BERTHA

D6.9. Data Management Plan (DMP) (II)

Main author: C. Arias-Perdomo (VED)

Co-authors: H.De Rosario (IBV); S. Oliveira (FI), J. Tarana (FI); S. Pechberti (VED); T. Bellet (UGE), J. Bornard (UGE); G. Kohatsu (CVC), A. Lopez (CVC); S. Govil (DFKI), J. Rambach (DFKI); K. Tovaas (AIT); S. Lacrampe (EUR); M. Bergström (SEYE), S. Finér (SEYE); E. Cervantes (CAP); F. Negin (CON); J. Romero (CID), M. Gomez (CID); D. Mlynarczyk (UV).

Version 1.0

25/04/2025

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

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.

IMPORTANT

This document serves as a template for deliverables and follows a proposal structure. The mandatory sections include: Executive Summary, Introduction and Objectives, and Conclusions. The remaining sections are customizable.

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	CAPGEMINI ENGINEERING DEUTSCHLAND SAS & CO KG	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
15	SMART EYE AKTIEBOLAG	Beneficiary	Sweden

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 several entities from different countries, 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	Data Management Plan
Deliverable n°	D6.9
Deliverable title	D6.9 Data Management Plan (DMP) (II)
Lead beneficiary	VED
Work package and task	WP6 Task 6.4
Document version	1.0
Contractual delivery date	M18
Actual delivery date	M18
Dissemination Level	PU - Public
Purpose	To establish the BERTHA framework for data management and exchange.

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

The Data Management Plan (DMP) delineates a comprehensive strategy for managing the data collected, processed, and generated throughout the project's lifecycle. A concise analysis within the DMP reveals the planned data collection, storage, and security methodologies, ensuring compliance with legal and ethical standards, particularly the General Data Protection Regulation (GDPR). The background of this document is rooted in the need for a framework to handle diverse data types—from raw data and processed information to software code and written documents. These data are integral to developing the Driver Behavioural Model (DBM). The main conclusions highlight the DMP's role in facilitating effective data sharing among project partners and the broader scientific community.

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	27/01/25	First draft.	C. Arias-Perdomo (VED)
V.0.2	05/02/2025	Marker modifications. Update on the introduction.	C. Arias-Perdomo (VED)
v0.3	11/02/2025	Modification in section 1 and 2. Inputs regarding data formats and sizes on table 1 and 2. Inputs regarding permissions access for the HUB.	C. Arias-Perdomo (VED). J.M. Belda, Hélios de Rosario Martínez (IBV); Thierry Bellet, Jean-Charles Bonard (UGE); Kristin Tovaas (AIT); Shreedhar Govil (DFKI), Jason Rambach (DFKI); Sébastien Lacrampe (EUR).
v0.4	26/02/2025	Updates and inputs on table 2. Addition of GitHub into the last section.	C. Arias-Perdomo (VED), Steve Pechberti (VED); Sébastien Lacrampe (EUR); Kristin Tovaas (AIT).
v.0.5	19/03/2025	Updates inputs on table by CVC. Re-check and reformulate section 3. SEYE first inputs. DFKI and CAP inputs in the last section.	C. Arias-Perdomo (VED); Gisele Kohatsu (CVC), Antonio Lopez (CVC); Martin Bergström (SEYE); Shreedhar Govil (DFKI); Eduardo Cervantes (CAP).
v.0.6	09/04/2025	SEYE's final inputs on DMP; AIT inputs on data security; General proofreading of document, modifications on the last section. FI input on data security.	Martin Bergström (SEYE), Svitlana Finér (SEYE); Kristin Tovaas (AIT); C. Arias-Perdomo (VED); Sofia Oliveira (FI), Joana Tarana (FI).
v.0.7	11/04/2025	Check of UGE input regarding data security. Final proof read and reference arrange. Index fix.	T. Bellet (UGE), J. Bornard (UGE); C. Arias-Perdomo (VED).

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.

v.0.8	25/04/2025	Comments addressed after the first review.	C. Arias-Perdomo (VED).
V.1.0	25/04/2025	Ready for submission	C. Arias-Perdomo (VED).

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
HAV	Human Autonomous Vehicle
BBN	Bayesian Belief Network
DMP	Data Management Plan
DPO	Data Protection Office
GDPR	General Data Protection Regulation
FAIR	Findable, Accessible, Interoperable, and Reusable
UC	Use Cases
TBD	To Be Determined
TBC	To Be Confirmed
VIN	Vehicle Identification Number

Disclaimer

1. INTRODUCTION AND OBJECTIVES

1.1. Objectives and scope of the document

The Data Management Plan (DMP) for BERTHA, a HORIZON Europe project [1], outlines the comprehensive strategy for managing the data collected, processed, and generated throughout the project's lifecycle. The DMP ensures that the research data is findable, accessible, interoperable, and reusable (FAIR) [2], which aligns with the Guidelines on FAIR Data Management in Horizon 2020 [3].

This document represents the second version of the Data Management Plan (DMP) of the BERTHA project, funded by the European Union under Grant Agreement n° 101076360. The first version was developed at the project's start, providing an initial framework based on available information. This version, D6.9, builds upon that foundation, integrating updates based on the project's progress, new insights, and refined strategies for data management and data security.

As a living document, the DMP will continue to evolve, with a third and final version planned at the end of the project (30/10/2026). Major revisions will be recorded systematically, ensuring transparency in data governance. The updates across versions are summarized in the following Table 1:

Table 1. Data management reports.

Deliverable	Name	Due date
D6.8	Data Management Plan (DMP) (I)	31/05/2024
D6.9	Data Management Plan (DMP) (II)	30/04/2025
D6.10	Data Management Plan (DMP) (II)	30/10/2026

The core objectives of the DMP on D6.9 remain unchanged, namely:

- Establishing protocols for standardizing metadata publication formats;
- Defining the primary components of the data management strategy applicable across all project datasets;
- Outlining the dataset units BERTHA will generate;
- Data preservation measures will be addressed in the final version of the DMP.

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.

Each project partner has outlined in D6.8 their respective institutions' data security and policy measures. In addition to ensuring that each partner's Data Protection Officer (DPO) is accessible to data subjects, the consortium has appointed Dr. Steve Pechberti from VED as the Data and Ethical Supervisor, overseeing compliance with the EU Charter of Fundamental Rights, GDPR (Regulation (EU) 2016/679), and other ethical obligations.

The DMP will be periodically reviewed and updated to reflect the evolving needs and objectives of BERTHA.

1.2. Structure of the deliverable

The deliverables consist of the following sections:

- Section 1 is the introductory section, which presents the objectives of the DMP on D6.9.
- Section 2 is the data summary of the project.
- Section 3 presents the legal and ethical frameworks of the DMP and how the FAIR principle is implemented.
- Section 4 is about the allocation of resources.
- Section 5 is about data security, storage, and recovery.

Section 6 has the last concluding remarks.

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. DATA SUMMARY

Data management is essential to the BERTHA project, encompassing data collection, storage, and secure utilisation. This second DMP provides an updated overview of the data types collected, and generated, ensuring compliance with FAIR principles and General Data Protection Regulation (GDPR) requirements. The data collected serves multiple purposes, including developing the Driver Behavioural Model (DBM), calibrating its parameters, and validation within autonomous vehicle systems.

The project involves collecting and generating different types of "data" including:

- **Datasets from Work Packages (WP)**: Data is gathered across WPs to support the Driver Behavioural Model (DBM) development, experimental validation, and real-world testing. These datasets originate from simulations, real-world observations, sensor-based recordings, and questionnaire responses collected by partners such as AIT (intended for internal consortium use).
- **Simulation data**: Generated from autonomous vehicles (AV) scenarios in CARLA, and other simulators, serving as a basis for DBM testing.
- **Sensor and Field Data**: Some project activities involve using sensor data from simulations or real-world driving experiments when available.
- **Metadata and Documentation**: Technical and descriptive metadata is produced to ensure data traceability and reusability.

In addition to these core datasets, the project also produces **data outputs** relevant to different stakeholders, including:

- **Project Deliverables**: These will be openly accessible to individuals interested in the project's insights. They include the DBM, the HUB, and the methodology for sharing the model with the scientific community.

- **Communication/dissemination**: We also consider **aggregate data** related to communication and dissemination efforts. This includes website analytics (e.g., number of visits, subscribers, and followers across social media and newsletters) to assess and optimize the effectiveness of communication channels. Additionally, data from event participants (e.g., name, surname, organization, email) is collected during workshop and final event registrations. Only first names and email addresses are stored for newsletter distribution, while contact form submissions (name, surname, organization, email) are used solely for responding to inquiries. All data collection follows GDPR compliance. Further details on dissemination and communication data management will be provided in D6.6 – Report on dissemination, communication activities, and materials (II) (M24) and D6.7 – Report on dissemination, communication activities, and materials (III) (M36).

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.

- **Scientific Publications**: Research results and methodologies will be documented through scientific papers, conference proceedings, and technical reports.

This second version of the DMP provides an updated overview of how this data is collected, structured, and managed within BERTHA. While long-term data preservation strategies will be addressed in the final DMP version, this deliverable focuses on the current data collection and processing activities.

2.1. General overview of existing datasets to be used in BERTHA

The Driver Behavioural Model (DBM) remains BERTHA's core model development effort. It aims to be a scalable, probabilistic representation of human driving behavior structured around the following (see Fig.1):

- Perception [4]
- Risk Awareness [5]
- Decision-Making [6]
- Affective State [7]
- Motor Control [8]

The goal is to make autonomous vehicles safer and more human-like by incorporating this model into their systems. The DBM implementation will be supported by the HUB, allowing for industrial validation and worldwide scalability. This DBM will also be translated into the CARLA simulator [9], an open-source driving simulator for autonomous vehicle research. This model development is central to achieving the objectives of the BERTHA project.

Figure 1. Structure of BERTHA modules. Each module contributes different features to the data model.

Europcar (EUR) provides access to an extensive vehicle fleet dataset to BERTHA task leaders, covering a wide range of vehicle signals, including localization, fuel levels, operational metrics, and some driving-related information such as speed, odometer readings, and trip distances depending on vehicle brand and model.

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 specific needs of the tasks determine the final selection of data used to improve the DBM. Partners such as IBV (on T1.7) and CID (on T2.4) are already reviewing EUR datasets to assess their suitability for DBM validation and use in Field Operational Tests.

Since the dataset includes multiple brands (e.g., Volkswagen, Renault, BMW, Ford, Toyota) and models, and the available information differs across manufacturers, the list of important data fields may vary. Commonly available inputs include:

- Position Data (latitude, longitude, odometer, timestamp)
- Trip Data (start/end time, trip distance, duration)
- Fuel and energy levels
- Maintenance and warning events (e.g., ABS warnings, service reminders)

Thus, EUR's dataset is valuable for increasing the realism and applicability of the DBM by providing access to large-scale, real-world operational data across multiple vehicle types and brands.

The data collected by EUR originates from rental cars, primarily communicated through the vehicle's telematics unit or occasionally via an added Geotab-type unit. Whether renters need to be informed about using their data for the DBM depends on the nature of the data collected. Specific customer notification is not required for metric data, which does not include personal information. However, the data collected can be traced to the customer through the Vehicle Identification Number (VIN). In that case, informing the customer is necessary to ensure compliance with privacy regulations.

The Table 2 below provides an overview of the datasets that BERTHA expects to use for the project:

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 2. List of existing datasets to be used in BERTHA..

Type of data	Description	Origin	Formats	Expected Size	Utility	License
Vehicle telematics data from EUR fleet	Data from EUR's fleet.	EUR via their platform.	Raw in a .CSV format [10]	Around 80000 connected vehicles raise data on every trip [10]	Used in WP2 and WP5 for real-world behaviour modelling and safety assessments.	EUR has contracts with each manufacturer or aggregator to collect this data. Therefore, the data is licensed under the terms specified in these contracts. In addition, data collection is covered by EUR's Terms and Conditions accepted by users at the time of vehicle rental, and all data provided is anonymized to prevent any link to individual drivers.

While this table is comprehensive based on current project needs, the BERTHA project remains open to incorporating additional datasets as the project evolves and if it becomes beneficial to its objectives, as well as to including a year collection range whenever this information becomes available

Based on the telematics dataset described above, EUR has indicated that certain *constructor-specific* information (e.g., detailed vehicle attributes from manufacturers such as Volkswagen, Stellantis, and others) and *aggregator-supplied* data can also be available upon request. While EUR's platform already applies some data harmonization, the aggregator's interface may offer further capabilities for delivering standardized or anonymized fields. This flexibility allows partners to request more specialized or granular data, if required, for specific analyses. The availability of these extended data elements—and the process for obtaining them—will be reviewed in the context of WP2, where the consortium will determine how best to filter and refine the large dataset

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.

for the purposes. The final approach, including any additional agreements for accessing constructor/aggregator-provided data, will be reflected in **the third version of this Data Management Plan**, ensuring proper documentation of licensing terms, anonymization procedures, and data use constraints.

2.2. Overview of research data to be generated in BERTHA

The Table 3 below provides an overview of the datasets that BERTHA expects to generate.

Table 3. List of datasets to be generated in BERTHA.

Type of data	Description	Origin	Formats	Dataset contents	Expected Size	Utility	License
T1.2 Survey Data	Responses from a global survey conducted in 5 countries, focusing on specific driving situations and driver reactions (no personal data).	International survey (WP1)	CSV [10]	See D1.2 [11]	~4,700 respondents (small file size in CSV)	Essential for calibrating Driver Behavior Models (DBM) in WP1 and WP2.	Restricted to consortium use; no formal license applies; no personal data included.
Perception data module	Eye-tracking and in-cabin monitoring to analyze driver attention and visual focus. Gaze data is mapped to CARLA's 3D environment.	Simulations in CARLA during Lab Test by DFKI (WP1, WP2)	CSV; CARLA; MP4; JPG [10]	Gaze orientation, fixations, blink patterns, head-pose; segmentation and scene mapping; ego-camera view (30 Hz); simulator images.	TBD.	Feeds into the Perception Module of DBM for analysing driver attention and awareness.	
Risk awareness data module	Situational awareness and risk estimation assessed from both objective	High Lab Tests implemented on the	CSV [10]	Driving simulator data: time-to-collision (objective risk	TBD.	Supports the development of the Risk Awareness module	

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.

	(e.g., time-to-collision) and subjective measures (e.g., likert scales).	UGE driving simulator (WP1, WP2: see D2.4 [12])		measure) and subjective risk rating (Likert scale).		of the DBM (modelling the drivers' perceived level of risk).
Decision-making data volume	Analysis of drivers' decision making in normal and critical situations, based on the UGE COSMODRIVE model background [13].	High Lab Tests implemented on the UGE driving simulator (WP1, WP2)	CSV [10]	Driving simulator data: driver tactical decisions (e.g., Go/No-Go) and drivers' vehicle command actions.	TBD.	Supports the development of the Decision-Making Module of the DBM, according to the assessed level of risk.
Affective data module	Data on drivers' emotional states, captured via physiological sensors and self-report measures during driving tests.	Simulations in CARLA during Lab Test by IBV (WP1, WP2)	CSV; Parquet; JSON.	Small data samples; large data samples; composite datasets.	TBD.	Integral to the Affective Module, influencing decision-making processes and driving style.
Motor control data module	Data on vehicle control actions by drivers, such as steering acceleration, under various test conditions.	Simulations in CARLA during Lab Test by IBV (WP1, WP2)	CSV; Parquet; JSON.	Small data samples; large data samples; composite datasets.	TBD.	Enhances the Motor Module by detailing the physical interactions with vehicle controls.
Field Operational Test (FOT) data	Participant evaluation data resulting from naturalistic driving studies. This includes everyday	Real-world settings (WP1, WP2)	TBD.	TBD	TBD.	The objective is to replicate and utilise diverse user profiles rather

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.

	driving behaviour and user profiles, capturing various parameters.					than conducting all these tests with the same group of volunteers.
Test Track Data	Data from controlled circuit environments under specific, programmed scenarios. It includes driver responses to predefined situations, safety test results, and regulatory analysis data.	Controlled Circuit Tests (WP5)	TBD.	TBD	TBD.	We assess driver reactions, improve safety features, inform new safety standards, and update certification processes for emerging automated vehicle technologies.
Scenario Generation and Simulation Data (WP4)	Data related to the generation of driving scenarios using SCENIC in CARLA, as well as the deep learning models controlling the vehicles.	SCENIC-based scenarios and AI-driven vehicle models (WP4)	.scenic; .scene; .pth	Use-case scripts; scenario files; deep-learning model weights.	Use Case scripts (~5KB total) Scenario files (~1MB total) Deep learning model weights (~4GB total) Large (~1TB for images and metadata generated from scenario executions)	Supports scenario-based testing and deep learning training for driving models.

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.

Besides T1.2 and EUR datasets, we currently do not have specific datasets available for the modules, as results have yet to be published, and the BERTHA experimental phase will finish by the end of 2025 (with WP4 generating the necessary analysis scripts). However, once the datasets are created—and after obtaining informed participant consent—we plan to license them to promote accessibility and reusability while ensuring data protection and compliance with sharing policies (e.g., GDPR, institutional review board requirements) and respecting privacy and proprietary constraints. Possible licenses could include Creative Commons Attribution (CC BY) or Open Data Commons Open Database License (ODbL), with the final selection to be made in consultation with project partners for future versions of the DMP.

In the case of the DMB, basic functionalities will be made available to the general public via the **MIT license**, and advanced software implementations will be licensed for use.

2.3. Data Storage Solutions and Access Control Mechanisms

Partners provide several data storage solutions to support the BERTHA project's data management needs. This subsection summarizes the storage providers, the technologies adopted, the nature of access (open or under authentication/authorization), and the implemented permissions and access control measures.

All the next points will undergo a detailed discussion in Section 5.

2.3.1. VED — Centralised Data Storage with Hadoop

- **Provider**: VED
- **Technology**: Centralised storage using a Hadoop-based infrastructure at VED, accessed remotely by partners via SSH authentication in its 1st version.
- **Access**:
  - Access is secured via **SSH authentication**.
  - Partners have shared their public SSH keys to enable secure connections.
  - Python scripts upload and download data directly to/from the server.
- **Permissions & Access Control**:
  - Access rights are managed based on predefined user roles.
  - Sensitive data is encrypted during transmission (e.g., TLS/SSL protocols complement SSH where needed).
  - Non-anonymized data storage requires a Data Processing Agreement with VED; anonymized data requires only a declaration of honor.
- **References**: Further public information is detailed in Deliverable D2.3; more technical specifics are available in D2.6 (restricted).

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.

2.3.2. IBV — Temporary Cloud Storage

- **Provider**: IBV
- **Technology**: Temporary secure storage through Nextcloud, SFTP-based solutions; GitLab for development versioning.
- **Access**: Authentication is required; secured access protocols are employed.
- **Permissions & Access Control**:
    - Access is limited to authorized consortium members.
    - Data management and security processes at IBV are certified under ISO/IEC 27001:2013 and the Spanish National Security Framework (Royal Decree 311/2022).

2.3.3. GitHub (IBV Maintained) — Software Code Repository

- **Provider**: IBV (management of GitHub organization)
- **Technology**: GitHub repositories (https://github.com/BERTHA-Project) for collaborative software development.
- **Access**:
    - Depending on sensitivity, private repositories (restricted to project members) or public repositories.
- **Permissions & Access Control**:
    - Sensitive or personal data must **not** be uploaded to public repositories unless fully anonymized.
    - Each partner remains responsible for complying with their institutional and project-level security and confidentiality requirements.

2.3.4. CAP — HUB Platform and Cloud Storage Solutions

- **Provider**: CAP
- **Technology**: The HUB platform, a community cloud-ready solution, facilitates the collaborative use and sharing of road user data and models. Cloud storage services (e.g., Amazon S3) support long-term storage, short-term simulation exchanges, and Human-in-the-Loop scenarios.
- **Access**:
    - Access is restricted through a user registration system (nickname and email) and enforced by secure authentication tokens.
- **Permissions & Access Control**:
    - The HUB applies a user–group–administrator model for permission management, with auditability and GDPR compliance ensured through platform-level security measures.

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. BERTHA FRAMEWORK

3.1. Legal Framework

Under the General Data Protection Regulation (GDPR), personal data encompasses any details connected to an identifiable or identified individual. This includes any means by which a person can be recognized whether directly or indirectly, through identifiers like names, ID numbers, location details, online identifiers, or through characteristics unique to the person's physical, biological, psychological, cultural, or social aspects.

In line with the General Data Protection Regulation (GDPR), the BERTHA project is committed to the responsible and ethical handling of personal data. The project's core output—the DBM—is built from several human-centric data modules. The summary below shows how GDPR principles are applied in practice for the data that needs to be collected for the modules; a complete breakdown is available in WP2 Deliverable D2.4 [12].

The DBM utilizes actual human-driver data, and we are fully aware of the sensitive nature of this information. Every BERTHA data-collection protocol is pre-approved by the relevant institutional ethics committee and Data Protection Officer (see section 3.3): the Ethics Committee of the Universitat Politècnica de València for IBV, the Ethics Committee of Université Gustave Eiffel for UGE, and the DFKI Ethics Board for DFKI. Only the minimum data needed for modeling are captured:

- IBV – driver-facing RGB video and vehicle signals; faces are blurred, and the encrypted files are stored on an access-controlled UPV server.
- UGE – anonymized CAN-bus and environmental data; no personal identifiers are collected.
- DFKI – eye-tracking gaze points; partners receive only anonymized heat maps, while raw eye-tracking videos remain on-site.

For the dataset from the modules, we will apply a common control framework: (i) anonymization and (ii) secure storage with role-based, logged access (see previous section 2.3 for the technical implementation). The detailed workflow for exercising the right to be forgotten—including timelines, confirmation procedures, and propagation of deletions to all replicas—will be discussed among partners and finalized and documented in the next DMP revision.

To ensure compliance and best practices, we also adhere to the FAIR principles and follow each partner's data collection and storage policies. This includes secure storage, access control, and regular audits to ensure ongoing compliance. The level of sensitivity depends on the data source and owner. For more details, check Section 5.3. Any updates to these measures will be reflected in the final version of the DMP (D6.10).

This commitment to data protection and privacy is not just about legal compliance; it is central to our mission of making autonomous vehicles safer and more human-like. By respecting the privacy and rights of individuals, we can build a model that genuinely serves society, a model that is not just about technology but about people.

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.

3.2. FAIR principles in BERTHA

As mentioned, personal data in BERTHA within the GDPR is based on the "FAIR" principles of data management, which imply making research data **Findable, Accessible, Interoperable, and Reusable**. Below is a review of each of these principles for BERTHA.

3.2.1. Making Data and Metadata Findable

Metadata is data about data, in other words, data that provides information about other data or on the research data themselves. Rich metadata helps people to determine whether a dataset is relevant and valuable. As described in the strategy for the management of the intellectual property of BERTHA's proposal, the metadata in publishing format should include the following:

- At least the terms "European Union (EU)" and "Horizon Europe".
- The project title, acronym, and prominently acknowledge grant agreement.
- Publication date.
- Persistent identification.

In addition to the metadata provisions outlined above, the BERTHA project will produce scientific publications as part of its data output. These publications will be assigned a Digital Object Identifier (DOI), a persistent identifier used to identify these digital resources uniquely. The DOI ensures that the publications are easily discoverable and accessible, further enhancing the findability and identifiability of our data.

On the other hand, the automotive industry currently lacks a universally accessible collaborative platform or concept. BERTHA data will be used to develop a working prototype for a decentralised, cloud-scalable platform that offers collaborative workflows as a building block to improve models of traffic participants (drivers, pedestrians, and others). This would be a Data Hub whose structure, facilities, and management comply with FAIR data principles.

The datasets produced in BERTHA include dynamic metadata. Alongside the Use Case (UC) parameters, each experiment contains its own set of common metadata instance consisting of a unique identifier for drivers (UUID). The list below describes the metadata that is provided in the data model and related to the UC parameters:

Identity: Unique for the driver, whose value is a string.
Environment: This is a reference to implementing a Static Data definition. It includes the number of objects in the scenarios.
Value: Integer, Float, String, and others.
Unit: If related, N/A otherwise.
Use Case parameters: speed, distance, and others.

The Dynamic MetaData in BERTHA follows the guidelines outlined in D2.1. While BERTHA does not explicitly reference the Common Evaluation Methodology (CEM), our approach to structuring data aligns with similar principles. We remain open to exploring potential synergies with CEM best practices to improve metadata structuring and evaluation if relevant.

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.

For the name of datasets:

Datasets will be versioned and structured using a name convention consisting of DS_BERTHA_DataController_Description_YYYMMDD_HHMMSS_DataNr.

Where DS stands for the dataset, DataController refers to the short name of the partner responsible for the dataset, Description refers to a brief description of the content of the dataset, YYYMMDD_HHMMSS considers necessary the year, month, day, hour, minutes, and seconds of the data, and DataNr is the version number.

Finally, as stated in D1.6 in terms of keywords for data findability, the project website is actively implementing SEO techniques to improve visibility, including optimizing content for high-performing keywords and enhancing metadata. However, keyword optimization for dataset findability (beyond website searchability) remains an ongoing discussion. The final version of the DMP should provide a more concrete plan, as BERTHA's experimental phase has just begun, and data collection is still in progress.

3.2.2. Making Data Openly Accessible

In alignment with the Open Access Mandate, the BERTHA project is committed to making its research data as accessible as possible while recognizing the need for certain restrictions due to privacy concerns, commercial interests, or security reasons.

Wherever legally and contractually feasible, the partners will make all data and software outputs from the BERTHA project openly available, free of charge, following the principle of "being *as open as possible and only as closed as necessary*."This includes project outcomes, news, and other relevant information, which project partners will regularly update on the project's website. The project also opts for open-source software, primarily **CARLA,** which will be available without restrictions to facilitate the reproducibility of results.

However, **not all project data** can be made openly accessible immediately. A **private GitHub repository** has been established for **internal use within the consortium**, with further details outlined in **Section 5**. For external stakeholders, **data access will be managed via the HUB**, ensuring a **structured and controlled dissemination environment** rather than relying on GitHub.

While the project aims for maximum openness, it also recognizes the need to safeguard legitimate interests or constraints. So, certain parts of the source code or outcomes may need to be kept confidential when necessary, adhering to the principle of being *'as open as possible, but as closed as necessary*.' This approach safeguards legitimate interests or constraints.

The data, associated metadata, documentation, and code will be deposited in various platforms, including the project's website and the Open Research Europe (ORE) [14] platform. The project also aims to support developing and consolidating the European Open Science Cloud (EOSC), facilitating synergies with other projects. All data deposited on the ORE platform are accessible to the public without restriction.

To ensure the data's security and integrity, BERTHA will adhere to secure communication protocols and will utilise SHA-256 [15] for data integrity checks.

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.

Regarding **data access governance**, the consortium has not identified a need for a dedicated Data Access Committee at this stage, and no formal discussions on this topic have taken place. However, access policies are subject to **regular review** as the project evolves.

While the project is active, working datasets will be available only to authorized consortium members through the HUB's role-based permission system; users authenticate via institutional single-sign-on, and all access is logged. In the future, partners will provide comprehensive documentation about any software required to access the data. In the future, the accompanying documentation will also spell out license terms and technical requirements in machine-readable form. With this process, the data remains restricted during development to protect personal or proprietary content, but it becomes openly accessible once those constraints no longer apply.

3.2.3. Making Data Interoperable

While there is no formal decision on adopting **standardized vocabularies** across all datasets, BERTHA remains open to their use. Several structured data categories—**vehicle state, environmental conditions, and driver behavior**—align with **Connected and Automated Mobility (CCAM) Taxonomy elements**. Further, if deemed relevant, interoperability with CCAM terminology may be explored to improve BERTHA's data standardization.
The **unified data approach** ensures that BERTHA's datasets can serve as a foundation for **future projects** requiring standardized data management, extending the impact beyond BERTHA's immediate scope.

For the data and the other research outputs mentioned in Section 1, the project BERTHA will publish documents in their original format and as PDFs if the software required to read the document is not available to everyone, making the data interoperable.

Table 4. Interoperability of BERTHA datasets..

Deliverable	Type	Interoperability	Responsible
D2.1. BERTHA data model	OTHER	It considers the meaning of the information, characteristics allowing representation, and principles of data recording campaigns.	VED
D2.2. BERTHA Data Format and Common Acquisition Principles	REPORT	It is directly related to ensuring that the collected data can be understood and integrated across different systems and platforms, a key aspect of data interoperability.	VED

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.

Deliverable	Type	Interoperability	Responsible
D3.1. HUB Architecture, Data Formats &; Requirements	REPORT	This deliverable now establishes the HUB architecture and data formats while incorporating an extended access permission model. In addition to the classic user–group–administrator framework, a new 'metadata-read' permission has been integrated.	CAP
D3.4. HUB's data governance and policies	REPORT	Adhering to GDPR guidelines and implementing data governance measures can contribute to interoperability by ensuring data exchange and re-use between researchers, institutions, organisations, and countries while adhering to standards and regulations for data formats and privacy.	VOR

The table regarding interoperability aims to detail how the BERTHA project ensures that its various data deliverables and research outputs are accessible and usable across different systems, platforms, and by diverse stakeholders. It specifies the data type, the interoperability measures taken, and the responsible partners, ensuring data can be exchanged and reused efficiently, promoting collaboration and integration across researchers, institutions, organisations, and countries while adhering to standards and regulations.

WP2 will be important in addressing the specific standards, methodologies, and vocabularies for data interoperability in the BERTHA project. Depending on the acquisition and usage circumstances, the definitions of indicators may undergo significant changes in terms of units, frame rates, and referential.

3.2.4. Increase Data Re-use

In line with Horizon Europe's **Guidelines on Open Access to scientific publications and research data** [16], all BERTHA datasets will be released under licences that maximise re-use while respecting intellectual-property rights, confidentiality obligations, and the legitimate interests of the data owners. However, the granting authority can object to transfers or

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.

licensing under certain conditions. Partners will notify the granting authority before any intended transfer or licensing, providing details about the results concerned and the planned exploitation of the results. These processes ensure that data licensing aligns with EU interests and the specifics of the project and data involved.

During the time the DMP is being prepared (from Month 6 to Month 36), decisions will be made on which data will be made open and which will be kept confidential. If an embargo is sought to give time to publish or seek patents, it will be specified in future versions of the DMP, bearing in mind that research data should be made available as soon as possible.

In subsequent versions, the DMP will describe the measures that will be utilised to ensure the long-term preservation of the data, thus ensuring that the data remains reusable for as long as possible.

Finally, in the case of quality assurance: the risk of data quality and quantity was already addressed in the Grant Agreement, with mitigation strategies embedded in WP2, WP4, and WP1. Datasets from demonstrator are supplemented by those obtained from prior partner projects to increase robustness.

- Specific data quality assurance processes will be detailed in D2.7, ensuring that all datasets follow rigorous collection methods.
- Additionally, each WP actively upholds quality standards, ensuring data meets the required reliability benchmarks before use.

3.3. Ethical aspects

Since the assessment will encompass user studies, securing ethical approval is necessary. Each partner is tasked with obtaining its ethical approval via the appropriate methods, notably from any (national or local) ethics committee or other bodies such as data protection authorities. This section will outline each partner's process.

- IBV has submitted its application to the ethical committee at the Universidad de Politécnica de Valencia, where the IBV is based. The application includes comprehensive details about the measurements they intend to make. While the endorsement to handle the ethical aspects of the project is anticipated, the status is that the application is under review and updates are awaited.
- VED will secure approval from the VED legal department regarding data privacy, adherence to GDPR, and all study documents.
- CVC will secure the endorsement of the ethical committee at the Universidad Autónoma de Barcelona.
- In the case of FI, no approval from the ethics committee is necessary. All approvals or validations of FI data processing are made by the Privacy and Information Security Committee. Ethics committee approval would only be needed if the data to be processed could involve any risk to the company's values or ethics, which is not the case with BERTHA's WP6-related data for the communication and dissemination activities.
- DFKI will secure approval for user studies from the Ethics Board at DFKI as well as the Data Protection Officer (DPO).

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.

- AIT will secure approval from the AIT Data Protection & Information Security department regarding the contents of the T1.2 survey, as well as the related data that will be collected from respondents and analysed.
- In the case of CID, the Data Protection Officer will be in charge of the anonymization of the data, in such a way that the personal data of individuals participating in various trials will be used strictly confidentially and published solely as aggregated statistics (anonymously) and under no circumstances as individual data sets. Considering this type of anonymized data to be registered and European GDPR regulations, there is no need to report the data registration to the Spanish Data Protection Agency. In any case, CID has a Data Protection Officer (DPO) at the Company who will collaborate with the personnel involved in the project, ensuring compliance with data protection at all times.
- UGE has secured approval of UGE experiments from the UGE ethical committee as well as the Data Protection Officer regarding data privacy and adherence to GDPR. CON does not need approval from the ethics committee since it would only be required if the data to be processed could involve any risk to the company's values or ethics, which is not the case with BERTHA's WP6-related data for communication and dissemination activities. However, it's important to mention that all legal sanctions inflicted on any partner responsible for sharing data should not reverberate on Continental or other partners if they already have used this data. The procedure in this case from Continental side is to ensure that we immediately stop using the mentioned data while being out of any further legal sanctions.
- CAP will secure approval from the Data Protection Officer and the legal department regarding data privacy, and also in adherence to GDPR, and all study documents.
- EUR will secure approval from the Group legal department and DPO regarding data privacy, adherence to GDPR, and all study documents.
- SEYE will secure approval from the Data Protection Officer regarding data privacy and adherence to GDPR.

In line with the principles of **Research Integrity and Transparency**, BERTHA is committed to conducting research transparently, fairly, and unbiasedly. We ensure that all data-sharing processes are transparent and communicated to all participants. This includes providing comprehensive information on how their data will be used, shared, and preserved. Specific methods for achieving transparency include detailed consent forms and accessible data management plans. These methods are detailed further in Work Packages 2 and 6.

Furthermore, we uphold the principle of **Respect for Participants and Society**. We treat all research participants respectfully and ensure that their data is handled carefully. We consider our research and data-sharing practices' societal, ecological, and cultural implications. This extends to our approach to data sharing and long-term preservation, where we prioritise the rights and interests of our participants and the broader society. Our commitment to these principles ensures that our research is robust, reliable, ethical, and respectful.

BERTHA's approach includes implementing secure data storage solutions. These practices are elaborated in WP2.

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. ALLOCATION OF RESOURCES

4.1. Estimations of Costs for Making the Data FAIR

Data management activities are limited to project management costs and will be covered by allocated resources in the project budget. Any additional costs for open access or early publishing will be paid from the project's allocated budget.

4.2. Responsible for the Data Management

The BERTHA consortium has agreed that **each institution will provide the contact details of their respective Data Protection Officers (DPOs)** to all data subjects involved in their internal research. This ensures that data protection is prioritised at the institutional level. Additionally, VED has appointed Dr. Steve Pechberti as the **Data and Ethical Supervisor** for the consortium. With this, BERTHA will establish all data requirements and the steps necessary to ensure GDPR compliance based on the various types of data collected.

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.

5. DATA SECURITY

Ensuring the security and integrity of data across multiple work packages and partner contributions is paramount in the BERTHA project. This chapter outlines the provisions for the secure storage, transfer, and long-term preservation of sensitive data.

The data storage and security of the data has to consider two stages:

- The data shared among partners. Covered by the Exchange Infrastructure and Temporary Storage.
- The data collected and stored by each partner.

On the rest of the Section, we discuss in more detail each of these points.

5.1. Data Security and Exchange Infrastructure

The BERTHA project employs a hybrid data storage solution led by Vedecom (VED), involving centralised data storage and decentralised systems maintained by various partners. This approach allows for flexible yet secure data sharing and accessibility across the project's spectrum, from WP1 to WP5.

An important component of this infrastructure is the establishment of secure gateways that facilitate access to data across different platforms, enhancing the efficiency of model realisation and validation tests for CCAMs. A homogeneous solution is pursued wherever possible to minimise the need for extensive interfacing between the BERTHA system and the data.

Sensitive data will be transmitted using encryption protocols (e.g., TLS/SSL) to prevent unauthorised access during transfer. A preliminary prototype of this system (D2.6) will be developed early in the project lifecycle, with subsequent enhancements detailed in D2.7. This phased approach ensures that the data exchange system evolves in line with project needs and security standards.

The specific architecture for HUB-hosted versus partner-hosted storage is still under discussion with CAP and VOR; the agreed solution will be documented in the next DMP revision. User access control is essential, involving robust policies to protect data against unauthorised access, granting access based on predefined rules and roles, and securing the HUB's capabilities. In addition, the HUB repository now incorporates an access permission model that builds upon the classic user–group–administrator framework with the introduction of a new 'metadata-read' permission. As explained on D3.1 [17], this permission enables users to view file descriptions—such as metadata about models—without granting full access to the underlying sensitive content, thereby allowing controlled usage (for example, in simulation or autopilot scenarios). Differentiated permissions are defined so that consortium members receive standardized access while external users can be granted tailored rights. This approach not only strengthens data protection and GDPR compliance but also promotes interoperability and ethical data sharing throughout the project lifecycle (for more reference, see D3.1 [17]).

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.

5.1.1. Contractual and Security Compliance

As detailed in Section 2.3.1, partners storing data on the VED infrastructure must comply with contractual and security requirements. Specifically, storing non-anonymized data requires a Data Processing Agreement with VED, whereas anonymized data only necessitates a declaration of honor.

5.1.2. Use of Third-Party Code Repository (GitHub)

As described in detail in Section 2.3.3, BERTHA uses a GitHub organization maintained by IBV (https://github.com/BERTHA-Project) to facilitate collaborative software development. Partners must adhere to guidelines and responsibilities regarding version control, access control, and compliance with institutional security, confidentiality, and data protection requirements outlined previously.

5.2. Temporary Data Storage Solution

In addition to the long-term solutions provided by VED, the Institute of Biomechanics of Valencia (IBV) offers a temporary storage solution during the initial phases of the project. This solution includes:

- **Secure Storage**: Data is hosted on secure servers at IBV's Valencia, Spain facilities. Access to this data is safeguarded through advanced security protocols, the specifics of which (e.g., Nextcloud, SFTP) will be finalised based on consortium requirements.
- **Development Workspace**: A secure workspace for software development is provided, featuring version control via Git, accessible through Gitlab and SFTP interfaces.
- **Certifications**: IBV's data management and security practices are certified under ISO/IEC 27001:2013 and the National Security Framework (Royal Decree 311/2022), ensuring compliance with high information security management standards.

Finally, BERTHA recognizes the need for long-term preservation. Therefore, we plan to:

- Identify certified repositories for data curation and preservation.
- Ensure compliance with best practices for data archiving.
- Document metadata and contextual information for future reference.

We acknowledge the potential value of integrating our preservation strategy with initiatives such as GAIA-X or other European data spaces, and we plan to explore this connection explicitly in the next version of BERTHA's DMP.

5.3. Data collected and stored by each partner

BERTHA collects data from individuals. Therefore, the data is initially linked to the contact details of the individuals and should be considered personal data. The partners that collect the data apply different policies about the contact information of the individuals:

- No collecting the data. Some partners will recruit participants through an external company.

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.

- Delete. Some partners will delete the contact details at the end of the project.
- Keep. Some partners keep the contact information because they maintain a database of individuals fulfilling the GDPR and national laws.

In any case, the data shared between partners will be pseudonymized or anonymized.

Table 5. Policies for deleting contact information.

PARTNER	Policy
IBV	Keep
VED	Delete
CVC	No data collecting
FI GROUP	Keep
DFKI	Keep
AIT	Delete
CIDAUT	Delete
UGE	Delete
CON	Keep
CAP	No data collecting
EUR	Delete
UV	Delete
SEYE	Keep

In the next subsection, we provide detailed partner-specific information regarding data management, security, and compliance. While Section 2 outlines overall project strategies and data types, Section 5 addresses individual partner responsibilities, reflecting institutional differences in data management and privacy policies.

5.3.1. Instituto de Biomecánica (IBV)

IBV's data management and security practices are certified under ISO/IEC 27001:2013 and the National Security Framework (Royal Decree 311/2022), ensuring compliance with high information security management standards.

IBV will keep the data it generates and processes in its storage system: the central institutional servers in IBV's facilities (Valencia). The servers are protected by a firewall and with restricted access permissions to project folders, where only the system administrators and the Principal Investigator of the project can grant access to read, write, and modify them. Weekly backup copies are only accessible by system administrators, and monthly additional backup copies are located in a separate building within IBV's private network in the Polytechnic University of Valencia. IBV's employees are forced to follow a security policy with strict rules for accessing the information systems to ensure security and data protection.

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.

IBV keeps people's contact information in a password-protected database that complies with Spanish law and GDPR. The contact information is only used by the "User Location Area" of IBV to recruit participants. Researchers do not have access to the contact information. Researchers will have access to pseudonymized data only.

5.3.2. Vedecom (VED)

Access to the data is restricted to the researchers working directly on the project. All the data is anonymous, and it cannot be traced back to individuals.

5.3.3. Universidad de Valencia (UV)

UV will work in the processing of data collected by other partners for the purposes of the BERTHA project, following the security and confidentiality protocols established by those responsible for each data set.

5.3.4. Computer Vision Center (CVC-CERCA)

CVC will code and execute simulation scenarios (based on the defined Use Cases) in the CARLA simulator, either for training or testing purposes, using DBMs or not. During the execution of such scenarios, simulated data may be collected locally. In the short term, access to this data is restricted to BERTHA partners. After reaching associated publications, the plan is to make this data publicly available, considering the best option given the data size (a significant number of terabytes is expected).

CVC has applied the appropriate technical and organisational measures to guarantee the confidentiality and integrity of the data and can demonstrate that the processing is following current legislation and that the protection of the rights of the interested parties is guaranteed.

To this end, the CVC has implemented a Personal Data Protection Management System to determine and apply the technical and organisational means at its disposal to prevent loss, misuse, alteration, unauthorised access and theft of the data provided, without prejudice to inform that the Internet security measures are not impenetrable.

Likewise, the CVC has implemented internal controls in order to regularly verify, evaluate and assess the effectiveness of the technical and organisational measures implemented to guarantee the security of the processing.

5.3.5. Fundación CIDAUT (CID)

CIDAUT applies a procedure to separate private data from test results data, so the data are anonymized. Thus, there will not be a direct relationship between personal data and data recorded during the BERTHA project. In any case, all personal data to which the Cidaut Foundation has access are treated following EU general data protection regulation (GDPR) procedures. In addition, the Data Protection Officer will coordinate the actions between

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.

Cidaut and the Spanish Data Protection Agency to solve any problem related to personal data attending to this regulation.

5.3.6. Deutsches Forschungszentrum für Künstliche Intelligenz (DFKI)

DFKI handles sensitive user data in the BERTHA project with strict security and compliance measures. The collected data includes eye images, ego-centric views, and gaze fixations and driving controls along with a questionnaire of driving characteristics. The data collection process has been approved by DFKI's Data Protection Officer and adheres to GDPR requirements. To ensure privacy protection:

- Retention & Anonymization: Raw and sensitive user data will be securely stored for up to 10 years, after which it will be permanently deleted. Only an anonymized dataset will be retained for further research.
- Storage & Access Control: Data is stored on DFKI's Augmented Vision Kaiserslautern server, protected by a firewall, and can only be accessed via the internal network using authorized credentials. Access is strictly limited to BERTHA project employees and system administrators.

DFKI remains committed to data security and privacy by implementing industry-standard safeguards and ensuring compliance with all applicable regulations.

5.3.7. Smart Eye (SEYE)

SEYE handles sensitive user data in the BERTHA project with strict security and compliance measures. The collected data includes face images, driver head- and eye-movement, driving controls along with a questionnaire of driving characteristics, some demographics related information relevant for the project reporting (for example, age, gender).

Storage & Access Control: Data is stored on SEYE's server, protected by a firewall, and can only be accessed via the internal network using authorized credentials.

Retention & Anonymization: In the event of sharing data with project partners, personal data (such as name, address, phone number or e-mail) will be anonymized or pseudonymized before sharing.

5.3.8. FI Group (FI)

The personal data that may be collected directly from the person concerned will be treated confidentially and will be included in the processing activities of the BERTHA project. The data controller is F-INICIATIVAS CONSULTADORIA E GESTAO, Unipessoal LDA (FI Group). DPD contact details: privacy@fi-group.com

The processing of data is carried out to comply with the legal obligations of the entity and the fulfilment of contractual relations or relations related to the BERTHA project. The legal basis for the processing of personal data collected through the contact forms on the website or the sending of e-mails is the consent of the data subject.

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 a rule, personal data will only be communicated to partners who are part of the BERTHA project. To request the exercise of any of the rights held by the interested parties established by the regulations, you can send an e-mail to the following address and your request will be processed by the FI Group team.privacy@fi-group.com

Types of personal data and processing:

Table 6. FI GROUP's specific types of personal data and processing.

Types of personal data and processing	
Collective	Every person who fulfils a form on the BERTHA project website.
Purpose of processing	Communication of developments in relation to the BERTHA project, subscription to the newsletter or any informative newsletter.
Category of data	Identification data: identification number (ID card, ID number or passport), position, name and surname. Contact details: telephone, e-mail and address.
Target	Marketing department and the team that carries out any other function within the project.
Period of retention	The time necessary to process the corresponding request, after which they are duly blocked for the period of time in which legal obligations may arise. In the case of subscription to newsletters, the necessary data is retained until the person concerned expresses his/her wish not to be included in the communication lists or withdraws his/her consent.
Source	Completion of the forms on the website itself.
Data retention	The personal data provided shall be kept for the time necessary to fulfil the purpose for which they are processed. They will also be kept for the period of time in which legal obligations may arise for FI GROUP. The periods established in the applicable regulations for the retention of files and documentation apply. As soon as the processing of personal data is no longer necessary, the data will be deleted or appropriately blocked.

5.3.9. Austrian Institute of Technology (AIT)

The data collected from the global survey, conducted in five countries, focuses on specific driving situations and driver reactions. It does not contain personal or sensitive data. Within AIT, access to the data is restricted to the internal project team, which consists of seven members. As the data is not personal in nature, no anonymization or pseudonymization measures are necessary before sharing it with project partners. The dataset remains restricted

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.

to consortium use in accordance with the project's data management and confidentiality policies.

5.3.10. Université Gustave Eiffel (UGE)

UGE handles participants' data involved in the experiment implemented on the UGE driving simulator with strict security and compliance measures. The collected data includes traffic situation parameters, driver's action on vehicle commands (pedals, steering wheel, indicators, horn), eye tracking and subjective data collected from questionnaires. The data collection process has been approved by the ethical committee of the University Gustave Eiffel, and adheres to GDPR requirements. Moreover, only anonymized data collected during the UGE experiment will be shared with the other partners involved in BERTHA, in full compliance with the privacy guidelines approved by the Ethical Committee of the UGE. As no sensitive personal data will be shared, there is no need for pseudonymization.

5.3.11. Continental (CON)

Access to the data in Continental will be restricted to the core project team members exclusively. This includes the Principal Investigators and designated system administrators. Data collected for validation tests will be recorded internally and will not be shared externally.

5.3.12. Capgemini (CAP)

Capgemini does not intend to store information from recordings when such data involves sensitive content. The data processed through the hub pipeline is entirely simulated and does not contain any sensitive information.

Additionally, as part of the hub's registration process, only a name and email address are required. Users will be presented with a consent box outlining the terms for data usage in compliance with applicable GDPR regulations. Users must accept these terms to access the hub; otherwise, access will be restricted.

5.3.13. Europcar Mobility Group (EUR)

Data access will be granted exclusively to the core project team and the P&T connectivity department to ensure data security and maintain confidentiality.

Vehicle Identification Numbers (VINs), which could potentially link data to individual customers, will be systematically excluded from the dataset prior to any processing or analysis. This measure ensures that the insights derived from the data pertain to trip patterns and connectivity performance without revealing any personally identifiable information. Furthermore, the processed, anonymized dataset will be securely transferred to Cidaut for the purposes of our collaborative project. Following this transfer, the original raw data will be securely purged from our systems within a 15-day retention period, adhering to our data governance protocols and minimizing data footprint.

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.

5.4. Summary of Partner Data Management Practices

This section provides a structured overview summarizing key data management aspects for each partner within the BERTHA project on Table 7.

Table 7. Data collected and stored by each partner summary.

Partner	Data Type	Anonymization	Storage	Access Control	Compliance
IBV	Personal (contact), research data	Pseudonymized	Institutional servers with firewall, regular backups	Restricted to system admins, PI approval	ISO/IEC 27001:2013, GDPR, National Security Framework (Spain)
VED	Research data	Anonymized	Centralized Hadoop storage	Restricted to project researchers	GDPR
UV	Processed research data (provided by partners)	Depending on source	Internal system per source specifications	Following guidelines provided by source partners	GDPR
CVC	Simulation-generated data	Anonymized	Local storage, future public access planned	Initially restricted to BERTHA partners, later public	GDPR, internal data protection management
CID	Test data, separate from personal data	Anonymized	Internal separation protocols	Restricted, controlled internally	GDPR, coordinated with Spanish Data Protection Agency
DFKI	Eye-tracking, driving controls, questionnaires	Anonymized after retention period	Augmented Vision server, firewall protected	Strictly project team, system admin	GDPR, internal data protection officer oversight
SEYE	Images, eye movement, demographics	Anonymized/Pseudonymized for sharing	Internal servers, firewall protected	Authorized project employees	GDPR
FI	Identification, contact data (web forms)	Not anonymized internally, shared as needed	Internal database, password-protected	Limited to marketing/project-related teams	GDPR, Portuguese law compliance

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.

Partner	Data Type	Anonymization	Storage	Access Control	Compliance
AIT	Survey data, non-personal	Not required (no personal data collected)	Internal restricted storage	Internal project team only	GDPR compliance measures
UGE	Simulator data (traffic, actions, eye tracking)	Fully anonymized before sharing	Internal secure storage	Project researchers, anonymized sharing	GDPR, Ethical Committee of UGE
CON	Validation test data	Restricted, internal use only	Internal systems, secure storage	Principal Investigators, admins only	GDPR, internal confidentiality protocols
CAP	Simulated data, minimal personal data for registration	Anonymized/simulated, consented registration	HUB platform secure storage	Consent-based, controlled access	GDPR
EUR	Vehicle/trip data excluding VIN	Fully anonymized before sharing	Internal secure temporary storage	Core team and connectivity department	GDPR, internal data governance

This standardized summary facilitates quick understanding and comparison of data handling practices across BERTHA project partners, demonstrating compliance with relevant legislation and best practices in data management.

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. CONCLUSIONS

This document describes the main principles of the BERTHA project for the Data Management Plan (DMP) after the first 18 months of the project. BERTHA's DMP provides a foundational framework ensuring data management aligns with the project's goals and adheres to the necessary standards and regulations.

The DMP release is a formal commitment by partners to adhere to the data management strategies it defines. As a living document, it will update and expand as the project evolves and new information on data collection, transfer, and handling arise. This report shall be considered the first release of the DMP for the data in the project. A revised version will be prepared towards the end of the project.

The last release should detail data licences, interoperability, and others. By continuously refining the DMP, we aim to address emerging challenges and incorporate best practices, ensuring robust data management throughout the project's lifecycle.

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. REFERENCES

[1] European Commission. (2024). Horizon Europe: Funding programmes and open calls. Retrieved May 7, 2024, from https://research-and-innovation.ec.europa.eu/funding/funding-opportunities/funding-programmes-and-open-calls/horizon-europe_en

[2] Stall, S., Yarmey, L., Cutcher-Gershenfeld, J., Hanson, B., Lehnert, K., Nosek, B., Parsons, M., Robinson, E., & Wyborn, L. (2019). Make scientific data FAIR. *Nature*, 570, 27-29. https://doi.org/10.1038/d41586-019-01720-7

[3] European Commission. (n.d.). Guidelines on FAIR Data Management in Horizon 2020. Retrieved May 7, 2024, from https://ec.europa.eu/research/participants/docs/h2020-funding-guide/cross-cutting-issues/open-access-data-management/data-management_en.htm

[4] 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.

[5] 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

[6] 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.

[7] 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).

[8] 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.

[9] Dosovitskiy, A., Ros, G., Codevilla, F., Lopez, A., & Koltun, V. (2017, October). CARLA: An open urban driving simulator. *In Conference on robot learning* (pp. 1-16). PMLR.

[10] D.C. Arias-Perdomo, S. Pechberti, J. M. Belda, H. de Rosario Martínez, T. Bellet, J.-C. Bonard, et al., "D2.2 Data Format and Common Acquisition Principles," *BERTHA European Project*, version 1.0, Sep. 17, 2024, Available: https://berthaproject.eu/wp-content/uploads/2025/02/BERTHA_D2.2_WP2_M10.pdf

[11] M. Zach, C. Rudloff, and K. Tovaas, "Deliverable 1.2: Stratification of population according to driving styles," BERTHA European Project, version 1.0, Dec. 10, 2024. Available: https://berthaproject.eu/wp-content/uploads/2025/02/BERTHA_D1.2_WP1_M12.pdf

[12] C. Perdomo, S. Pechberti, U. Martínez-Iranzo, J.-M. Belda, V. de Nalda, T. Bellet, J.-C. Bonard, S. Govil, J. Rambach, F. Negin, M. Bergström, S. Finér, S. Lacrampe, K. Tovaas, and M. Zach, "D2.4 Updated Methodology for Advanced Technology Acquisition," BERTHA European Project, version 1.0, Mar. 20, 2025, [Unpublished deliverable].

[13] T. Bellet, J.-C. Bornard, P. Mayenobe, and G. Saint Pierre, "A computational model for car drivers situation awareness simulation: Cosmodrive," in Proc. First Int. Symp. Digital Human Modeling, Lyon, France, Jun. 14–16, 2011, pp. 1–8.

[14] Open Research Europe. (n.d.). Retrieved May 7, 2024, from https://open-research-europe.ec.europa.eu/

[15] Gupta, S., Goyal, N., Aggarwal, K.: A review of comparative study of md5 and ssh security algorithm. *Int. J. Comput. Appl.* 104(14) (2014)

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.

[16] European Commission, Directorate-General for Research & Innovation. (2017). Guidelines to the rules on open access to scientific publications and open access to research data in *Horizon 2020 (Version 3.2) [PDF]. H2020 Programme*. Retrieved May 7, 2024, from https://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-pilot-guide_en.pdf

[17] S. Pretscher, E. Cervantes Duenas, and F. Baptista Alves, "D3.1: HUB Architecture, Data Formats & Requirements," *BERTHA European Project*, version 1.1, Sep. 24, 2024. [Unpublished deliverable].

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.