D2.1. BERTHA Data Model

Main autors: Dr. Steve Pechberti, Dr. Carolina Perdomo (VED)

Co-authors: J.M. Belda (IBV), Antonio M. López (CVC), Thierry Bellet, Jean-Charles Bornard (UGE), Kristin Tovaas (AIT), Javier Romo (CID)

Version 1.0

05/07/2024

Funded by the European Union

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Climate, Infrastructure and Environment Executive Agency (CINEA). Neither the European Union nor the granting authority can be held responsible for them.

TABLE OF CONTENTS

EXECUTIVE SUMMARY 4

BERTHA's details........ 4

The BERTHA consortium........ 4

Project's summary........ 5

Document details........ 5

Document's abstract........ 6

Document's revision history........ 6

Terminology and acronyms........ 6

Disclaimer 7

1. INTRODUCTION AND OBJECTIVES........ 8

1.1. Bertha Data Model........ 8

1.2. Employed Reference Systems........ 9

1.3. Dynamic, Meta and Static Data 9

2. EMPLOYED REFERENCE SYSTEMS 11

2.1. Time Reference System........ 11

2.2. Spatial Reference System........ 12

2.2.1. WGS84 12

2.2.2. OpenDrive........ 13

2.2.3. Unreal Engine Coordinate System........ 14

3. DYNAMIC, META-, AND STATIC DATA........ 15

3.1. Static Data 15

3.1.1. Common Specification........ 15

3.1.2. Specification for Simulation Environment Description........ 15

3.1.3. Specification for FOT Environment Description 15

3.2. Metadata 16

3.2.1. Contextual Information Related to Acquisition........ 16

3.2.2. Components of Metadata........ 16

3.3. Dynamic Data 17

3.3.1. Driver Information 17

3.3.2. Ego Vehicle Information........ 18

3.3.3. Environment Information........ 19

Funded by the European Union

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Climate, Infrastructure and Environment Executive Agency (CINEA). Neither the European Union nor the granting authority can be held responsible for them.

3.3.4. Integration of the DBM with CARLA simulator for Dynamic Data Collection 19
3.3.5. Vehicle Dynamic Data Collection from Field Operational Tests (FOT) 21
3.3.6. Integration of CARLA in WP4 for assessing CCAM systems using DBM 22
4. OTHER EMPLOYED INFORMATION 23
4.1. Questionnaires 23
4.1.1. AIT Questionnaires 23
4.1.2. IBV Questionnaires 23
4.2. KPIs 24
5. CONCLUSIONS 24
6. REFERENCES 25
7. ANNEXES 27
7.1. Annex 1: Bertha Data Model 27

Funded by the European Union

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Climate, Infrastructure and Environment Executive Agency (CINEA). Neither the European Union nor the granting authority can be held responsible for them.

EXECUTIVE SUMMARY

BERTHA's details

Project name	BEhavioural ReplicaTion of Human drivers for CCAM
Project acronym	BERTHA
Grant Agreement number	101076360
Duration and dates	36 months (1 November 2023 – 31 October 2026)
Call and topic	HORIZON-CL5-2022-D6-01-03: Safe, Resilient Transport and Smart Mobility services for passengers and goods
Granting authority	European Climate, Infrastructure and Environment Executive Agency (CINEA), under the powers delegated by the European Commission
Official project website	berthaproject.eu

The BERTHA consortium

N°	NAME	ROLE	COUNTRY
1	INSTITUTO DE BIOMECANICA DE VALENCIA (IBV)	Coordinator	Spain
2	INSTITUT VEDECOM (VED)	Beneficiary	France
3	UNIVERSITE GUSTAVE EIFFEL (UGE)	Beneficiary	France
4	DEUTSCHES FORSCHUNGSZENTRUM FUR KUNSTLICHE INTELLIGENZ GMBH (DFKI)	Beneficiary	Germany
5	CENTRE DE VISIO PER COMPUTADOR (CVC-CERCA)	Beneficiary	Spain
6	ALTRAN DEUTSCHLAND SAS & CO KG (CAP)	Beneficiary	Germany
6.1	VORTEX - ASSOCIACAO PARA O LABORATORIO COLABORATIVO EM SISTEMAS CIBER-FISICOS E CIBERSEGURANCA (VOR)	Affiliated entity	Portugal
7	CONTINENTAL AUTOMOTIVE FRANCE SAS (CON)	Beneficiary	France
8	FUNDACION CIDAUT (CIDAUT)	Beneficiary	Spain
9	AIT AUSTRIAN INSTITUTE OF TECHNOLOGY GMBH (AIT)	Beneficiary	Austria
10	UNIVERSITAT DE VALENCIA (UVEG)	Beneficiary	Spain
11	EUROPCAR INTERNATIONAL	Beneficiary	France
12	F. INICIATIVAS, CONSULTADORIA E GESTAO, UNIPESSOAL, LDA (FI)	Beneficiary	Portugal
12.1	F. INICIATIVAS ESPANA I MAS D MAS I SLU (FI_ES)	Affiliated entity	Spain
13	PANASONIC AUTOMOTIVE SYSTEMS EUROPE GMBH (PAN)	Beneficiary	Germany
14	THE KOREA TRANSPORT INSTITUTE (KOTI)	Associated partner	Korea

Funded by the European Union

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Climate, Infrastructure and Environment Executive Agency (CINEA). Neither the European Union nor the granting authority can be held responsible for them.

Project's summary

The main objective of BERTHA is to develop a scalable and probabilistic Driver Behavioural Model based mostly on Bayesian Belief Networks (BBN). The DBM will be implemented on an open-source HUB (repository) to validate the technological and practical feasibility of the solution with industry, and provide a distinctive approach for the model worldwide scalability. The resulting DBM will be translated into a simulating platform, CARLA, using various demonstrations which will allow the construction of new driving models in the platform.

BERTHA will also include a methodology which, using the HUB, will allow to share the model with the scientific community, in order to facilitate its growth.

The project includes a set of interrelated demonstrators to show that the DBM can be used as a reference to design human-like, easily predictable and acceptable behaviours of automated driving functions in mixed traffic scenarios.

BERTHA is expected to go from TRL 2 to TRL 4. The requested EU contribution is €7,981,801. The consortium, formed by 14 entities from 6 countries, including South Korea, deems this Project as vitally relevant to the CCAM industry due to its impact for safer and more human-like CAVs and its market and societal adoption.

Document details

Deliverable type	Other
Deliverable n°	D2.1
Deliverable title	BERTHA Data Model
Lead beneficiary	VED
Work package and task	WP2 Task 1
Document version	V1.0
Contractual delivery date	M6
Actual delivery date	M9
Dissemination Level	PU - Public
Purpose	The deliverable describes all information that will be employed during the project, from DBM definition and validation to simulation construction for CCAM evaluation.

Funded by the European Union

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Climate, Infrastructure and Environment Executive Agency (CINEA). Neither the European Union nor the granting authority can be held responsible for them.

Document's abstract

This document presents the data model developed as part of the BERTHA project, designed to enhance the understanding and evaluation of Driver Behavioural Models (DBM) and Connected, Cooperative, and Automated Mobility (CCAM) systems through advanced simulations. Grounded in a review of existing standards and methodologies, including OpenDrive, this model systematically organises and categorises dynamic, meta, and static data essential for capturing the complexity of human and vehicle interactions in simulated environments. The data model is structured to support data collection, processing, and analysis, facilitating detailed experiments and validations. Key outcomes include the integration of sophisticated simulation platforms like CARLA and adapting standards for accurate environment replication.

Document's revision history

The following table describes the main changes done in the document since it was created.

REVISION	DATE	DESCRIPTION	AUTHOR (PARTNER)
V.0.1	2024/02/05	Initial Draft	Steve PECHBERTI (VEDECOM)
V.0.2	2024/05/02	Second Draft	Carolina PERDOMO (VEDECOM)
V.0.3	2024/05/16	Addressing initial feedback from internal review and partners' corrections.	Carolina PERDOMO (VEDECOM)
V.0.4	2024/06/28	CID's input	Carolina PERDOMO (VEDECOM)
V.1.0	2024/07/05	Submission check	Helios DE ROSARIO (IBV)

Terminology and acronyms

TERM/ACRONYM	EXPLANATION
CAV	Connected Autonomous Vehicles
CCAM	Connected, Cooperative, and Automated Mobility
CINEA	Climate, Infrastructure and Environment Executive Agency
DBM	Driver Behavioral Model
EC	European Commission

Funded by the European Union

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Climate, Infrastructure and Environment Executive Agency (CINEA). Neither the European Union nor the granting authority can be held responsible for them.

HAV	Human Autonomous Vehicle
DBM	Driver Behavioural Model
FOT	Field Operational Test
UTS	Coordinate Universal Time
WGS	World Geodetic System
ADAS	Advanced driver-assistance systems

Employed Units

Unit	Symbol	Definition
Second	s	
Meter	m	
Meter per second	$m.s^{-1}$	
Meter per second squared	$m.s^{-2}$	
Decibels	dB	

Disclaimer

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Climate, Infrastructure and Environment Executive Agency (CINEA). Neither the European Union nor the granting authority can be held responsible for them.

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Climate, Infrastructure and Environment Executive Agency (CINEA). Neither the European Union nor the granting authority can be held responsible for them.

1. INTRODUCTION AND OBJECTIVES

1.1. Bertha Data Model

Generally, a data model represents experimental data undergoing several processing and analysis stages [1]. Data modelling aims to illustrate the types of data used and stored within the system, the relationships among these types, how the data can be organised, and its formats and attributes.

This document is inside the first phase in the research data lifecycle according to the UK Data Archive [2]: the **Planning**. Therefore, in the present document, we intend to identify and characterise the research data to be collected.

In the case of BERTHA, the Data Model lists all information employed during the development of the Driver Behavioural Model (DBM) modules and the evaluation of the Connected, Cooperative, and Automated Mobility (CCAM) system with simulation.

This DBM is intended to be a robust and scalable probabilistic model that comprehensively encompasses critical facets of human driving performance. The model will be structured around distinct modules: **Perception** [3], **Risk Awareness** [4], **Decision-Making** [5], **Affective** [6], and **Motor** [7].

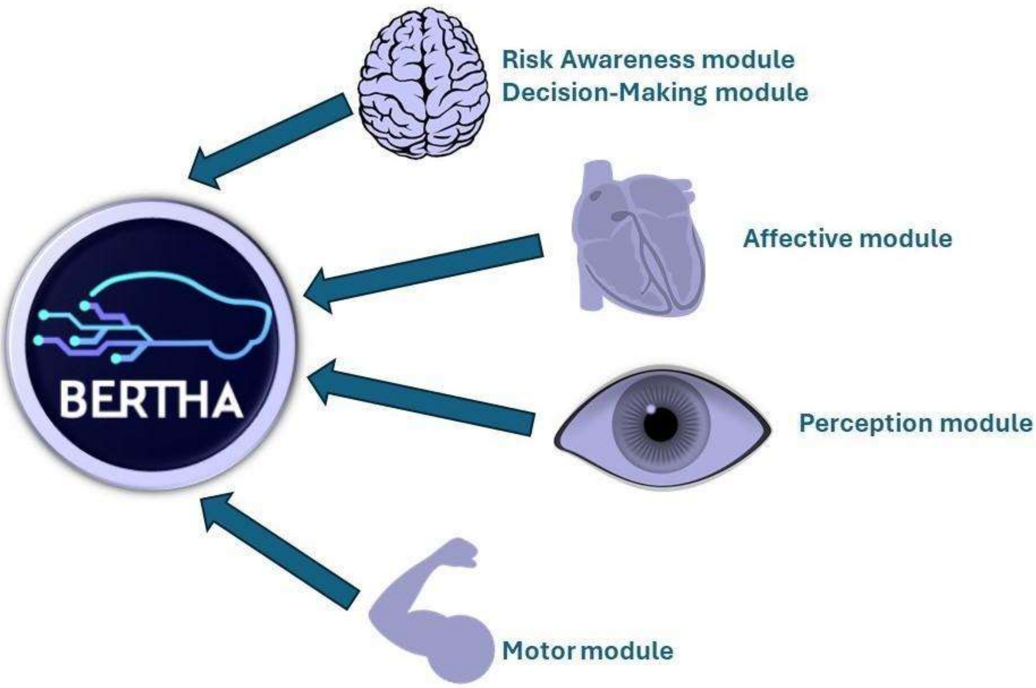

Figure 1. The DBM stands surrounded by four arrows pointing toward it, each representing a crucial module: Perception, Risk Awareness/Decision-Making, Affective, and Motor.

Funded by the European Union

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Climate, Infrastructure and Environment Executive Agency (CINEA). Neither the European Union nor the granting authority can be held responsible for them.

The DBMis the foundation for the project's ambitious goal of understanding human driving performance and comprehensively evaluating CCAM systems.

The methodology outlined in **this report** aims to define the set of data used in the project and the methods necessary for collecting them, facilitating relevant exploitation. Therefore, partners such as VED, IBV, UV, CON, DFK, UGE, AIT, and CVC, who are engaged in various aspects of the project, would benefit from the data model's comprehensive representation of driver behaviour and its integration into the project's objectives.

1.2. Employed Reference Systems

Depending on acquisition type, several reference systems are defined; these systems are explained in section 2.

1.3. Dynamic, Meta and Static Data

In the case of project BERTHA, different types of data will be generated, manipulated, analysed, transformed, and published. To build the data model effectively, we need to define three categories of information:

- **Data acquired during experiments**: This category encompasses the raw data gathered during experimental runs, crucial for fine-tuning the models and algorithms. Then, this data acquired is the general information to describe the experience. These data include sensor readings, vehicle trajectories, and other real-time information directly obtained during the experiment.
- **Contextual information related to acquisition**: This category encompasses data necessary to understand the experiment's circumstances. This contextual information includes details such as the type of driver involved, meteorological conditions, road conditions, and other factors that might impact the experiment's outcome.
- **Common information across experiments**: This includes consistent data across multiple experimental runs. Such information typically includes details about the simulated use case, characteristics of the road infrastructure, geographical location, and any other fixed attributes essential for understanding the experiment's context.

These three categories of information can be further categorised into dynamic, meta, and static data, each serving distinct but complementary roles in constructing a comprehensive data model. By understanding and categorising these types of information into dynamic, meta, and static data, we can develop a structured approach to data management, ensuring that all aspects of the experiment are adequately captured and utilised for model development and analysis:

- **Dynamic Data**: The data evolving during the experiment is categorised into three primary categories: **human driver data** (including physiological information, gaze, and

Funded by the European Union

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Climate, Infrastructure and Environment Executive Agency (CINEA). Neither the European Union nor the granting authority can be held responsible for them.

others), **vehicle state**, and **environmental factors**. This data is crucial for understanding the real-time dynamics of the driving scenario and is structured to be collected and analysed during the experiment.

- **Metadata**: Data about the data, providing contextual information about the experiment, including parameters related to the use case, time references (e.g., UTC), and space references (e.g., WGS or OpenDrive format). Metadata also includes information about questionnaires conducted with drivers.
- **Static Data** refers to the global information about the experience, such as the use case description and environment characterization. This data remains constant throughout the experiment and serves as a foundational aspect of the data model.

In section 4, the document will delve into each data category concerning BERTHA, providing a comprehensive explanation of their significance and relevance.

Funded by the European Union

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Climate, Infrastructure and Environment Executive Agency (CINEA). Neither the European Union nor the granting authority can be held responsible for them.

2. EMPLOYED REFERENCE SYSTEMS

There is a need for information and data that can be aggregated for standard processing. However, several prerequisites exist for such aggregation. The first of these, as presented in this section, pertains to the fundamental requirement of using:

- Time reference system.
- Spatial reference system.

This employment emphasises the importance of understanding the context in which data are acquired.

2.1. Time Reference System

Maintaining a cohesive temporal framework for all experimental data in BERTHA is important for facilitating comprehensive analysis and evaluation. To streamline this process and ensure seamless integration of data from simulations and Field Operation Tests (FOT), they must share a unified temporal space. This simplifies the relationship between different datasets and enables precise treatment and evaluation of temporal gaps down to milliseconds.

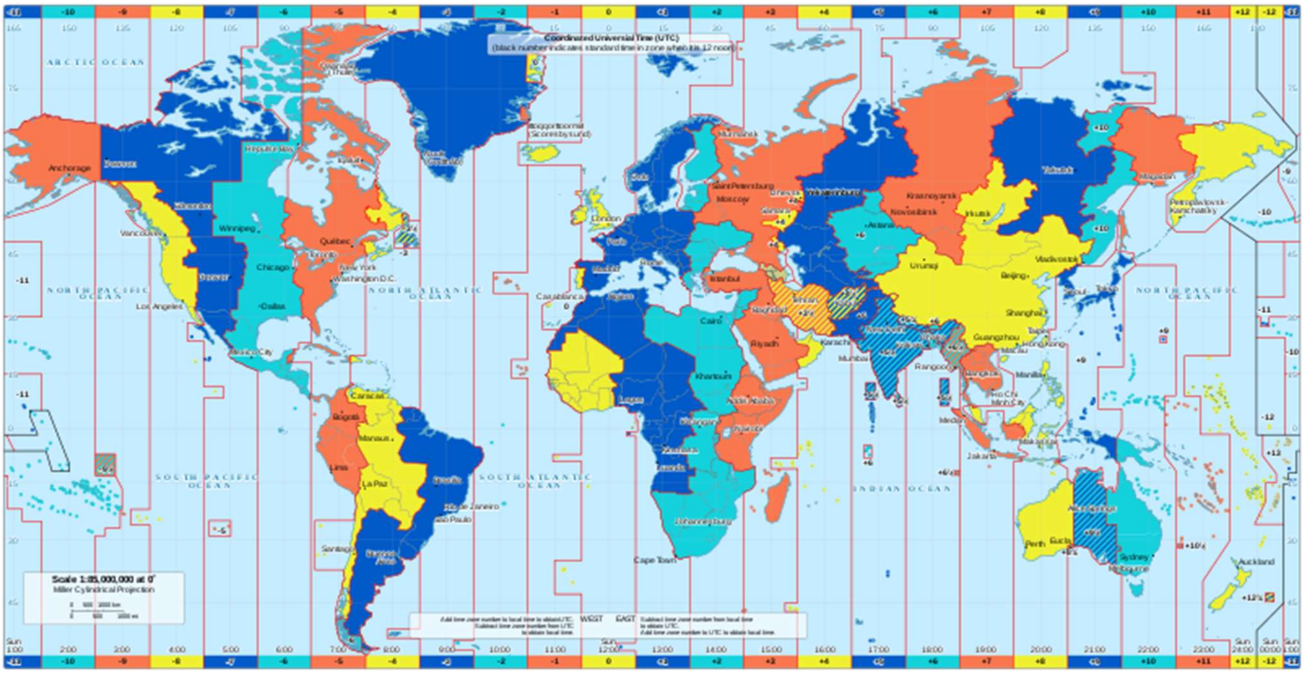

Figure 2. Time zones of the world in UTC.

At the heart of the BERTHA project's time reference system is adopting a standardised time reference: **Coordinate Universal Time** (UTC) [8]. As our primary temporal framework, UTC establishes a consistent and reliable basis for timestamped data across all recording solutions. This approach guarantees synchronisation and accuracy and enhances the efficiency and reliability of our data management practices. In conjunction with recording equipment

Funded by the European Union

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Climate, Infrastructure and Environment Executive Agency (CINEA). Neither the European Union nor the granting authority can be held responsible for them.

equipped with a time server, UTC ensures the maintenance of a synchronised time base, facilitating robust analysis and decision-making processes.

While UTC remains our foundational time reference system, we acknowledge the relevance of **Local time** for certain driving scenarios, particularly in the context of environmental and operational conditions. Local time is also important for accurately determining day and night cycles, which can significantly impact data interpretation and decision-making processes. For instance, local time considerations could be integrated to analyse driver behaviour during different times of the day. This ensures that our approach remains flexible and adaptable to the specific needs and contexts of the project as it evolves.

2.2. Spatial Reference System

The BERTHA project employs a specific spatial reference framework: a **global reference** frame [9] (the **Geo-Reference Frame**), in which all spatial data is universally representable.

Most geolocation systems, including those in the BERTHA project, are used for autonomous vehicle research. Therefore, depending on the experimental context in BERTHA, the spatial reference frame utilised could be:

- WGS84 for Field Operational Tests (FOT).
- OpenDrive Framework, and Unreal engine coordinate system for simulation testing.

2.2.1. WGS84

WGS84 [9] serves as the global reference framework employed by GPS technology via the broadcast ephemerides. Initially established based on the coordinates of stations within the GPS Ground Segment ascertained through various space geodesy techniques, it has since undergone periodic updates.

WGS84 is associated with GPS receivers, widely used across Europe to address localization issues. The data is represented within a spherical reference system in their non-projected form.

Funded by the European Union

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Climate, Infrastructure and Environment Executive Agency (CINEA). Neither the European Union nor the granting authority can be held responsible for them.

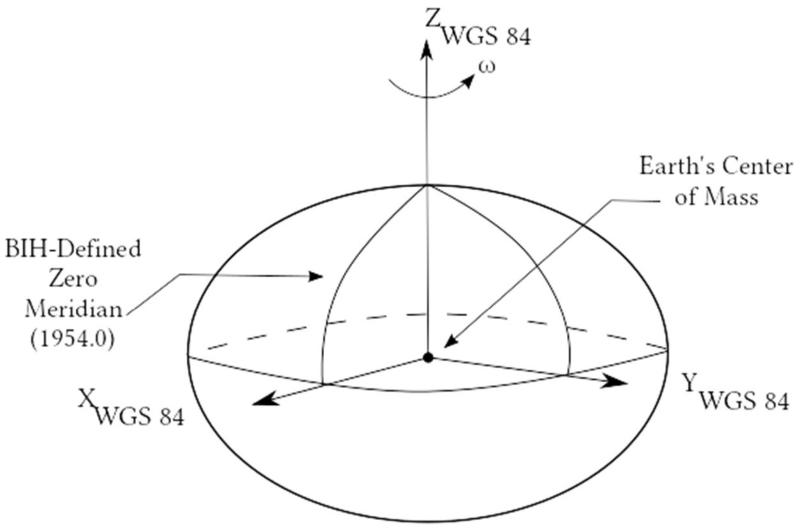

Figure 3. WGS84 reference frame.

2.2.2. OpenDrive

The OpenDrive reference system, often known as OpenDrive, is a standard specification used predominantly to simulate and model road networks and environments for driving simulations. This standard provides a structured way to define road network geometry, topology, and semantics [10].

OpenDrive aims to facilitate the exchange and use of digital road information between software applications and systems [10]. It allows for detailed and accurate representation of road elements such as lanes, road signs, traffic lights, and other infrastructure relevant to driving and traffic management [11]. The specification covers both urban and rural road settings. It is adaptable to various levels of detail, making it suitable for different types of simulations, from high-level strategic modelling to detailed traffic and driving simulations.

One of OpenDrive's key features is its ability to define road networks in a hierarchical structure [12]. The OpenDrive reference sytem is essential in developing autonomous driving systems, advanced driver-assistance systems (ADAS), and traffic engineering and simulation.

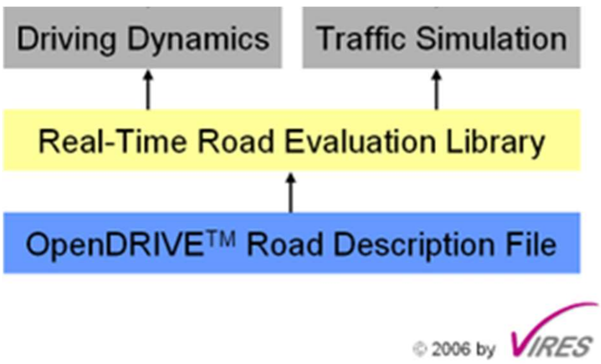

Figure 4. Typical incorporation of an OpenDrive file into a simulation application.

Funded by the European Union

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Climate, Infrastructure and Environment Executive Agency (CINEA). Neither the European Union nor the granting authority can be held responsible for them.

Nevertheless, it is essential to clarify that Work Package 4 (WP4) has no plans to create new maps using OpenDrive for simulations. Instead, the focus will be on allocating specific scenarios within the existing CARLA maps, which have already been tailored to fit the defined use cases from Work Package 1 (WP1).

2.2.3. Unreal Engine Coordinate System

CARLA is developed as an open-source framework on top of Unreal Engine [13], as a consequence, it adheres to the coordinate conventions of Unreal Engine. According to the CARLA documentation [14], all sensors use Unreal Engine's coordinate system, where the x-axis points forward from the vehicle, the y-axis points right, and the z-axis points up. Sensor coordinates are returned in local space. In other words, CARLA uses a **left-handed coordinate system**.

Funded by the European Union

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Climate, Infrastructure and Environment Executive Agency (CINEA). Neither the European Union nor the granting authority can be held responsible for them.

3. DYNAMIC, META-, AND STATIC DATA

As mentioned in the previous Section. 2.3, we have three families for the information in BERTHA.

3.1. Static Data

This data describes the experiment in terms of use case and road topology. The experiment is designed around the use cases defined in WP1. The road topology for the experiment is defined according to the OpenDrive standard, which includes specific descriptions of road geometry, lane widths, lane numbers, and signalling information. This setup is essential to simulate real-world driving scenarios accurately and assess automated driving systems' behaviour and safety.

3.1.1. Common Specification

3.1.1.1. Associated Use Cases

The associated use cases were selected in the WP1. The selection of use cases includes the following use cases with two use cases for highway settings, two use cases for urban settings, and one use case that can be applied both in highway or urban settings:

- Collision risk avoidance – highway settings.
- Insertion on highway – highway settings.
- Pedestrian crossing – urban settings.
- Left turn at urban intersection (with traffic lights) – urban settings.
- Pull back in – highway or urban settings.

The full description of the selected use case can be found in the deliverable 1.1 of BERTHA.

3.1.2. Specification for Simulation Environment Description

It specifies the relevant attributes to create a detailed and accurate representation of road systems. For more information about how this is done in OpenDrive, check reference [11].

3.1.3. Specification for FOT Environment Description

For Field Operational Tests (FOTs), the environment description aims to replicate real-world conditions as accurately as possible to ensure representative data.

Characterising the FOT environment is categorised under 'Static Data' as it encompasses unchanging infrastructure elements essential for maintaining consistent test conditions. These elements include road topology, geometry, lane width, lane number, signalization, and the positioning of UUID objects. This static data has already been documented in the accompanying Excel spreadsheet. The stable nature of this information ensures that each

Funded by the European Union

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Climate, Infrastructure and Environment Executive Agency (CINEA). Neither the European Union nor the granting authority can be held responsible for them.

FOT scenario can be consistently replicated and analysed, providing reliable and accurate data.

A comprehensive methodology to fully define the FOT environments will be detailed in the upcoming report D2.5.

3.2. Metadata

As mentioned in Section. 2, metadata provides essential contextual information that aids in understanding the data acquired during the experiment. It describes the circumstances and parameters of data collection, which are crucial for interpreting the results. It is the "*data about data*," encompassing various details that offer insights into the nature and context of the experiments in BERTHA. This includes use case parameters, time and space references, and information from driver questionnaires. It is fundamental to understand the nuances of the experiment's setup and execution.

3.2.1. Contextual Information Related to Acquisition

This Section includes crucial details needed to understand the experimental setup:

- **Driver Profile**: Information such as age, class, and experience level.
- **Environmental Conditions** include meteorological conditions, road conditions, and other environmental factors that might affect the experiment.
- **Experiment Settings**: Use case parameters like speed approach and initial distance, as well as environmental settings like time of day and weather conditions.

3.2.2. Components of Metadata

Based on previous details, in the Excel spreadsheet that accompanies this document, we can find the following information:

I. **Human Driver Information**:
  A. *Identity*:
    1. UUID (Unique User Identifier): A unique string that identifies each driver involved in the experiment.
      a) Description: The unique identifier for the driver.
      b) Value Type: String.
      c) Unit: N/A
  B. *Condition*: Details regarding the driver's condition are recorded here, though specifics are yet to be defined.

II. **Environment**:
  A. *Reference*: Links to the detailed implementation of static data definitions used during the experiment.

Funded by the European Union

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Climate, Infrastructure and Environment Executive Agency (CINEA). Neither the European Union nor the granting authority can be held responsible for them.

III. **Use case**:

A. *Reference UC*: References the specific use case implemented during the experiment.

B. *Use Case Parameters*: This details the specific parameters associated with the Reference UC. Each parameter is crucial for defining the experiment's scope and expected outcomes.

3.3. Dynamic Data

Dynamic data evolves during an experiment, capturing variable changes over time. This data is instrumental in understanding behaviours and interactions within the experimental environment. To ensure consistency and comprehensiveness across different experimental setups, dynamic data is structured into three primary categories: **Human Driver**, **Vehicle State**, and **Environmental Factors**. Each category includes subsets of parameters that must be acquired in tandem to maintain data integrity and relevance.

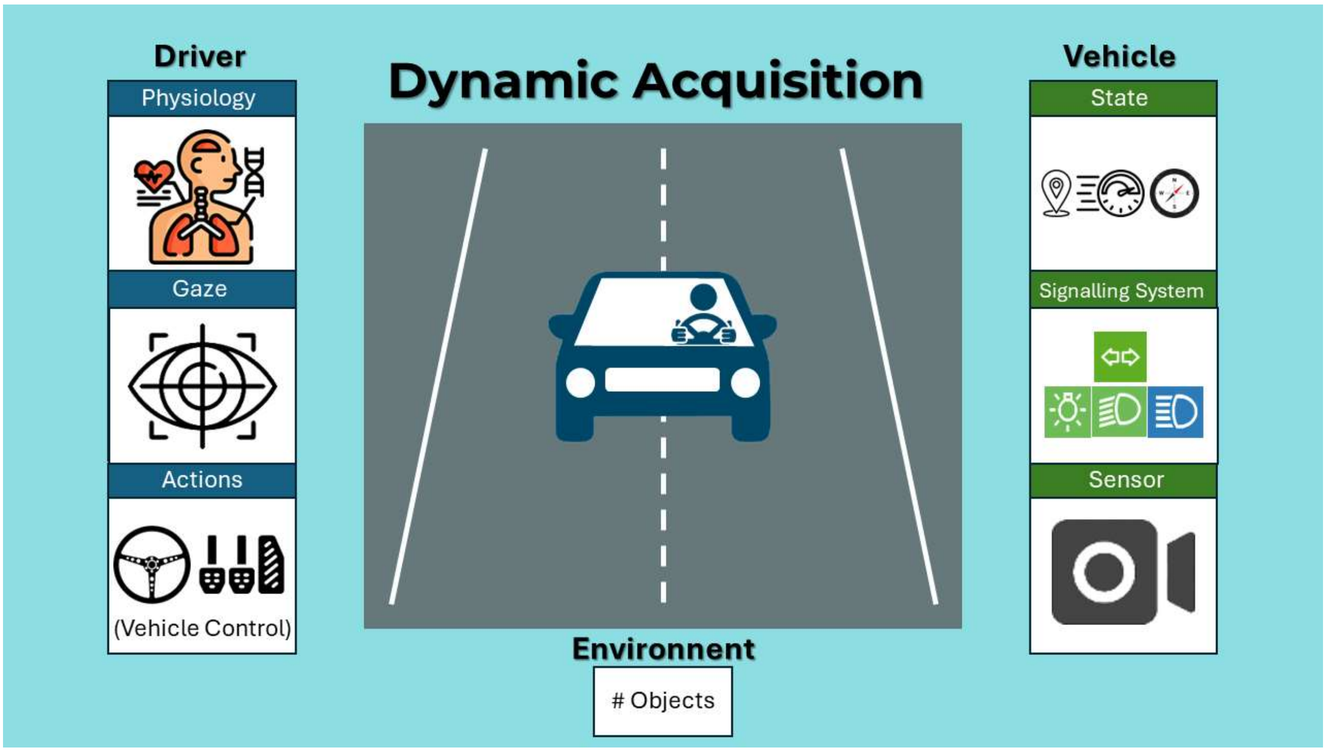

Figure 5. Schematic representation of Dynamical data in BERTHA.

3.3.1. Driver Information

Dynamic data on the human driver is divided into physiological and gaze information, providing insights into the driver's condition and reactions during the experiment.

Funded by the European Union

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Climate, Infrastructure and Environment Executive Agency (CINEA). Neither the European Union nor the granting authority can be held responsible for them.

3.3.1.1. *Physiological Information*

- **Definition:** Measurements that reflect the physiological state of the driver, such as heart rate variability, EEG signals, and cortisol levels.
- **Usage**: These parameters help assess the driver's stress levels, attention, and overall health during driving tasks.
- **Associated To**: Driver's performance and reaction to different driving scenarios.
- **Requested Acquisition**: High precision and frequency are required to capture transient physiological responses accurately.

Example:

I. **Parameter**: Heart Rate Variability (HRV)
   A. *Definition*: Variation in the time intervals between consecutive heartbeats.
   B. *Value*: Integer.
   C. *Unit*: ms.
   D. *Usage*: Indicates the driver's stress level and autonomic nervous system activity.
   E. *Associated To*: Driver's stress and alertness.
   F. *Requested Acquisition*: Measurement precision of ±1 ms, acquisition frequency of 1 Hz.

For the rest of the parameters, refer to the Excel document that accompanies this report.

3.3.1.2. *Gaze Information*

- **Definition**: Data concerning the direction and focus of the driver's gaze, including metrics like saccade rates and pupil diameter.
- **Usage**: Gaze data is crucial for understanding the driver's attention distribution and cognitive load.
- **Associated To**: Driver's visual attention and situational awareness.
- **Requested Acquisition**: High-resolution data capturing is necessary to monitor subtle changes in gaze orientation and eye movement.

3.3.2. Ego Vehicle Information

Dynamic data from the vehicle includes metrics related to its state, such as position, velocity, and control inputs.

- **Definition**: Continuous data capturing of the vehicle's dynamics, system status, and interaction with the control inputs.
- **Usage**: To analyse the vehicle's behaviour under various operational scenarios and its interaction with the driver's inputs.
- **Associated To**: The vehicle's performance and safety metrics.
- **Requested Acquisition**: Data must be precise and frequent to reflect vehicle dynamics and control responsiveness accurately.

Funded by the European Union

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Climate, Infrastructure and Environment Executive Agency (CINEA). Neither the European Union nor the granting authority can be held responsible for them.

Example:

I. **Parameter**: Vehicle Velocity
   A. *Definition*: Speed of the vehicle along longitudinal and lateral axes.
   B. *Value*: Float; Unit: m/s
   C. *Usage*: Critical for assessing the vehicle's dynamic response and stability.
   D. *Associated To*: Driver inputs and road conditions.
   E. *Requested Acquisition*: Measurement precision of ±0.05 m/s, acquisition frequency of 10 Hz.

3.3.3. Environment Information

This includes data about the surrounding environment, focusing primarily on the interactions and dynamics of surrounding objects.

- **Definition**: Information about static and dynamic objects within the vehicle's operational environment, such as their position, speed, and heading.
- **Usage**: Essential for understanding the context of vehicle operation and developing safety systems.
- **Associated To**: The vehicle's navigation and collision avoidance systems.
- **Requested Acquisition**: Dynamic descriptions must reflect real-time changes in the environment with high fidelity.

Example:

I. **Parameter**: Surrounding Objects
   A. *Definition*: Data on nearby objects, including their unique identifiers, positions, and movement patterns.
   B. *Value*: Multiple data types.
   C. *Unit*: varies (e.g., metres for position, m/s for speed)
   D. *Usage*: To monitor nearby objects' proximity and relative velocity for collision prediction.
   E. *Associated To*: Environmental awareness and driver assistance features.
   F. *Requested Acquisition*: Real-time tracking with precision and resolution adequate for safe operational decisions.

3.3.4. Integration of the DBM with CARLA simulator for Dynamic Data Collection

The DBM is connected to the CARLA simulator at a global level, enabling the model to be fed with dynamic data. This connection is established through the **PERCEPTION module**, which understands the scene within CARLA via the privileged information or by analysing sensor raw data. Additionally, the DBM produces outputs via the **MOTOR CONTROL module**, which influences the CARLA ego-vehicle controls, such as pedals, steering wheel, and turn signals.

Funded by the European Union

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Climate, Infrastructure and Environment Executive Agency (CINEA). Neither the European Union nor the granting authority can be held responsible for them.

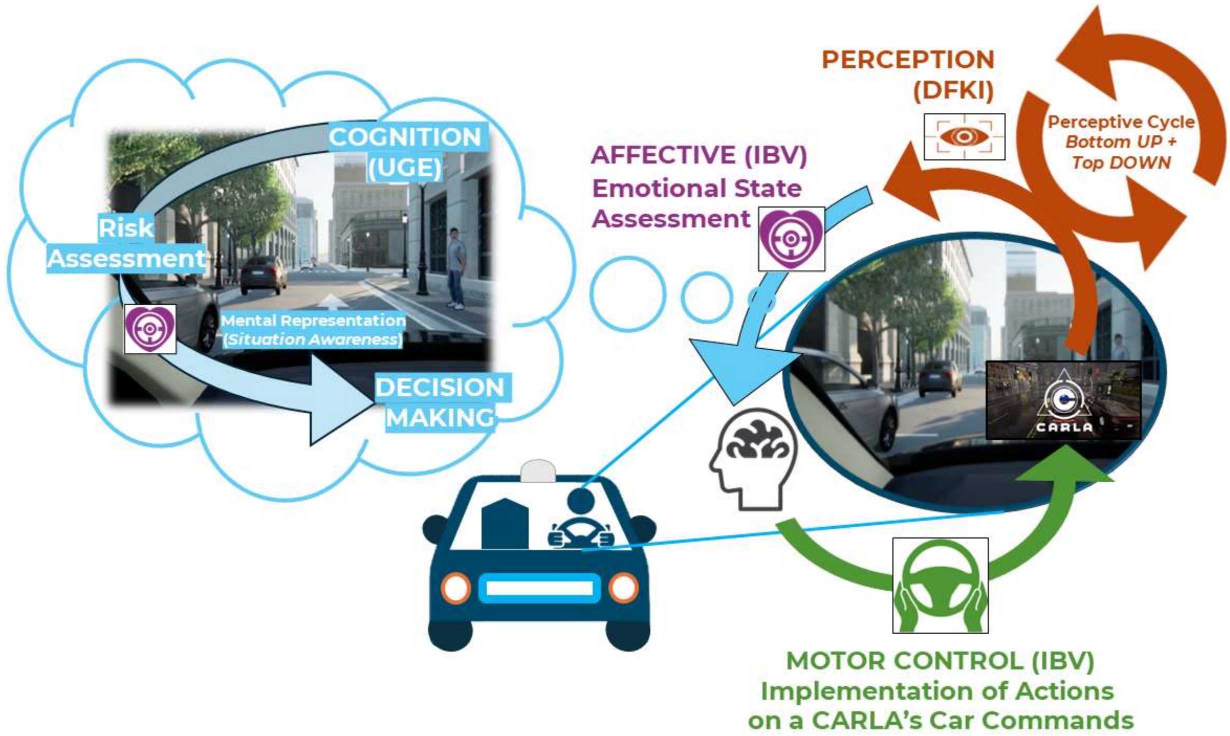

*Figure 6. DBM functional architecture: The DBM's main inputs (the dynamic data perceived in the CARLA simulation environment) are collected by the **PERCEPTION module**. The DBM's main outputs are generated by the **MOTOR CONTROL module**.*

As shown in Fig.7, the DBM comprises five interconnected modules that exchange data internally. These exchanges are crucial for analysing situations, assessing risks, making decisions, planning actions, and evaluating effects. Each module within the DBM can use inputs from other modules and generate outputs for them, facilitating a comprehensive and dynamic interaction.

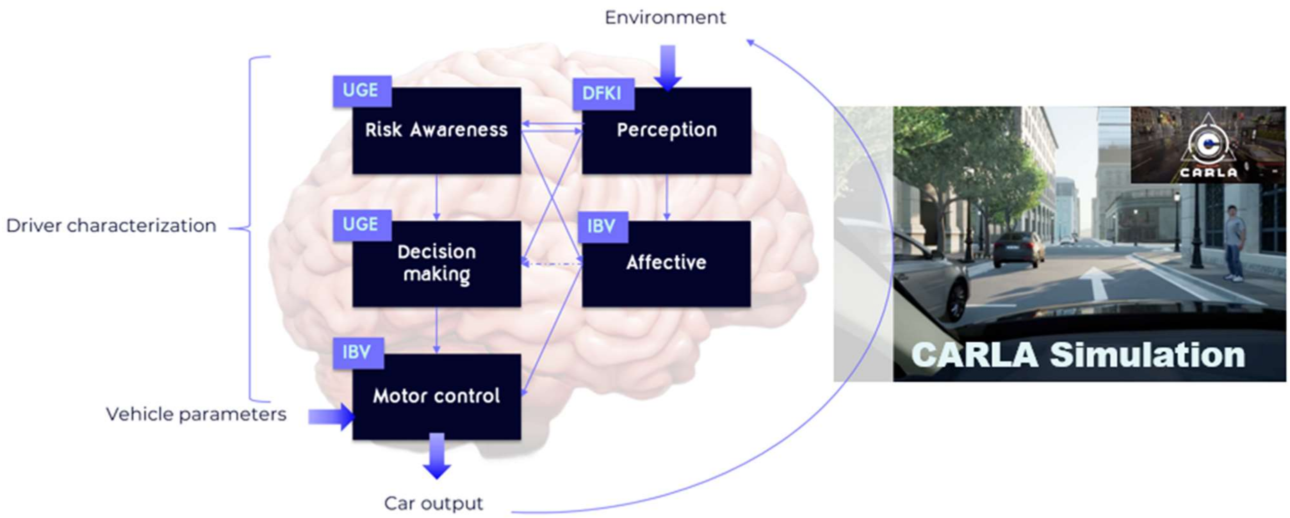

Figure 7. Data exchanges between DBM modules.

Funded by the European Union

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Climate, Infrastructure and Environment Executive Agency (CINEA). Neither the European Union nor the granting authority can be held responsible for them.

As shown in Fig. 7, the **PERCEPTION module** is designed to model a driver's visual understanding of their surroundings. It will rely on a neural network trained to predict the driver's visual attention, using either privileged information from CARLA or by analysing sensor raw data. Additionally, inputs from the risk awareness and decision-making module can be used to adjust the quality of the attention outputs

The **RISK AWARENESS and DECISION MAKING modules** evaluate traffic situation risks and stimulates cognitive processes that support drivers' decision making based on situational risk levels. Inputs from the perception module such as road infrastructure details and road environment hazards. Output may direct visual attention, indicate the perceived level of risk to the **AFFECTIVE module**, and suggest driving manoeuvres (example: braking) to the motor control module, which execute these actions.

Now, the **AFFECTIVE module** addresses psychological and emotional responses influenced by internal factors (driver characteristics) and external factors (traffic, environment). As shown in Fig. 7, the emotional state is influenced by risk and perception inputs. Finally, the **MOTOR CONTROL** module manages the driver's physical actions, controlling speed and steering. Calibration and testing of this module require use cases with situations demanding quick reactions, such as braking or avoiding obstacles.

It must be emphasised that the internal data exchanges within the DBM are beyond the focus of this deliverable, which primarily addresses CARLA-DBM exchanges. These internal interactions are explained in the deliverable D1.5: " *Model Framework*", scheduled for submission by T12.

3.3.5. Vehicle Dynamic Data Collection from Field Operational Tests (FOT)

In the context of FOT, CID specialises in field measurements using vehicular platforms. This subsection delineates the key dynamic data metrics that can be provided to partners conducting simulations, aiding in the comprehensive analysis and validation of simulated models against real-world scenarios.

Until now, one of the primary methods tested for data collection involves using sensors and instruments integrated within the vehicle. This approach allows for the capture of a wide array of dynamic metrics. Some of the notable metrics include:

- **Brake Pressure**: Measurement of the pressure applied to the brakes is essential for understanding braking performance and response under various conditions.
- **Steering Angle**: The angle of the steering wheel can be used to determine the turning angles of the vehicle's left and right wheels.
- **Other Vehicle Dynamics**: Data encompassing the acceleration position and the car's acceleration in multiple axes (X, Y, Z).
- **Lights Activation**: Monitoring of the car's lights, which can indicate various operational states and driver actions.

Additionally, a plan is to enhance the vehicle's localization capabilities by implementing a professional GPS, providing referential data in the WGS84 coordinate system, as mentioned in Sub-section 2.2.1. This will include **localization data** as position coordinates (X, Y, Z).

Funded by the European Union

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Climate, Infrastructure and Environment Executive Agency (CINEA). Neither the European Union nor the granting authority can be held responsible for them.

This comprehensive dataset will provide simulation partners with a foundation to validate their models against real-world data, ensuring the accuracy and reliability of their simulations. As of the date this report is written, CID continues to enhance its measurement capabilities, focusing on delivering the most relevant metrics.

3.3.6. Integration of CARLA in WP4 for assessing CCAM systems using DBM

In Work Package 4 (WP4), CARLA is strategically employed to test the Driver Behavioural Models (DBMs) developed in BERTHA. The objective in WP4 is to integrate DBMs from WP1 into CARLA simulations to assess their potential to enhance CCAM systems, including autonomous driving systems and specific ADAS features. It is important to note that WP4 does not enhance the DBMs; instead, it focuses on evaluating how these models perform within simulated environments to understand their impact on system effectiveness and reliability.

CARLA provides a sophisticated simulation platform that enables the detailed modelling of scenarios. This includes representing pedestrians and other objects within the simulation environment. In WP4, vehicles and pedestrian behaviour will be represented using standardised tools or standards, e.g; the SCENIC tool [15] or the OpenSCENARIO format [16], that can be connected with the available CARLA capabilities.

In using CARLA within WP4, while the simulator adeptly manages various environmental conditions such as sunlight, rain, and clouds, it currently does not support all weather scenarios (like snow).

Funded by the European Union

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Climate, Infrastructure and Environment Executive Agency (CINEA). Neither the European Union nor the granting authority can be held responsible for them.

4. OTHER EMPLOYED INFORMATION

4.1. Questionnaires

Questionnaires play a crucial role in gathering subjective data directly from participants, providing insights into their emotions, habits, preferences in driving, and various other factors. These tools are instrumental in understanding individual differences and contextualising other data types within the experiment.

4.1.1. AIT Questionnaires

In principle, the questionnaire forms for **driver characterization** in BERTHA are currently being formalised. These questionnaires will be primarily applied in T1.2 to gather sociological data and in T5.4 to assess the level of acceptance by different types of drivers in selected situations. In T1.2, this data will be captured from an international survey conducted across six countries, targeting at least 1,200 drivers to gather diverse insights into driver behaviour and reactions in specific scenarios. This data is essential for calibrating the Decision-Based Model (DBM) in various work packages. In T5.4, for each driver type, at least five representatives will test the CCAM systems in a driving simulator to collect driving performance data, followed by completing a questionnaire to assess the perception of the new systems and their influence on driver performance.

4.1.2. IBV Questionnaires

These questionnaires capture various variables, providing a deeper understanding of the drivers' profiles and conditions during simulations. Below is the proposed structure and categorization for these questionnaires from IBV. They will follow the next structure:

- **Driver UUID**: Unique identifier for each participant.
- **Data Type**: Specifies whether the data is gathered through **Questionnaires** or **Scales**.
- **Levels/items**: Indicates the complexity or depth of questions/scales used.
- **Privacy**: Level of sensitivity associated with the data.

We can also divide them into two types with different scales:

- **Baseline and Punctual Measurements**: These include assessments of emotions, fatigue, somnolence, mental stress, mental workload, and attention. Both baseline (general condition) and punctual (specific moment or condition) measurements are taken to compare changes over time or in response to specific conditions.
- **Health-Related Data**: In-depth questionnaires cover demographic data, health risk factors, and detailed health assessments to profile drivers comprehensively.

Funded by the European Union

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Climate, Infrastructure and Environment Executive Agency (CINEA). Neither the European Union nor the granting authority can be held responsible for them.

The detailed structure of the data collection by these questionnaires is in Annex 1 of the BERTHA data model report. This comprehensive data collection strategy ensures a framework for analysing and understanding the diverse factors influencing driver behaviour in general conditions and specific instances.

4.2. KPIs

Key Performance Indicators (KPIs) are metrics used to quantitatively assess driver performance and state aspects, such as alertness and response to driving tasks.

Example:

I. **PERCLOS (Percentage of Eye Closure)**:
   A. *Description*: A measure of the percentage of time a driver's eyes are closed, either fully or partially, during a specified observation period. It is commonly used to assess driver fatigue or drowsiness.
   B. *Equation*.
   C. *Unit*: Percentage (%)

II. **Usage**: PERCLOS monitors driver alertness during the experiment and can be a critical indicator of fatigue, influencing safety and performance outcomes.

All the Key Performance Indicators (KPIs) established during BERTHA's assessments and extended operation will be addressed in upcoming deliverables concerning the data methodology in Work Package 2 (WP2).

5. CONCLUSIONS

This deliverable defines the data of project BERTHA, its definition, classification, and common reference systems.

As this deliverable outlines, the data model described herein is fundamental to the BERTHA project's progress and success. It guides the experimental approaches and ensures all project partners have a solid foundation for data collection. This framework is vital for achieving the nuanced understanding required to advance CCAM technologies.

Funded by the European Union

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Climate, Infrastructure and Environment Executive Agency (CINEA). Neither the European Union nor the granting authority can be held responsible for them.

6. REFERENCES

[1] Antoniou, A. (2021). What is a data model? *European Journal for Philosophy of Science*, 11, 101. https://doi.org/10.1007/s13194-021-00412-2

[2] UK Data Archive. (n.d.). Life cycle of data. Retrieved from https://www.data-archive.ac.uk/

[3] Kapitaniak B, Walczak M, Kosobudzki M, Jóźwiak Z, and Bortkiewicz A. Application of eye-tracking in drivers testing: A review of research. *International Journal of Occupational Medicine and Environmental Health*. 2015;28(6):941-954. doi:10.13075/ijomeh.1896.00317.

[4] Horswill, M. S., & McKenna, F. P. (2004). Drivers' hazard perception ability: Situation awareness on the road. *A cognitive approach to situation awareness: Theory and application*, 155-175

[5] Bornard, J. C., Sassman, M., & Bellet, T. (2016). Use of a computational simulation model of drivers' cognition to predict decision making and behaviour while driving. *Biologically Inspired Cognitive Architectures*, 15, 41-50.

[6] Belda-Lois, J. M., Iranzo, S., Silva, J., Mateo, B., Palomares, N., Laparra-Hernández, J., & Solaz, J. S. (2020). The Estimation of Occupants' Emotions in Connected and Automated Vehicles. *In CHIRA* (pp. 262-267).

[7] Belda-Lois, J. M., De-Rosario, H., Pons, R., Poveda, R., Morón, A., Porcar, R., ... & Gómez, A. (2010). Can human movement analysis contribute to usability understanding? *Human movement science*, 29(4), 529-541.

[8] G Panfilo, F Arias. The Coordinated Universal Time (UTC). Metrologia, 2019, 56 (4), pp.042001.

10.1088/1681-7575/ab1e68. hal-02183608

[9] Crespi, M., Mazzoni, A., & Colosimo, G. (2015). Global and local reference frames. *Rendiconti Lincei, 26*. https://doi.org/10.1007/s12210-015-0435-0

[10] Navigation Data Standard. (2021, April 22). Data for all: Simulating logical road networks with OpenDRIVE. Retrieved from https://nds-association.org/opendrive/

[11] Dupuis, M., & et al. (2010). Public license statement for OpenDRIVE format specification. In OpenDRIVE Format Specification (VI2010.002) (p. 3). Retrieved from https://www.asam.net/standards/detail/opendrive/

[12] Bellet, T., Mayenobe, P., Bornard, J. C., Gruyer, D., & Claverie, B. (2012). A computational model of the car driver interfaced with a simulation platform for future virtual human-centered design applications: COSMO-SIVIC. *Engineering Applications of Artificial Intelligence*, 25, 1488-1504.

[13] Dosovitskiy, A., Ros, G., Codevilla, F., Lopez, A., & Koltun, V. (2017). CARLA: An open urban driving simulator. In **Proceedings of the 1st Annual Conference on Robot Learning** (pp. 1-16).

[14] CARLA Simulator Team. (n.d.). CARLA documentation (Version 0.9.11). Retrieved from https://carla.readthedocs.io/en/0.9.11/

Funded by the European Union

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Climate, Infrastructure and Environment Executive Agency (CINEA). Neither the European Union nor the granting authority can be held responsible for them.

[15] Fremont, D. J., Vin, E., Kim, E., Dreossi, T., Ghosh, S., Yue, X., Sangiovanni-Vincentelli, A. L., & Seshia, S. A. (2020-2024). Getting Started with Scenic. Scenic Documentation. https://docs.scenic-lang.org/en/latest/quickstart.html

[16] CARLA Simulator. (2023). Scenario runner [GitHub repository]. Retrieved from https://github.com/carla-simulator/scenario_runner

Funded by the European Union

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Climate, Infrastructure and Environment Executive Agency (CINEA). Neither the European Union nor the granting authority can be held responsible for them.

7. ANNEXES

7.1. Annex 1: Bertha Data Model

Link to the Excel with the Data Model: https://docs.google.com/spreadsheets/d/1-xvn7fdD1PJKsJhUgHg2kU8fLJ3_EuIi/edit?usp=sharing&ouid=112506360149916110599&rtpof=true&sd=true

Funded by the European Union

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Climate, Infrastructure and Environment Executive Agency (CINEA). Neither the European Union nor the granting authority can be held responsible for them.