different methods will be used to that end. These are indicated in Table 2.

Table 2: Data items to be collected and methods of collection.

KPI id	Statement / Goal	Data items	Data collection method
KPI-1a	Achieving 10% longer range per energy consumption in comparison with the state-of-the-art similar EVs …	Energy consumption (from high-voltage battery current, voltage and timestamp), distance (from speed and timestamp), gradient (from GPS), ambient temperature	Vehicle telemetry
KPI-1b	... and for at least 90% loaded case in comparison to the ICE trucks for similar missions	Payload (from axle loads or derived from motor toque, depending on vehicle)	Vehicle telemetry
KPI-2	For all use cases, achieving at least 95% of the ZEV truck operation range while heating is used, compared to no heating	Heating, ventilation and air conditioning (HVAC) energy consumption (from compressor current and voltage or resistance heater, depending on vehicle), ambient temperature	Vehicle telemetry
KPI-6a	20% TCO reduction of ZEV is achieved when compared with a modern e-truck	Capital and operational expenditure (CAPEX and OPEX). For the demonstrators, energy consumption will be used.	Vehicle telemetry, literature
KPI-6b	TCO parity with ICE 2020 trucks.		
KPI-7	Up to 4% more average charging efficiency increase is achieved at WBG-based power modules.	Power entering high-voltage battery (from voltage, current, timestamp) and power exiting charging modules at chargepoint	Vehicle telemetry, chargepoint telemetry
KPI-8	Vehicle thermal efficiency is improved by 15%.	HVAC energy consumption (from compressor current and voltage or resistance heater, depending on vehicle), ambient temperature	Vehicle telemetry
KPI-11	Double shift operation for the fleet operator (thus, improving the TCO by minimizing the investment in rolling stock and the charging time) is made possible	Distance and timestamp	Vehicle telemetry
KPI-13a	Goods distribution e-truck in Istanbul with FORD as OEM is successfully in operation for a daily range of more than 200km, for at least 6 months	Distance and timestamp	Vehicle telemetry
KPI-13b			
KPI-14a	ZEV refuse truck in Barcelona with IRIZAR as OEM is successfully in operation for a daily range of more than 200km, for at least 6 months.	Distance and timestamp	Vehicle telemetry
KPI-14b			
KPI-15a	Back to base logistics e-truck for express transport in the UK with DAF as OEM is successfully in operation for a daily range of more than 200km, for at least 6 months.	Distance and timestamp	Vehicle telemetry
KPI-15b			
KPI-20	NextETRUCK demonstrates a reduction in operational CO2 emission relative to BEV and ICE baseline vehicles.	Energy consumption (from high-voltage battery current, voltage and timestamp), distance (from speed and timestamp)	Vehicle telemetry
KPI-21	The operating hours of NextETRUCK are similar to ICE truck	Distance and timestamp	Vehicle telemetry
KPI-22	The charging rate of the NextETRUCK will be as close as possible to the predicted by the MCS system in the range from 20% to 80% state of charge.	Power entering high-voltage battery (from voltage, current, timestamp) and state of charge	Vehicle telemetry

4.2 Data collection methods

4.2.1 Vehicle telemetry

4.2.1.1 Irizar use case

For both baseline and demo vehicles, Irizar collect data using their own telemetry connected to several controller area network (CAN) buses. This data is very complete and contains all relevant information required for KPI calculations. There are only some caveats around the global positioning system (GPS) data collected by this logger:

- For the baseline vehicle, the GPS antenna in the Irizar logger was malfunctioning so the baseline GPS data is collected from the local fleet operator, who had a GPS tracker.
- For the demo vehicle, the GPS in the Irizar logger can sometimes lose signal in densely built areas of the city, so this data is complemented with GPS data collected by a logger deployed by Datik.

4.2.1.2 Ford use case

For both baseline and demo vehicles, Ford collect data using their own telemetry connected to several CAN buses. At the time of writing, data from eight journeys of the baseline vehicle were shared, while no data is available from the demo vehicle yet.

- For the baseline vehicle, only eight journeys were provided because Ford considered that this baseline data was the best representative data for comparison with the demo vehicle. Data from some journeys is lacking some variables so more data might be requested in due course.
- For the demo vehicle, Ford has communicated that the format of the data will be very similar to that of the baseline vehicle. Ford will start testing the demo vehicle on their premises in March 2026 and can make test data available as soon as they have it for Cenex and TNO to adjust their data processing algorithms.

4.2.1.3 DAF use case

- For the baseline vehicle, data from a past project was used. This is Home | BETT - Battery Electric Truck Trial, which was an Innovate UK-funded project in which 20 DAF Electric LFs (19-tonne GVW) were deployed in 9 public sector fleets across England between April 2022 and September 2023. Cenex led the study aspects of the research and deployed third party loggers in all vehicles to collect high frequency CAN data encompassing all variables required for the NextETRUCK KPI comparison. These loggers were remotely connected to the Cenex server database and the data collected will be used as a baseline in NextETRUCK.
- For the demo vehicle, DAF have two systems collecting data. Firstly, the Paccar Connect system reports data remotely to fleet operators but can also be used to monitor telemetry data required for this project. Secondly, DAF also have a logger in

the demo vehicle which collects additional data not collected by Paccar Connect, such as ancillary energy usage (e.g. HVAC).

4.2.2 Chargepoint telemetry

4.2.2.1 Irizar use case

- The baseline vehicle was charged using a Jema charger, which records high frequency data and has the data variables required to calculate charging efficiency.
- The demo vehicle is also being charged using a Jema charger, which K2 Mobility is communicating with via the standard open charge point protocol (OCPP) 1.6J. The K2 portal can be accessed to export logs per day into a csv file. At the time of writing the charger was being commissioned so no data has been received yet.

4.2.2.2 Ford use case

At the time of writing, it is still unclear what charger was used in the baseline vehicle or what charger will be used in the demo vehicle. However, most chargers nowadays use the OCPP and therefore data is expected to be available in a standard format.

4.2.2.3 DAF use case

- Baseline charger data was recorded in the BETT trial through the chargepoint management system (CPMS) provider and will be used in this project.
- Mobile chargers will be used for the demo vehicle due to short timeframes that DAF had to deploy the vehicles. At the time of writing, it is not clear whether any charger data will be available.

4.2.3 Questionnaires

Several questionnaires were developed by CERTH during the production of D7.1:

- One for vehicle drivers to collect comprehensive information from individuals who have tested the NextETRUCK vehicle. The form covers general information about the respondent, initial impressions, driving experience, technology and features, comfort and interior, and an overall impression of the vehicle.
- One for the human-machine interface (HMI, i.e. the tablet installed in the vehicles) to gather insights into the usability and perceived constraints of the system being evaluated, providing feedback for improvements and enhancements. It consists of two parts: the System Usability Scale (SUS) and the Driving Activity Load Index (DALI), which evaluates the constraints and difficulties experienced by participants during a driving task.
- One for fleet managers to gather information and feedback about the NextETRUCK fleet management tool. It covers key areas such as general information about the company and its fleet, the current fleet management situation, user interface and experience, vehicle monitoring and maintenance, route optimization and planning, reporting and analytics, and overall satisfaction.

For more information on these questionnaires and their content, please refer to D7.1.

4.2.4 Literature

Where data cannot be collected from the above methods, publicly available literature and data will be used. This will be mostly the case with the TCO-related KPIs in order to protect the confidentiality of OEMs and technology developers in the project regarding the sensitivity of costs from vehicles and components, and anonymisation techniques will be applied where possible (e.g. applying percentage differences in costs). Publicly available data was used to a large extent in D3.5 “Report on first TCO assessment with mission profiles” (led by Cenex), so the methodology and data from that deliverable can be further leveraged.

5 INWARDS DATA TRANSFER AND STORAGE

Most data will be transferred to Cenex via either email or a Microsoft SharePoint folder shared amongst project partners. These systems ensure that data owners can share their data only with the project partners that require it for evaluation and validation purposes, i.e. Cenex and TNO, hence preserving confidentiality and instilling confidence in data owners.

Raw data is collected using the various collection methods previously described and is stored on Cenex’s data processing workstation, which is physically located in Cenex’s UK office. We aim to retrieve the data in the rawest form we can to allow maximum flexibility with how it is processed, and we always keep a copy of this original data for the duration of the project so that we can re-process it if necessary.

5.1 Irizar use case

Vehicle telemetry is automatically uploaded to a shared SharePoint folder every day, with the previous day uploaded during the current day. A day’s worth of data contains as many files as key on-off events occur. The Datik logger collecting complementary GPS data is accessed live via an API that refreshes every 5 seconds. Chargepoint telemetry will be shared via SharePoint or a similar system. Questionnaires are filled in manually, scanned and emailed to CERTH and Cenex.

5.2 Ford use case

Vehicle telemetry will be shared via SharePoint, while the sharing method for chargepoint telemetry is still to be decided but it is likely to be SharePoint or a similar system. Questionnaires will be filled in manually, scanned and emailed to CERTH and Cenex.

5.3 DAF use case

As DAF joined the project recently, the exact methods of data sharing are still being decided. However, we expect to use SharePoint or a similar system to be aligned with the other OEMs. Questionnaires will be filled in manually, scanned and emailed to CERTH and Cenex.

6 DATA PROCESSING

A data processing pipeline is generally defined as the process through which data flows in its life cycle from its creation to the dissemination and reporting of its insights. In this project it has been designed as a modular system with its parts being bespoke to accommodate different logging systems and varying requirements of each data type. The number of different vehicle models and resulting disparate data sets mean that data is processed in a variety of ways to be able to report results on a like-for-like basis. The modular nature of our data processing pipeline will enable this to happen.

6.1 Data handling and reliability

The raw data is processed via Python scripts to extract the information of importance to the project into a working data set. This working data is used as the source for all higher-level analysis and reporting. The working data set is stored in a database hosted on Cenex's data processing server. It is structured in a standard but extensible way to ensure it is compatible with our data processing pipeline, but also flexible enough to be able to meet differing project requirements.

As part of the data handling process, Cenex will set up systems to detect missing data, faulty loggers and invalid data. These checks will comprise both systems already in place as part of our existing suite of data collection and processing software, and custom checks specific to this project. We will receive alerts if there is a problem so we can investigate and organise a solution to the issue. The checks we will perform, and actions we will support include:

- Identifying patterns of usage and flagging if data has not been received for an unusually long time, which might indicate a non-functional logger. Action can be taken to verify a faulty logger (as opposed to the vehicle not being used, e.g. due to maintenance) and to repair or replace the logger as necessary.
- Ensuring that all signals are present and received at the expected rate. Any deviations may point to a faulty connection to a specific CAN bus, or a potential logger malfunction.
- Validating the values of data received to ensure they are within sensible limits. Telemetry systems often produce spurious data, e.g. a default very high or low value, an invalid identifier if a sensor is not working, a signal not applicable at a certain time, or simply junk data at the start or end of a session. Cenex's data processing pipeline identifies such invalid data and either automatically removes it or highlights it as suspicious so that such data does not find its way into the final analysis.

6.2 Data harmonisation

The project demonstrators produce multiple data sets from various sources, and therefore harmonisation will be required to enable appropriate extraction of insights in future steps. Examples of harmonisation include:

- Timestamp alignment of different data sources on the same use case, e.g. in the Irizar use case: vehicle CAN data logger, Datik's GPS data logger, and K2's chargepoint OCPP data.
- Alignment in data frequency, as even data from the same source (e.g. CAN bus) will have different refresh frequencies depending on the signal being logged.
- Variable name standardisation across the three use cases. For example, vehicle telemetry signals have different names for each OEM, but they will be standardised to facilitate sharing the processed data with WP8 partners for KPI evaluation.
- Data format standardisation. Data items will be received in a variety of file types, therefore they will be converted to common formats suitable for both data processing using Cenex's algorithms, and for onwards data sharing with WP8 partners for KPI validation.

6.3 Resilience

Cenex will ensure there are multiple copies of data in multiple server locations to reduce the impact of failures. Copies of the raw and processed data sets are all made as part of nightly backups to our internal on-site file systems. This means if there is a failure of the data processing server, we have quick access to all data for easy recovery. Furthermore, our entire internal filesystem is backed up live to a commercial off-site mirror which includes versioned changes ensuring that no data will be lost even in the case of a catastrophic failure. Finally, the data processing pipeline is designed in a way that we can regenerate all working data and results from the raw data automatically, so providing we have a copy of the original raw data, all processed results can easily be recalculated.

6.4 Data security and privacy

Cenex has considerable experience handling data from many projects and much of this data is regarded as personal. We pay particular attention to the metadata linking the data to specific vehicles, and GPS location in telemetry data as it is usually the most sensitive signal in such data sets. However, we store and process all data with the same high level of protection to meet or exceed General Data Protection Regulation (GDPR) requirements. Where possible, reported data is aggregated and anonymised to avoid identifying individual participants. Furthermore, Cenex has Cyber Essentials Plus certification for our IT security procedures.

7 OUTWARDS DATA TRANSFER

At the time of writing, Cenex is working with TNO to agree on the best way and format for sharing clean, processed, standardised, aligned, harmonised and, where applicable, anonymised data. This data should be in the best position to input into the KPI calculations as part of WP8 (T8.2 and 8.3). The expected format is csv, and the expected sharing platform is Microsoft SharePoint. The expected data sharing frequency is once a week as KPIs are not required to be calculated more frequently and are based on historical datasets.

8 TIMELINES

Cenex develops data processing scripts as soon as sample datasets are shared by data owners. As stated in section 4.2, some datasets are not available yet due to delays in vehicle deployments. Once all data processing scripts are tested and ready, the actual data processing (cleaning, harmonisation, etc.) takes reasonably little effort and time, as most of the effort is usually allocated to writing data analysis algorithms. These algorithms will continue to be written between December 2025 and March 2026.

At the time of writing, Cenex has already started to transfer processed data to TNO to check if it has the suitable format and quality for T8.2 activities. Weekly processed data transfers are expected to start in full swing during January 2026. A data reception cut-off has tentatively been set as 30th April 2026 to allow enough time for KPI evaluation in tasks T8.2 and T8.3 before the project finishes in September 2026. This means that data received after the cut-off date will not be processed and hence will not count towards KPI evaluation.

9 REFERENCES

References were only made to previous project deliverables. Please refer to the appropriate deliverable: either on the project SharePoint if you are a project partner and the deliverable is marked as “sensitive”, or on the European Commission portal for publicly accessible deliverables.