

AutoDL – Secure and Privacy-Aware Data Lake for Vehicle Data Storage and Processing

1. Overview

Secure and Privacy-Aware Data Lake for Vehicle Data Storage and Processing (AutoDL) is a joint RD&I initiative by LISHA, mobway, Bosch, Renault and Stellantis, and promoted by *Line VI Rota 2030 Vehicular Connectivity*.

The project aims at developing a Data Lake Platform based on LISHA's [IIoT Platform](#), along the platforms developed by LISHA [IASE](#) and [SDAV](#), to support efficient handling of structured and non-structured data from the domain of Automotive Big Data. The Automotive data will be communicated to the platform using Machine-to-Machine (M2M) communication protocols within the paradigm of the Industrial Internet of Things (IIoT) over 5G, envisioning Security and Privacy. The project includes two main applications of Artificial Intelligence and data analysis in the scope of Automotive Big Data, namely, Ecologic Driving Profiling and Predictive Maintenance.

1.1. AutoDL Consortium



The **Software/Hardware Integration Lab (LISHA)** was founded in 1985 to promote research in the frontiers between hardware and software. Since then, it has dedicated considerable efforts to research in areas such as computer architecture, operating systems, computer networks and the related applications. Currently, the laboratory focuses on innovative techniques and tools to support the development of embedded systems.



mobway is a Brazilian startup that maintains a vehicle data platform connected to automakers in order to standardize access to such data, offering owners the possibility of connecting vehicles to products, with LGPD compliance, using a single data standard and free from the informality of telemetry.



BOSCH

Grupo Bosch is a global leader in technology and services for the Mobility, Industrial Technology, Consumer Goods and Energy and Building Technology sectors. As a leading IoT company, Bosch provides innovative solutions for smart homes, Industry 4.0 and connected mobility. The company strives for mobility that is sustainable, safe and fascinating and uses its expertise in sensors, software and services, as well as its own IoT cloud to offer its connected consumers multiple solutions from a single source.



Renault do Brasil is one of the largest vehicle manufacturers in the Brazil, with a vehicle factory in the city of São José dos Pinhais - PR since 1998. This vehicle plant is one of the most modern in Latin America and, in 2020, was recognized by the World Economic Forum as a benchmark in Industry 4.0.



Stellantis is a constellation of 14 iconic automotive brands. Stellantis aim to develop, engineer, manufacture, and scale the best breakthroughs in all facets of sustainable mobility from autonomous, connected, electrified, shared and pre-owned vehicles to micro-mobility, commercial vehicles, and even electric aircraft.

1.2. AutoDL Architecture

An overview of the AutoDL architecture is depicted in Figure 1 and illustrates the major entities:

- **Connected Vehicles** that communicate with an IoT Platform to send telemetric information about their operating status (e.g. ECUs and other CAN-connected components, non-CAN components) and also sensing data (e.g. GPS, IMU, LiDAR, Cameras) that might be of interest to fleet managers.

- **5G M2M services such as URLLC** as the connectivity technology.
- **SmartData** to provide secure, georeferenced and timed communication.
- **IoT Platform** to implement the microservices used by the vehicles to send data and by the fleet managers to securely store, retrieve, and process such data.
- **Analytics** algorithms to extract useful management and operation information from the acquired dataset, including visualization.
- **Machine Learning Models** built on the acquired data to support predictive maintenance and ecologic driving profile.

1.3. A Review of the Solution Proposed for Data Lake in AutoDL Project

The AutoDL project aims to develop a **secure automotive big data infrastructure in a data lake format**, which is capable of aggregating and processing large volumes of data from various sources related to vehicles. This infrastructure will be an essential enabler for creating new business for project partners, allowing the application of artificial intelligence algorithms and data analysis tools to properly organized, cataloged and structured vehicle information.

The focus of the project is on two application scenarios in cooperation with partner companies: the development of a tool to **identify the ecological driving profile of drivers** and the **creation of algorithms for predictive vehicle maintenance**. Both applications use semi-structured and unstructured data, requiring data anonymization to ensure driver privacy and efficient correlation of unstructured data with vehicular data.

The proposed infrastructure will use data lakes to deal with the diversity of data, meeting the performance requirements generated by the acquisition, storage and analysis of vehicle data in both application scenarios. Cyber-security will be a priority, implementing strict measures to protect user data and ensure the privacy of collected information. The use of secure M2M communication, especially in the context of 5G technologies with the URLLC protocol, and IEEE 802.11p to forward data collected from vehicles to the cloud, will ensure the security and privacy of owners. Secure connections to all vehicle-related data sources, including automaker systems, will be established for secure integration with the data lake.

UFSC's Software and Hardware Integration Laboratory (LISHA) will play a crucial role in the formation of automotive big data, aggregating various sources into the data lake and developing essential components for data processing and cyber-security. The participation of the companies that make up the project arrangement will enable everything from generation to consumption of data, with the provision of sources for information and the use of infrastructure in the listed use cases.

1.4. Project Scope at Rota 2030

The ultimate goal of the project is to raise the Technology Readiness Level (TRL) from 3 to 6, allowing the implementation of a working prototype ready for testing and deployment in a limited production environment. Project evaluation will occur through iterations with project partners, ensuring performance validation and compliance with security and privacy requirements. Meeting these objectives will enable the creation of a robust automotive data infrastructure, driving innovation, safety and new product development in the automotive industry.

This project fits into the thematic area “Data Privacy and Security Technology” and thematic lines “Protection of data related to vehicle connectivity”, “Privacy of user and vehicle data” and “Detection and prevention of security attacks in the environment vehicle”, as it will ensure that the entire data transfer process, from the vehicle's communication system to the user and the generation of intelligence, is secure, preserves the privacy of the information and is done with the consent of the vehicle owner.

The project is also has relation with the thematic area “Vehicle Connectivity with the External

Environment” and thematic lines “IoT applied to vehicular connectivity”, “New smart city services resulting from the use of vehicular connectivity” and “New services based on vehicle georeferenced information ”, as it aims to develop a data lake that will be an essential enabler for creating new revenues and new businesses for project partners. Together with secure M2M communication, especially in the context of 5G technologies to forward data collected from vehicles to the cloud, ensuring the security and privacy of owners.

2. Related Projects

- [SmartData on Wheels - a Safe and Secure Runtime Support System for Autonomous Vehicles](#)
- [Auto5G — Intelligent Vehicle Telemetry and Supervision System](#)

3. Technical Documentation

- [EPOS](#)
- [IoT Platform](#)
- [IoT Platform Internals](#)
- [SmartData Series Semantics](#)

4. Publications

<https://lisha.ufsc.br/pub/index.php?key=AutoDL>