IoT4CPS – Trustworthy IoT for CPS

FFG - ICT of the Future
Project No. 863129

Deliverable D5.2
Product Lifecycle Data Management (PLCDM) Stakeholder Perspectives

The IoT4CPS Consortium:
AIT – Austrian Institute of Technology GmbH
AVL – AVL List GmbH
DUK – Donau-Universität Krems
IFAT – Infineon Technologies Austria AG
JKU – JK Universität Linz / Institute for Pervasive Computing
JR – Joanneum Research Forschungsgesellschaft mbH
NOKIA – Nokia Solutions and Networks Österreich GmbH
NXP – NXP Semiconductors Austria GmbH
SBA – SBA Research GmbH
SRFG – Salzburg Research Forschungsgesellschaft
SCCH – Software Competence Center Hagenberg GmbH
SAGÖ – Siemens AG Österreich
TTTech – TTTech Computertechnik AG
IAIK – TU Graz / Institute for Applied Information Processing and Communications
ITI – TU Graz / Institute for Technical Informatics
TUW – TU Wien / Institute of Computer Engineering
XNET – X-Net Services GmbH

For more information on this document or the IoT4CPS project, please contact:
Mario Drobics, AIT Austrian Institute of Technology, mario.drobics@ait.ac.at

Document Control

Title: Product Lifecycle Data Management (PLCDM) Stakeholder Perspectives
Type: public
Editor(s): Violeta Damjanovic-Behrendt
E-mail: violeta.damjanovic@salzburgresearch.at
Author(s): Violeta Damjanovic-Behrendt (SRFG), Christos Thomos (IFAT)
Doc ID: D5.2

Amendment History

Version	Date	Author	Description/Comments
V0.1	01.04.2019	Violeta Damjanovic-Behrendt	Document organization
V0.2	04.05.2019	Violeta Damjanovic-Behrendt	Finalizing Section 2
V0.3	20.05.2019	Violeta Damjanovic-Behrendt	Public datasets – overview
V0.4	25.05.2019	Christos Thomos	Added contributions, e.g. Figures and descriptions
V0.5	10.06.2019	Violeta Damjanovic-Behrendt	Finalizing overview on public datasets
V0.6	20.06.2019	Violeta Damjanovic-Behrendt	Overview of related ontologies
V0.7	22.06.2019	Violeta Damjanovic-Behrendt	Finalizing of the report
V0.8	29.06.2019	Violeta Damjanovic-Behrendt	Adding conclusion remarks; template
V0.9	30.06.2019	Violeta Damjanovic-Behrendt	The report sent out for the quality assurance (QA)
V0.10	02.07.2019	Silvio Stern (X-NET)	QA
V1.0	08.07.2019	Violeta Damjanovic-Behrendt	Final version of D5.2

Legal Notices

The information in this document is subject to change without notice.

The Members of the IoT4CPS Consortium make no warranty of any kind with regard to this document, including, but not limited to, the implied warranties of merchantability and fitness for a particular purpose. The Members of the IoT4CPS Consortium shall not be held liable for errors contained herein or direct, indirect, special, incidental or consequential damages in connection with the furnishing, performance, or use of this material.

The IoT4CPS project is partially funded by the "ICT of the Future" Program of the FFG and the BMVIT.

Federal Ministry
Republic of Austria
Transport, Innovation
and Technology

FFG
Forschung wirkt.

Content

List of Figures

List of Tables

Abbreviations

AGV	Automatic Guided Vehicle
CAD	Computer Aided Design
CAM	Computer Aided Manufacturing
CC	Cyclomatic Complexity
CDT	Custom Datatypes
CIM	Computer Integrated Manufacturing
CPS	Cyber Physical System
DFM	Design For Manufacturing
DSRC	Dedicated Short-Range Communications
ECLIF	Extended Common Logic Interchange Format
FFT	Fast Fourier Transform
GDPR	General Data Protection Regulation
IIoT	Industrial Internet of Things
IoT	Internet of Things
ITS	Intelligent Transport Systems
KBS	Knowledge Based Systems
LiDAR	Light Detection and Ranging
LOC	Lines of Code
ML	Machine Learning
MSDL	Manufacturing Service Description Language
MSO	Manufacturing Systems Ontology
NASAMDP	NASA Metric Data Program
NIS	Network and Information Systems
OSLC	Open Services Lifecycle Collaboration
OWL	Web Ontology Language
PLCDM	Product LifeCycle Data Models
PROV-O	Provenance Ontology
PSO	Production Systems Ontology
RAMI	Reference Architecture
SAREF	Smart Appliances REFerence Ontology
SIR	Software-artifact Infrastructure Repository
SKOS	Simple Knowledge Organization System
SLAM	Simultaneous Localization and Mapping
SSNO	Semantic Sensor Networks Ontology
UGV	Unmanned Ground Vehicles
VSSo	A Vehicle Signal and Attribute Ontology

Executive Summary

This report captures multi-tenancy aspects related to connected cars and emerging standards for data and information exchange in the Automotive industry. The report creates a ground for the research in WP5 and "digital twinning" of the real-world situations and processes related to various lifecycle stages in the Automotive Driving sector. The automotive companies typically do not publicise data and hence, in this report, we look at relevant open-source systems and public datasets, in order to construct own data repository to enable the required data analyses for the Digital Twin demonstrator. The presented data model will be further enriched through task T5.4 "Identity, Security and Safety in Product Lifecycle Data Management", to additionally address identity, security, privacy and safety aspects in the Automotive industry. In addition, the IoT4CPS data model will be semantically enriched by using standard ontologies and some newly proposed ontologies for formal description of car signals and sensors, e.g. VSSo (Vehicle Signal and Attribute Ontology) (Klotz et al., 2018). Such approach will contribute to the cross-interaction of the connected car system with the external stakeholders in the cloud.

1. Introduction

The predecessor report **D5.1 "Data Models for the IIoT and Industry 4.0, and in the Automotive Sector"** explores the current state of technology progress that impacts Product LifeCycle Data Models (PLCDMs) and methods for data management in the Industrial Internet of Things (IIoT). More concretely, the D5.1 report overviews a set of the most popular standards and data models with the potential to further enhance ongoing developments in the two sectors central to IoT4CPS: Industry 4.0 and Automotive Driving. The identified standards and data models serve as a basis for the IoT4CPS data model design, which underlines the Digital Twin demonstrator to be implemented in task T5.5 of IoT4CPS.

The report D5.1 also presents a preliminary scenario created to identify potential assets in the automotive sector. Based on that, several preliminary lifecycle data models corresponding to e.g. design, operation and maintenance phases of the identified assets are defined (see Appendix 4 and Appendix 5 of D5.1). The report D5.1 is published in month M06 of the project, and shortly afterwards, in month M09, the business needs of the Automotive Driving applications have been defined in the report **D2.2 "Business needs consolidation – competitive intelligence"**. Hence, starting from D5.2 and in the remaining reports of WP5, we align our models and methods for the design and implementation of Digital Twin demonstrators towards the scope of business needs presented in D2.2.

The focus of this task **T5.2 "Data Model for Multi Stakeholder Lifecycle Data Management"** is to capture multi-tenancy aspects related to connected cars and actual legislations and emerging standards for data and information exchange in the Automotive industry, along the entire lifecycle data management. In addition, the data model created in this task needs to consider the roles of various stakeholders included through automotive lifecycle phases, from production of connected cars to their end-of-life phase. In task T5.4 "Identity, Security and Safety in Product Lifecycle Data Management", the same data model will be further extended to address identity, security and safety aspects, along multi-tenancy and the lifecycle data management in the Automotive industry.

In order to create a ground for the research in WP5 and for "digital twinning" of the real-world situations and processes related to various lifecycle phases in the Automotive Driving sector, there is a natural strong request for reasonable quality data to be available in the project, in order to support the data analyses. The automotive sector companies typically do not publicise data (neither production nor driving (operational) data) and hence, in IoT4CPS, we look at relevant open-source systems and public datasets, in order to either reuse available datasets or construct own data repository based on partially relevant public datasets.

The report is organized as follows: Section 1 describes our motivation to address both multi-stakeholder (multi-tenant) and Product LifeCycle Data Management (PLCDM) perspectives when constructing the data model that fits the scope and challenges of IoT4CPS. Here, we also consider BMVIT's "Austrian Action Programme on Automated Mobility" (see BMVIT (2019)) that defines several guiding principles related to Automated Mobility for the ongoing period 2019 – 2022:

- safety;
- technological bases and infrastructure to support the future mobility;
- building trust throughout the entire product lifecycle; and
- impact assessment and access to data (emphasizing clearly a distinction between research and commercial data).

Section 2 describes a multi-stakeholder perspective applied to Automotive Driving use cases, which are further aligned to the Device.CONNECT™ business case presented in the report D2.2 "Business needs consolidation – competitive intelligence". Section 3 describes our methodology for data management strategy in IoT4CPS. Section 4 brings several relevant public datasets for Machine Learning (ML), and in Section 5, we look for relevant ontological data representation models to support PDLCM relevant for the Automotive Driving. Section 6 designs the data models that fit PDLCM and multi stakeholder perspectives in IoT4CPS. Section 7 concludes this report and presents the next steps, e.g. extending the created datasets to enable automated security and safety measures and defect predictions (see D5.4.1 for the details).

1.1 Motivation

Our motivation in IoT4CPS is to research some of the major challenges and business cases for future mobility to take place in Austria. Here, our particular interests relate to cybersecurity and safety conditions of the Automotive Driving, observed through the entire product lifecycle, plus evaluation and impact assessment through the specifically designed Digital Twin demonstrator (task T5.5). New mobility technologies, in particular Automotive Driving, Smart Cars, Connected Cars, Intelligent Transport Systems (ITS) become commercially viable and, in order to allow for new mobility experiences to happen, these technologies call for many open issues to be solved at various levels, including the levels of governments and businesses. Such technologies are expected on the road by 2020, offering improved road safety, reduced congestion and emissions, and increased accessibility to personal mobility.

Some major challenges to be addressed in the Automotive industry to enable vehicle connectivity and Automated Mobility to happen, are summarized as follows (FNC-2018, 2018):

- Holistic infrastructure development policies: In the future, each element of the Smart Automotive infrastructure will transmit data about traffic, weather, road works information, etc. to the cloud, and will pro-actively participate in decision making (e.g. driving and navigational decisions) segmented at the level of various stakeholders (e.g. roads (including highways, crossings, parking areas), buildings (and bus stops), public vehicles like busses and trams, cyclists, pedestrians, manufacturers and car sharing dealers), and many more). For example, to enable safe navigation, roads need to transmit traffic, weather and road works information to the cloud.
- International standards for vehicle communication: International bodies such as the UNECE and ITU are working with global automotive players towards a harmonization of communication standards for vehicle connectivity (i.e., 802.11p, 5G, DSRC).
- Faster and wider network connectivity: Major service providers are expected to provide mainstream 5G services by 2020.
- Cybersecurity: In both the Industry 4.0 and the Automotive Driving sectors, each point of communication can be seen as a possible vulnerability point for cyberattacks. In January 2017, the UNECE initiated a task force to address cybersecurity issues relevant to the Automotive industry. In December 2016, ENISA published a report summarizing good practices and recommendations related to cybersecurity and resilience of smart cars (ENISA, 2016).
- Improved data sharing and data transparency: In both the Industry 4.0 and the Automotive sectors, data must be accessible to all stakeholders in the mobility ecosystem to ensure continuous safety conditions. One of the key promises of these two sectors is to support automated decision making based on data addressing the entire lifecycle phases, from smart car production to its disposal.

Our motivation in WP5 is to address the last two above-mentioned challenges: (i) cybersecurity aspects in the Automotive Sector (task T5.4 and T5.5), and (ii) improved data sharing and data transparency provided for various stakeholders in the sector (tasks T5.2 and T5.3). Therefore, WP5 designs and implements the Digital Twin demonstrator (task T5.5), featuring cybersecurity aspects and threats of Automotive Driving use cases.

1.2 Regulations and Initiatives for the Automated Mobility in Austria

The most prominent international governmental bodies, regulations and initiatives at present, that strongly influence the Automotive Driving sector can be summarized as follows:

- eCall system (from the 1 of May 2018, all new cars sold in the EU are equipped with the eCall alarm systems that automatically calls the emergency services and, in case of an accident, send the location of the car);
- GDPR (from the 25 of May 2018, GDPR regulatory requirements related to privacy data protection within the EU, entered into force),

- the Directive on Security of Network and Information Systems (NIS) (from the 09 of May 2018, the NIS Directive affects search engines, cloud providers and online marketplaces, and sets cybersecurity regulations, incident response procedures, etc.),
- the new UNECE's Working Party on Automated/Autonomous and Connected Vehicles (GRVA),
- the IEEE 802.11p, International Standard for Wireless Access in Vehicular Environments,
- the Dedicated Short-Range Communications (DSRC), a wireless communication technology that enables vehicles to communicate with each other and other road users directly, without involving cellular or other infrastructure,
- 5G connectivity promises to reach peak speed 20 times faster in comparison with current 4G. It would enable an autonomous car to know of traffic obstructions kilometres away, based on data sharing from other vehicles and via the cloud. It would enable safe driving by reducing, and will possibly eliminate traffic fatalities in the future,
- and more.

In June 2016, the Austrian Federal Ministry of Transport, Innovation and Technology (BMVIT) published the first Austrian roadmap on Automotive Driving called "Automated – Connected – Mobile. Action Plan Automotive Driving – Executive Summary" (BMVIT (2016)). This roadmap sets priorities and steps for the future use of automated vehicles and mobility services in Austria. It sets the guidelines to ensure safe, efficient and environmentally sound mobility, towards strengthening of the Austrian economy for the period between 2016 – 2018, including:

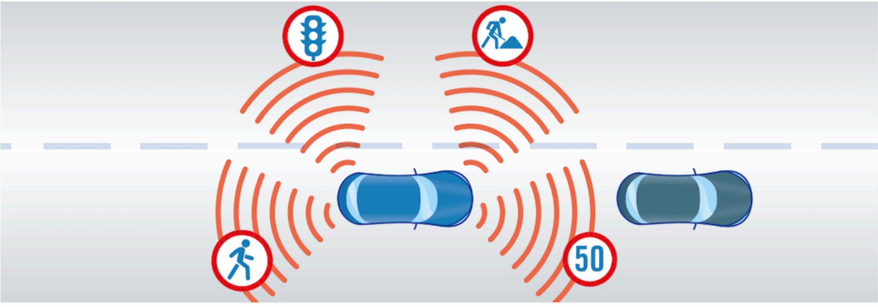
- the continuous adaptation of the legal framework in Austria, in close coordination with international legislation,
- the expansion of the digital infrastructure,
- the expansion to various modes of transport,
- stronger international networking and collaboration, and
- the integration of electrical and environmentally efficient drive systems.

The BMVIT's latest publication called "Austrian Action Programme on Automated Mobility" (BMVIT (2019)) defines next steps and guiding principles for Automated Mobility in Austria, for the currently ongoing period 2019 – 2022. In this roadmap, BMVIT (2019) identifies the following key trends: the electrification of powertrains, the increase of automation, and shared mobility and multimodality in the context of publicly accessible mobility. The basic guiding principles related to Automated Mobility are summarized as:

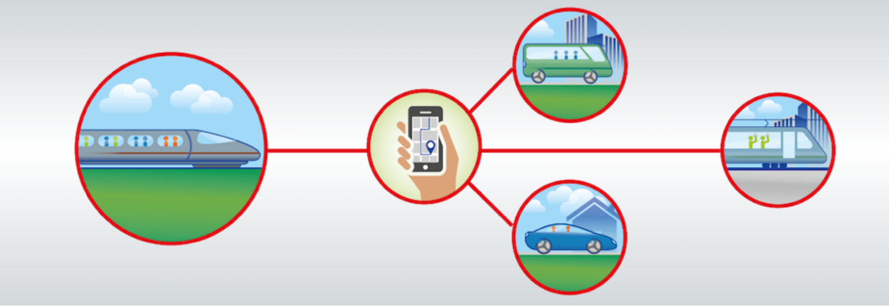
- safety,
- technological bases and infrastructure to support the future mobility,
- building trust throughout the entire product lifecycle, and
- impact assessment and access to data.

Finally, the major use cases of interest for the coming years in Austria include the following three cases (BMVIT (2016), BMVIT (2019)):

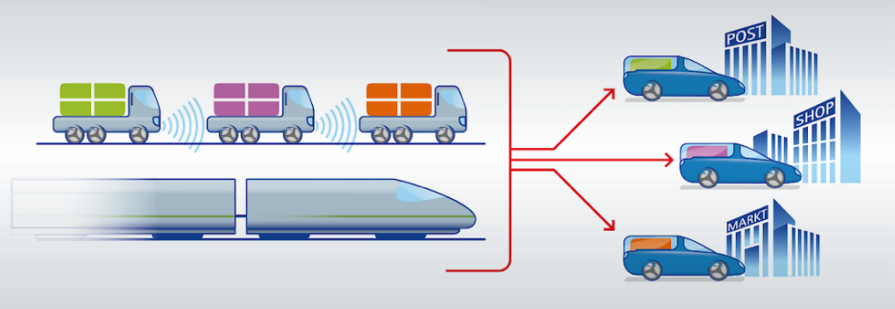
- "Safety+ through an all-round view": This use case is about driver assistance systems that use predictive sensors to intervene in traffic situations whenever danger is imminent. The information from other road users and from the infrastructure itself is used too. This enhances road safety in the immediate environment of the vehicle.



- “New flexibility”: This use case is about automated vehicles that offer new, on-demand services that can increase the flexibility of mobility users (e.g. route optimization, driving times tailored to personal preferences, secure and convenient connection mobility with intermodal transfer points, booking services, etc.) and ease the burden on the environment (decreasing the environmental impact).



- “Well supplied”: This use case is about the automated freight transport, optimized feeder services with efficient long-distance transports and suitable concepts for the “last mile”.



The first two of the three above-presented use cases are taken as starting points for the design of the Digital Twin demonstrator in IoT4CPS, next to the selected business case from the report D2.2 (use case with the Device.CONNECT™ system).

In the following section, we firstly discuss the twofold perspective of WP5, and secondly, we adapt the two selected use cases to serve the IoT4CPS’ business needs as discussed in the report D2.2.

2. Multi-Stakeholder and Cybersecurity Focus in the Automotive Driving in IoT4CPS

The WP5 perspective in IoT4CPS is twofold: it is about both multi-stakeholders and cybersecurity needs in the Automotive Sector, along the PLCDM stages.

The Automotive Driving applications are designed to assist the owners of smart, connected cars in a variety of ways, from the enhancement of the driver's user experience (e.g. lane changing, parking assistance, night vision, traffic sign and traffic light recognition, map navigation support, etc.) to reducing driver's distraction and improving the overall safety on the roads (ENISA, 2016). This kind of applications is beneficial not only to the smart car's owners, but also to other stakeholders linked to the connected cars in various ways, e.g. passengers sharing the car, smart cities, smart roads, car sharing dealers, etc. However, the Automotive Driving applications are based on cloud technologies, and are therefore, exposed to a vast attack surface in which every asset is a potential target that can compromise its security, safety and privacy. Hence, we additionally look at the design of cybersecurity, safety and privacy capabilities for the Digital Twin demonstrator.

Figure 1 illustrates the IoT4CPS major concepts: product lifecycle (or PLCDM; in green), security aspects (in blue); trustworthy connectivity (in orange), and Digital twin demonstrator (in pink). In relation to other tasks and WPs, the concepts are covered through the following:

1. Cybersecurity Lifecycle (joint work through WP3, WP4, WP5);
2. Digital Twin modelling (WP5);
3. Trustworthy connectivity (WP7);
4. Traceability through lifecycle phases (WP7);
5. Security by isolation (WP7);
6. Smart production use case (Device.CONNECT™) (WP2, WP7);
7. Autonomous vehicles (WP6).

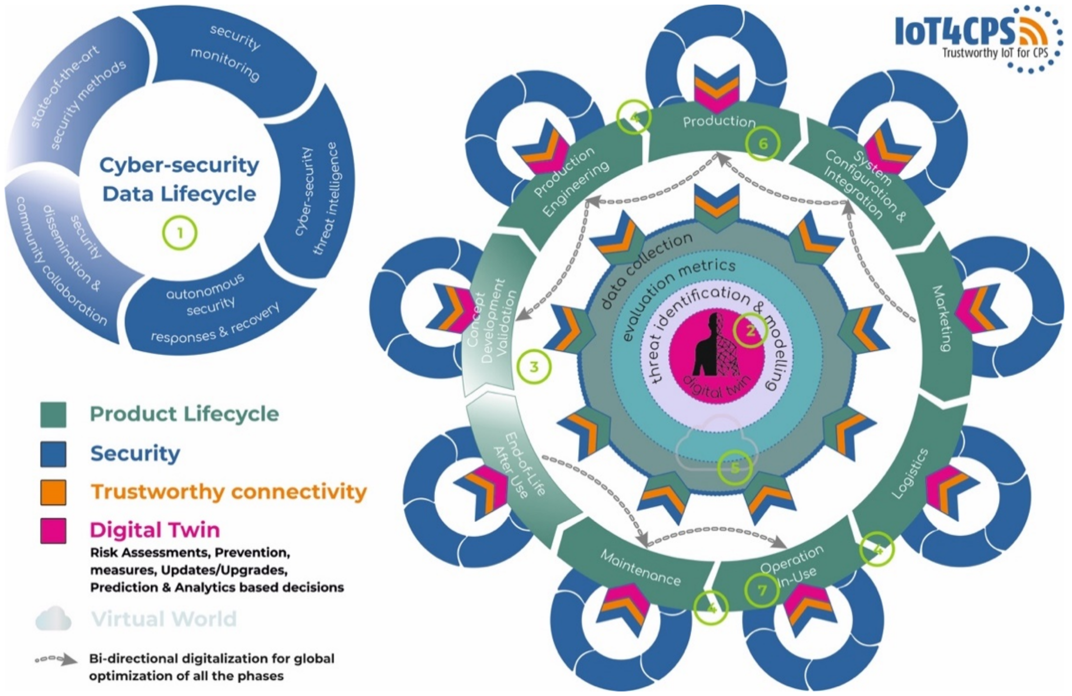

security monitoring
cyber-security threat intelligence
autonomous security responses & recovery
community collaboration & security dissemination
state-of-the-art security methods
Cyber-security Data Lifecycle
1
Product Lifecycle
Security
Trustworthy connectivity
Digital Twin
Risk Assessments, Prevention, measures, Updates/Upgrades, Prediction & Analytics based decisions
Virtual World
Bi-directional digitalization for global optimization of all the phases
IoT4CPS
Trustworthy IoT for CPS
Production
6
Production Engineering
4
System Configuration & Integration
Marketing
Logistics
4
Operation In-Use
7
Maintenance
4
End-of-Life After Use
Concept Development Validation
3
data collection
evaluation metrics
threat identification & modelling
digital twin
2
5

Figure 1 – The major concepts in IoT4CPS, in relation to the definition of the Digital Twin data models

Figure 2 shows the conceptual view of the IoT4CPS use cases to be designed for the Automotive Driving applications. One of the very first steps in the design of the Digital Twin demonstrator is to identify assets that are involved and functional in the use cases. Hence, we set the authorization boundaries for the system to be designed; note this is a simplification of the connected car system that include a variety of sensors and their CPSs (illustrated as subsystem 1 to subsystem n, where each of the subsystems can have many subsystem's elements).

The connected car has its lifecycle stages, e.g. initiation phase (design, engineering and production), operational phase (driving and interacting), maintenance and end-of-life (decommission) phase. The connected car system is furthermore linked to various external systems, which could be considered either from the multi-stakeholder perspective (stakeholders with various roles and interests to exploit the system) or from cybersecurity, safety and privacy perspective (each asset could be vulnerable and expose the whole system to various threats).

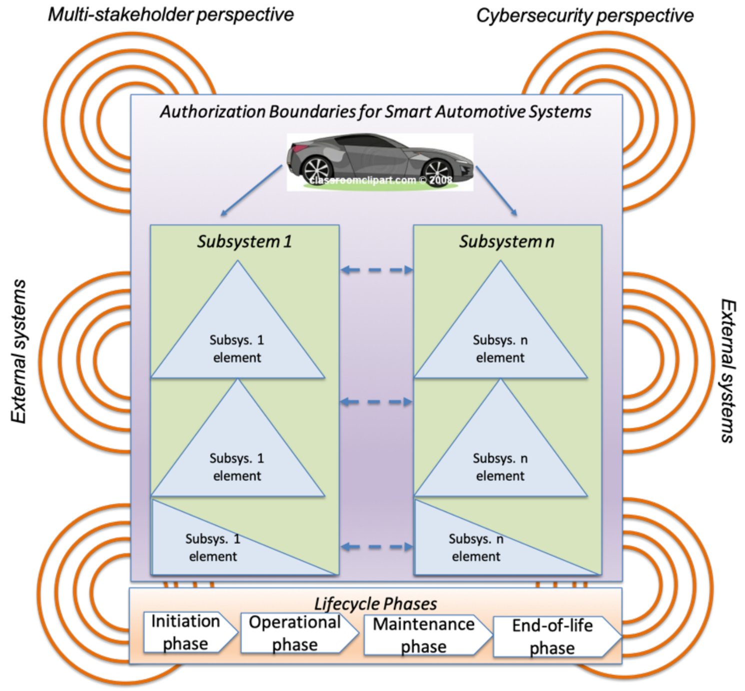


Figure 2 – A conceptual view on IoT4CPS use cases for the Automotive Driving applications

2.1 Relation to D2.2 Business Case on Security Verification Along the Lifecycle

The IoT4CPS D2.2 “Business needs consolidation – competitive intelligence”, Section 3.2 “AVL: Security Verification Along the Full Life Cycle of IoT-based Industrial Instrumentation Systems” reports on the role of the Device.CONNECT™ system, which enables communication links with the external systems through, e.g. smart/ predictive maintenance services in the cloud. It provides connectivity to a multitude of cloud-based commercial products, such as emission analysers, particle samplers, instrumentation systems, etc. At the same time, the cloud-based nature of the Device.CONNECT™ puts this device into a category of highly vulnerable and critical assets that need to be continuously monitored and checked against common threat intelligence indicators, regulatory compliance obligations and stakeholder’s governance rules, in order to effectively responds to both cyber incidents and regulatory challenges.

In the rest of this section, we adapt the AVL’s business case (the Device.CONNECT™ system) in the context of two selected use cases, previously discussed in Section 1.2 (c.f. BMVIT (2019)). In this way, we enrich our use cases by combining cases defined in IoT4CPS D2.2 and BMVIT (2019). This allows us to widen the range of major CPS-based assets and stakeholders of interest in WP5, and to better “frame” the Digital Twin demonstrator and its data model for the creation of datasets, which are necessary to support PLCDM. The datasets providing historical and real-time data for each of the lifecycle phase of CPSs (as shown in 2) are crucial for our research in WP5.

We “redefine" the two use cases, which are identified as being of a major interest in Austria (c.f. BMVIT (2019)) for the purpose of IoT4CPS. Practically speaking, we engage the business case from the report D2.2 into the two selected major use cases discussed in BMVIT (2019), by addressing both multi-stakeholder and cybersecurity perspectives to widen the complexity of the real-world Automotive Driving applications to be “digitally twinned” in IoT4CPS.

- We "redefine" the case "Safety+ through an all-round view" (see Section 1.2) to include (i) the Device.CONNECT™ for the data acquisition and management and (ii) its Digital Twin counterpart (demonstrator) that will be implemented in IoT4CPS for security and safety evaluations related to the Automotive Driving applications and verifications.
- Furthermore, we "redefine" the case "New flexibility" by adding the Digital Twin demonstrator (as a counterpart of the Device.CONNECT™) to enable assistive intelligence capabilities.

2.2 Combined Use Case 1: Safety & Cybersecurity+ through the Lifecycle Stages

The use case "Safety+ through an all-round view" (see BMVIT (2019)) is about the driver assistance system that uses information from various stakeholders, e.g. road users, smart city traffic regulations, passengers, and from the infrastructure itself.

The goal of this use case is to enhance road safety in the environment of the vehicle.

Use case description. In IoT4CPS, the sensor data are collected using the Device.CONNECT™ system, that gathers data related to the road and environmental conditions, e.g. air pollution, temperature near the surface of the road, humidity. This data can be combined with the data from the car's powertrain and chasses controls. For example, the powertrain controls receive sensor information from electrical engines, transmission, wheels. Chasses control receive sensor information related to both the car's frame and car's environment, including the steering and brakes, airbags, embedded cameras, real-view mirrors, windshield wipers (ENISA, 2016).

Note that the sensor data collected through the Device.CONNECT™, or through electrical engines, wheels or chasses controls, are all generated during the operational (driving) phase of PLCDM (see Figure 3). These data are neither about the initiation phase (design, engineering, production) nor maintenance of the connected car nor its end-of-life stage. The creation of such missing datasets in IoT4CPS will need to be managed by using either the most probable value to fill in the missing value, or using the attribute mean for all samples belonging to the same class as the given tuple.

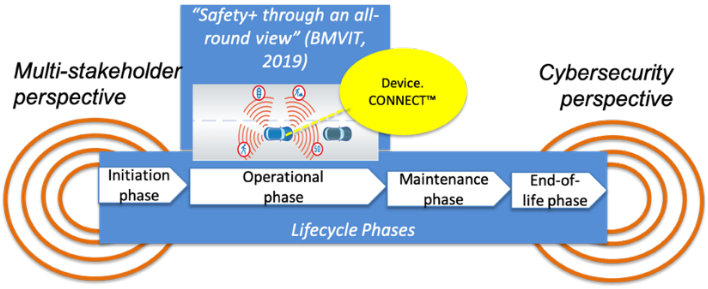

"Safety+ through an all-round view" (BMVIT, 2019)
Multi-stakeholder perspective
Device. CONNECT™
Cybersecurity perspective
Initiation phase
Operational phase
Maintenance phase
End-of-life phase
Lifecycle Phases

Figure 3 – The enrichment of the "Safety+ Through an All-Round View" use case in IoT4CPS

Identification of assets. Table 1 identifies various types of assets involved in the above illustrated use case 1.

Table 1: Assets involved in the use case 1

Connected car's device/ sensor/ CPS	Type of sensor data
Initiation phase	
CAD uploader	Computer Aided Design (CAD). It assures that the design of the CPS-based product is analysed, optimized and sent for further manufacturing
Collaborative analysis checker	It enables collaborative design and further improvements of CPS-based products to be manufactured

CAM/CIM initiator	Computer Aided Manufacturing (CAM) / Computer Integrated Manufacturing (CIM). It enables the manufacturing flow from raw materials to finished products, with quality assurance and automated assembly.
Robotic assembly checker	It checks the production of completed assemblies, part size, part defects (e.g. based on feeder jam data)
Supply chain status control	It checks for the delivery terms in order to meet the demand
Operational phase	
Device.CONNECT™	Air pollution, temperature near the surface of the road, humidity
Powertrain control	Data from electrical engines, transmission data, wheels data
Chasses control	Data about the steering and brakes conditions, airbags, embedded cameras, real-view mirrors, windshield wipers
Maintenance phase	
CIM Remote Monitoring Service	It monitors for unauthorized access and changes to the files and products (connected cars)
Integrity Monitoring Service	It detects and reports changes made in files or detects manipulations
End-of-life phase	
Privacy Data Monitoring Service	It ensures that retained privacy data are removed from the connected cars (FPF, 2018), e.g. mobile apps that are used, mobile apps log-in data, location, the driver's daily route, phone contacts and address books, garage door codes, various digital content, subscription services, wi-fi hotspots, data services, etc.
Other Data Monitoring Services	It enables other data to be removed from the connected cars, e.g. on-board diagnostic information.

Identification of stakeholders. Sharing data amongst various stakeholders can open numerous privacy issues leading to reputational damage for the users, car manufacturers, suppliers, garages, network service providers, software and application providers, etc. Table 2 identifies the potential stakeholders that are (directly or indirectly) involved in the use case 1.

Table 2: Stakeholders involved in the use case 1 (analysis partly based on (ENISA, 2016))

Stakeholder	Description of the stakeholder's role in the use case
Initiation phase: Manufacturers & Suppliers	
Manufacturer	Provides the production and assembly of the car components.
Aftermarket Supplier	Provides components with additional features, e.g. media player.
Operational phase: Car Users & Internal Services	
Driver	Drives and uses the connected car's gadgets and apps/services. Connects via smartphone. Uses external cloud applications.
Passengers	Use gadgets and apps or are exposed to apps and services running on other user's devices.
Powertrain control services	Transmission controls; wheels controls; services for monitoring of the engine features, etc.
Operational phase: External Services	
Road services	Monitoring road and traffic conditions; Safety recommendations and contextual insights, e.g. speed limit changes, roadway conditions. eCall services.
Testing and certification services	Monitoring driving habits; Contextual insights.
Insurance services	Pay-How-You-Drive insurance plan.
Network connectivity providers & services	Network access and services.

Smart cities & services	Economical use of the road infrastructure. Smart city weather station and road speed controls. Environmental impact evaluation.
Maintenance phase: External Services	
Insurance services	Pay-How-You-Drive insurance plan.
Road services	Monitoring traffic conditions; Safety recommendations.
Manufacturer	Evaluation of part's functionality and safety
End-of-life phase: External Services	
Smart city services	Economical use of the city infrastructure. Environmental impact evaluation.

2.3 Combined Use Case 2: Assistive Intelligence+ through the Lifecycle Stages

The use case "New flexibility" (see BMVIT (2019)) is about automated vehicles offering on-demand services that can increase the flexibility of mobility users (e.g. route optimization, driving times tailored to personal preferences, secure and convenient connection mobility with intermodal transfer points, booking services, etc.) and ease the burden on the environment (decreasing the environmental impact).
The goal of this use case is to improve the mobility of users and the positive environmental impact.

Use case description. In IoT4CPS, the Device.CONNECT™ system collects data such as air pollution, temperature near the surface of the road, humidity. The other telematics data from the connected cars refers to control data streams affecting critical functionality of the vehicle, e.g. braking, engine performance, collision detection and emergency calling, vehicle diagnostics, vehicle speed, GPS data. The infotainment data refers to non-critical systems, e.g. music and video streaming, Bluetooth connectivity, wi-fi connectivity and wi-fi hotspots, SMS texting, etc. The power of data lies in its combination. For example, the connected car can recognize the intention of another car to change lanes, based on data related to the car's speed and position adjustment. Based on light signals, it can know which connected car will turn and which will continue straight. This type of scenarios shows a potential to eliminate traffic fatalities in the future. Another case of possible privacy misuse of combined data obtained from the connected car's infotainment and other CPSs, is presented in (Damjanovic-Behrendt, 2018).

In order to support assistive intelligence capabilities, we "redefine" the use case on "New flexibility" (see BMVIT, 2019) by adding the Device.CONNECT™ system and the Digital Twin demonstrator to automatically process data in a way that answers the needs of various stakeholders involved in the lifecycle stages and verifies the system overall safety and cybersecurity conditions (see Figure 4).

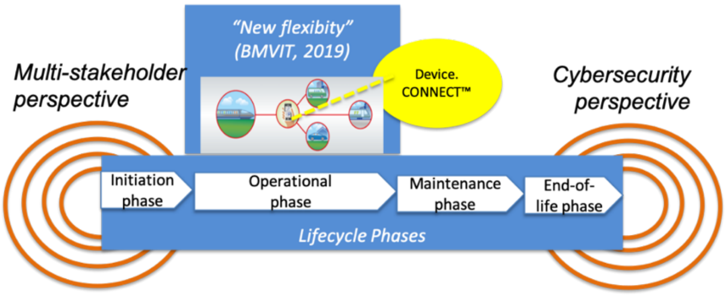


Figure 4 – The enrichment of the "New Flexibity+" use case in IoT4CPS

Identification of assets. Table 3 identifies various assets involved in the use case 2.

Table 3: Assets involved in the use case 2

Connected car's device/sensor/ CPS	Type of sensor data
Initiation phase	
Robotic assembly checker	It checks the production of completed assemblies, part size, part defects (e.g. based on feeder jam data)
Supply chain status control	It checks for the delivery terms in order to meet the demand
Operational phase	
Device.CONNECT™	Air pollution, temperature near the surface of the road, humidity
Powertrain control	Data from electrical engines, transmission data, wheels data
Chasses control	Data about the steering and brakes conditions, airbags, embedded cameras, real-view mirrors, windshield wipers
Infotainment control	Music and video streaming, Bluetooth connectivity, wi-fi connectivity and wi-fi hotspots, SMS texting …
Maintenance phase	
Integrity Monitoring Service	It detects and reports changes made in files or detects manipulations
End-of-life phase	
Privacy Data Monitoring Service	It ensures that retained privacy data are removed from the connected cars (FPF, 2018).
Non-Privacy Data Monitoring Services	It enables other data to be removed from the connected cars, e.g. on-board diagnostic information.

Identification of stakeholders. Table 4 identifies the stakeholders involved in the use case 2.

Table 4: Stakeholders involved in the use case 2 (analysis partly based on (ENISA, 2016))

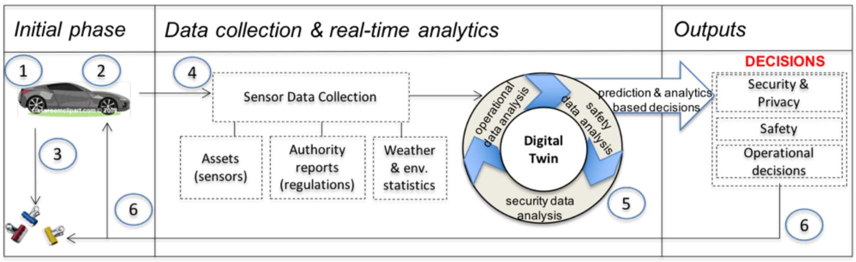
Stakeholder	Description of the stakeholder's role in the use case
Initiation phase: Manufacturers & Suppliers	
Supplier	Provides car components and /or operating system for connecting car components.
Aftermarket Supplier	Provides components with additional features, e.g. media player.
Operational phase: Car Users & Internal Services	
Driver	Drives and uses the connected car's gadgets and apps/services. Connects via smartphone. Uses external cloud applications.
Passengers	Use gadgets and apps or are exposed to apps and services running on other user's devices.
Cross-collaborative services and data exchanged among connected cars	Data received from another connected cars, e.g. the location of a car accident that another connected car spotted on the road, or received accident information from other cars, or smart city info services.
Operational phase: External Services	
Smart cities & services	Economical use of road infrastructure.
Road services	Monitoring road and traffic conditions; Safety recommendations and contextual insights, e.g. speed limit changes, roadway conditions.
Insurance services	Pay-How-You-Drive insurance plan.
Energy/fuel services	Energy/fuel supply.
Marketing services	Monitoring driving habits and user's preferences to create personalized offers.
Maintenance phase: External Services	
Insurance services	Pay-How-You-Drive insurance plan.

Road services	Monitoring traffic conditions; Safety recommendations.
End-of-life phase: External Services	
Smart city services	Environmental impact evaluation. Weather data.

2.4 Digital Twin-based Enhancements of Automated Mobility

The report D5.1 “Data Models for the IIoT and Industry 4.0, and in the Automotive Sector” presents a conceptual view of the Digital Twin-based security and safety evaluations along the identified lifecycle phases (see Figure 8 in D5.1). Simulating the manufacturing and operational behaviour and evaluating performances of connected cars need to be automated through Digital Twins that are technologically capable of replacing manually performed data analytics and reporting with automated decision-making based on (near) real-time measurement and distributed data analytics.

Figure 5 shows that the collection of (near) real-time sensor data (asset data) and data obtained from the external sources, including environmental statistics and government regulations, could be automatically processed through the Digital Twin-based analytics (e.g. location- and temporal behavioural analyses) and correlation of captured behaviour and performances through ML algorithms. At present, the decisions on how to improve the manufacturing and operational procedures that may affect various stakeholders, are mainly under control of human engineers. With the evolution of digital manufacturing and better automated mobility, more intelligence and automation could be pushed to business processes (services), from manufacturing, operational, to administrative services in the connected car ecosystem (see Figure 5).



1 Smart Car data, e.g. vehicle model, on-board sensor data, steering wheel feedback tuning data, data required for tyre/surface modelling

2 Driver’s perception data, e.g. driving skills, inertial cues (orientation), environmental cues (optical and acoustic systems)

3 Data about the connectivity to external systems, e.g. manufacturer, importer, retailer, insurance, other third-party networks

4 Adding newly collected sensor data to the relevant historical data (past data) and KPIs for their analysis and creating security, safety and operational decisions in real-time

5 Prediction & analytics based decisions obtained from the Digital Twin

6 Decisions sent to various stakeholders, including the driver

Figure 5 – Automated Mobility trends with Digital Twins for security, safety and privacy evaluations

Clearly, to achieve any form of a digital transformation and extract the value out of processes and data flows, requires an effectively designed and adopted data-centric approach for decision making. For all of the above to happen, creating a data methodology to enable comprehensive datasets to be acquired for each asset and each identified stakeholder, and to be integrated, consolidated and exploited together with engineering and production data, plant design, commissioning history, maintenance history, etc. is a must in IoT4CPS WP5. But this is easier said than done, as IoT data comes with different access controls and permissions, and is often stored in a range of data pools, from proprietary CRMs, to confidential spreadsheets, cloud data storages with different regulatory compliance and security controls, etc. While extracting value from so called data value chains could be enforced through some emerging forms of value-driven data governance, tools and services to monitor and analyse datasets in order to quantify their value are still missing.

3. IoT4CPS Data Methodology for Digital Twins

Our methodology to support lifecycle data acquisition and data management in IoT4CPS, includes the following steps (see Table 5).

Table 5: Data methodology in IoT4CPS

Description of steps/ activities of the proposed data methodology	To be provided in task / WP…
Searching for the relevant public data sources that are currently available on the web. Public data sources of interest to IoT4CPS need to cover the following topics: • **automotive lifecycle data**, including design, engineering and manufacturing of IoT devices and their CPSs involved in the use cases, as well as operational (driving) data, maintenance and disposal data, and • **automotive cybersecurity, privacy and safety related data**. This data needs to address various cybersecurity lifecycle phases, from cybersecurity monitoring, to discovery, incident response and mitigation management methods. The public datasets need to be analysed, normalised and fused to identify relationships, trends and anomalies to help in reacting to security- and/or safety- related vulnerabilities in the infrastructure. The public datasets need to enable predictions of possible emerging threats against the infrastructure that is monitored by the IoT4CPS Digital Twin demonstrator.	T5.2 TZ5.4.1 WP4 WP7
Searching for the relevant ontologies in the Automotive Driving and Smart Manufacturing sectors, and for cybersecurity, safety and privacy related concepts. Ontologies are used to additionally describe the specific domain, capture the relationships, axioms, rules and restrictions among individuals (instances of objects), classes (concepts) and attributes. Ontologies also enable automated reasoning about data and an easy navigation through the ontological concepts and easy data integration.	T5.2 T5.3 T5.4.1
Design methods to assimilate the diverse data sets, are used to support the decision-making processes.	T5.5.1 T5.5.2 WP4
Developing security, safety and privacy models and risk-based metrics to help security and safety analysts decide on relevant hypothesis (to be described in D5.4.1). These models and metrics need to enable different measures to support stakeholders around the IoT- and CPSs-based systems.	T5.4.1 T5.4.2 WP3 WP4
Considering techniques to address emerging threats in the cloud security. We will look at threat techniques at the two layers: • At the software layer, we will focus on developing security mechanisms against recent malware strains, such as ransomware, and against social engineering attacks, such as phishing (to be described in D5.4.2). • At the infrastructure layer, we will focus on fighting infotainment frauds.	T5.4.1 T5.4.2 WP3 WP4
Assess the increase in the value that is added to the overall security and safety of the system through providing relevant datasets for Digital Twin demonstrator performances. Asses how much improvement can be expected by taking the proposed approach (to be described in D5.5.2 and D5.5.3). Asses the value of collected datasets and data value chains (in D5.5.2 and D5.5.3).	T5.5.2 T5.5.3 WP6 WP7

Apart the pure domain-driven approach to data methodology, e.g. by collecting datasets related to the specific domains, we also look at the IEC 62890 standard, which is about establishing basic principles for lifecycle management of products and systems for industrial-process measurement, control and automation (see the report D5.1 for more details on standards around the two sectors of interest in IoT4CPS). Such PLCDM-based principles are incorporated in RAMI 4.0, that differentiates among the following types of a single IoT product:

Connected World, Enterprise, Work Centres, Station, Control Device, Field Device, and finally, a single device that is used in production (i.e. Product) (see RAMI (2016), Adolphs (2015)).

- The analysis of a single product and its functionalities is provided by RAMI 4.0' IT Layers, e.g. both the Business Layer and the Function Layer describe the conditions and business processes that the system needs to follow.
- The Information Layer provides all data and data models to enable various single product capabilities.
- The Communication Layer enables the connections and information exchanges between a single product and the system.
- Finally, the Integration Layer connects and represents the physical objects of the Asset Layer.

In order to facilitate the requirements identification for Industry 4.0 use cases, including specific data requirements, we perform the mapping (see Figure 6) between the IoT4CPS conceptual view on the Automotive Driving applications, and RAMI 4.0 layers. For example, Figure 6 illustrates mappings between lifecycle axes of both IoT4CPS and RAMI 4.0, as well as at the level of asset management.

Layers which still need to be addressed in the project's use cases are those related to the integration, communication, information, functions and business processes, as well as classification of use case subsystems and their elements according to RAMI 4.0 hierarchy levels, e.g. product, field device, control device, station, etc.
These are all elements that need to be represented by the IoT4CPS data model.

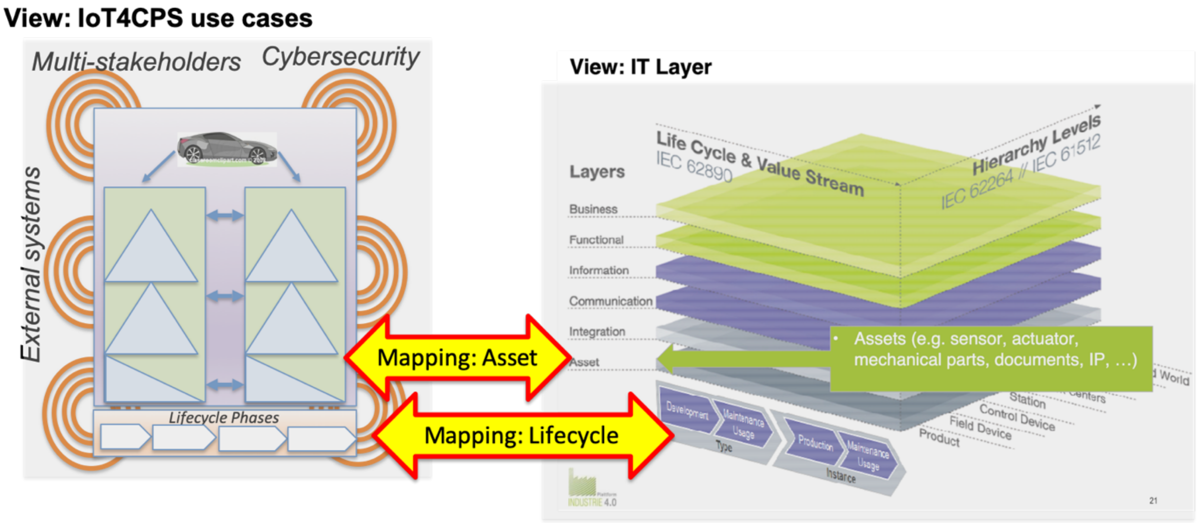


Figure 6 – Identifying Datasets Requirements in IoT4CPS through Mapping with the RAMI 4.0

4. Relevant Public Datasets

4.1 Public Datasets for the Automotive Driving Sector

Leveraging the power of data analytics for Digital Twins and Automotive Driving applications is not as simple. Large scale visual datasets (containing both images and videos) are necessary in computer vision for recognition tasks. Although several large visual datasets exist, they are often irrelevant for the Automotive Driving or not sufficiently diverse to support ML algorithms. For example, they are often limited in describing scene variation, richness of annotation, geographic distribution and road or car conditions to be analysed. To overcome such limitations, the authors in Yu et al. (2018) collect and annotate a new, large-scale dataset of visual driving scenes, called BDD100K datasets. Some other existing datasets focus on particular objects e.g. pedestrians (Dollar et al., 2009; Zhang et al., 2017).

In IoT4CPS, the Digital Twin demonstrator and its applications in the Automotive Driving sector need to learn from a variety of car's and road conditions related to different cities (and countries), changing weather conditions, even changing regulatory policies, etc. High-quality training datasets are essential to create meaningful ML models. Hence, in the following, we analyse several prominent quality datasets that can be at present publicly download for the research experimentations.

4.1.1 Berkeley DeepDrive BDD100k Datasets: Self-Driving AI

The datasets are available from: https://bdd-data.berkeley.edu/
The images in BDD100k come from New York and San Francisco areas.

Berkeley DeepDrive BDD100k is currently the largest and most diverse dataset for self-driving cars. The dataset contains over 100K videos of over 1.100-hour driving experiences across different times of the day and weather conditions. The videos contain diverse kinds of annotations including image level tagging, object bounding boxes, drivable areas, lane markings, and full-frame instance segmentation (Yu et al., 2018; Xu et al., 2017; Xia et al., 2018). The dataset possesses geographic, environmental, and weather diversity, which is useful for training models in a way that they are less likely to be surprised by new conditions.

The dataset format is stored as a JSON file and contains the following details:

```
[
   {
      "name": str,
      "timestamp": 1000,
      "category": str,
      "bbox": [x1, y1, x2, y2],
      "score": float
   }
]
```

4.1.2 The Marulan Datasets: Multi-Sensor Perception for Unmanned Ground Vehicles

The datasets are available from: http://sdi.acfr.usyd.edu.au/datasets/
Other useful links: http://sdi.acfr.usyd.edu.au/

The Marulan datasets are gathered from a multi-sensor Unmanned Ground Vehicles (UGV) called Argo, as described in Peynot et al. (2009), Peynot et al. (2010). The Marulan datasets are large, accurately calibrated and time-synchronised datasets, gathered in controlled environmental conditions using an UGV that is equipped with a wide variety of sensors, e.g. multiple laser scanners, a millimetre wave radar scanner, a colour camera and an infra-red camera, in addition to a cm accuracy dGPS/INS system for localization. The Marulan datasets also include some environmental data acquired from sensors for identifying the presence of dust, smoke and rain in the air. Some other datasets are collected at night.

The authors in Peynot et al. (2009), Peynot et al. (2010) describe the calibration parameters for the sensors in details, and specify the format and content of the data, as well as the environmental conditions in which the data have been gathered. Furthermore, the datasets include static and dynamic tests:

- The static tests consist of sensing a fixed 'reference' terrain, containing simple known objects, from a motionless vehicle.
- The dynamic tests consist of data acquired from a moving vehicle in various environments, including an open area, a semi-urban zone and a natural area with different types of vegetation.

In the Argo UGV, the spatial transformations between sensors and reference frames (navigation frame, body frame and sensor frame) have been estimated through calibration methods (see Figure 7).

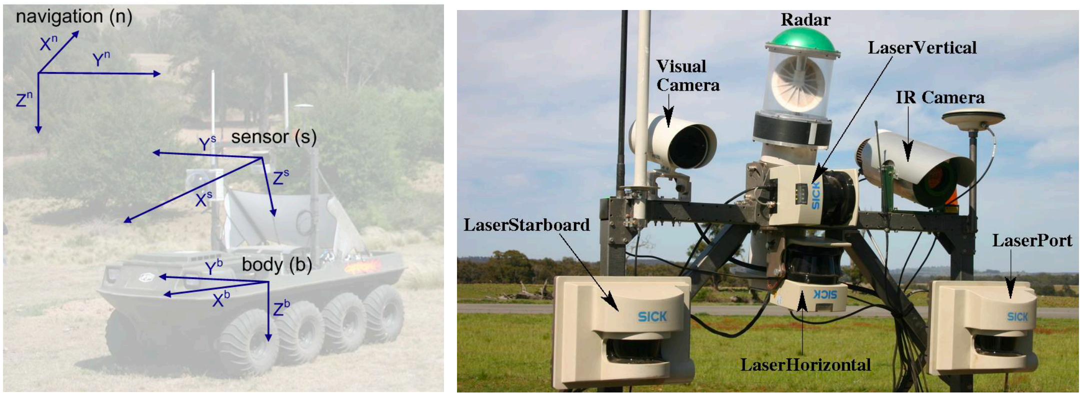


Figure 7 – Argo UGV with sensor, body and navigation frames (on the left). Variety of sensors (on the right)

Each dataset of Marulan has its directory containing data from sensors. A regular dataset directory typically contains ten sub-directories corresponding to different sensors involved, namely:

- LaserHorizontal
- LaserPort
- LaserStarboard
- LaserVertical
- Nav
- Payload
- RadarRangeBearing
- RadarSpectrum
- VideoIR
- VideoVisual

The four lasers: LaserHorizontal, LaserPort, LaserStarboard and LaserVertical, give sensor range data that are actual range values returned by the laser sensor.

<< timestamp, RANGE_DATA, StartAngleRads, AngleIncrementRads, EndAngleRads, RangeUnitType, NScans, Range$_i$>>

The Nav (Navigation, localisation) consists of the three translations (North, East, Down) and the three rotations around the same axis (RollX, PitchY, YawZ), and variations of these entities (dNorth, dEast, dDown, dRollX, dPitchY, dYawZ) and the corresponding covariances matrix ($C_{i,j}$).

<< timestamp, NAV_DATA, North, East, Down, dNorth, dEast, dDown, RollX, PitchY, YawZ, dRoll, dPitch, dYaw, $C_{i,j}$>>

The RadarRangeBearing contains range information from the radar, which is estimated from the spectrum.

<< timestamp, RANGE_REFLECTIVITY_DATA, StartAngleRads, AngleIncrRads, EndAngleRads, RangeUnitType, NScans=1, $Range_1$, $Reflectivity_1$>>

The RadarSpectrum contains the radar spectrum, described as the bins of a Fast Fourier Transform (FFT).

<< timestamp, Angle(degrees), $Reflectivity_i$>>

The VideoIR and VideoVisual relate to the camera images. Both contain the same type of data.

<< timestamp, VISION_FRAME, filename>>

The Payload include additional vehicle internal data, such as status of braking or wheel velocity. That this category of data is only relevant for the dynamic tests (moving vehicle).

<< timestamp, TEXT_TYPE, data>>

Here, <<TEXT_TYPE>> can have various possible data, e.g.:
<< timestamp, BRAKE_DATA, leftBrakePosition(Left Wheel AngularVelocity (rad/s))>>

4.1.3 The UQ St Lucia Datasets: Unaided Stereo Vision Based Pose Estimation

The datasets are available from: http://asrl.utias.utoronto.ca/datasets/index.html
Other useful links: http://asrl.utias.utoronto.ca/~mdw/uqstluciadataset.html

The authors in Warren et al. (2010) publish the UQ St Lucia Dataset as is a vision dataset gathered from a car driven in a 9.5km circuit around the University of Queensland's St Lucia campus on the day of 15.12.2010. The data consists of visual data from a calibrated stereo pair, translation and orientation information as a ground truth from an XSens Mti-g INS/GPS and additional information from a USB NMEA GPS. The dataset traverses local roads and encounters a number of varying scenarios including roadworks, speed bumps, bright scenes, dark scenes, reverse traverses, a number of loop closure events, multi-lane roads, roundabouts and speeds of up to 60 km/h.

This dataset has been released for public use in testing and evaluating stereo visual odometry and visual SLAM algorithms. The dataset formats include the following details:

GPS
<<Timestamp (seconds), Timestamp (millsec), UTC Time Hours, UTC Time Minutes, UTC Time Second, Latitude (deg), Longitude (deg), Altitude (deg), Horizontal Position Error, Vertical Position Error, Heading, Speed (m/s), Climb rate (m/s)>>

INS (Inertial Navigation System)

<<Timestamp (seconds), Timestamp (millisec), Latitude (deg), Longitude (deg), Altitude (m), Height AMSL (m), Velocity East North Up (m/s), Velocity East North Up (m/s), Velocity East North Up (m/s), Roll (rad), Pitch (rad), Yaw (rad)>>

IMU (Inertial Measurement Unit)

<<Timestamp (seconds), Timestamp (milliseconds), X Acceleration (m/s^2), Y Acceleration (m/s^2), Z Acceleration (m/s^2), Gyro X Velocity (rad/s), Gyro Y Velocity (rad/s), Gyro Z Velocity (rad/s)>>

Multicamera datasets

<< Total Number of Cameras, Cam 0 Expected Image Width (pixels), Cam 0 Expected Image Height (pixels), Cam 0 Data format, Cam 0 Frame Rate (Hz), Cam 0 X offset (m), Cam 0 Y offset (m), Cam 0 Z offset (m), Cam 0 Roll offset (rad), Cam 0 Pitch offset (rad), Cam 0 Yaw offset (rad), Cam 0 Focal Length Fx (pixels), Cam 0 Focal Length Fy (pixels), Cam 0 Principle Point cx (pixels), […continuing data for Cam 1, ….]>>

<<Cam 0 Timestamp (seconds), Cam 0 Timestamp (millisec), Cam 0 Image Width (pixels), Cam 0 Image Height (pixels), Cam 0 Data format, […continuing data for Cam 1, ….]>>

4.1.4 The Malaga Datasets: A Collection of Outdoor Robotic Datasets

The datasets are available from: https://www.mrpt.org/malaga_dataset_2009

The Malaga datasets collect outdoor data from a large and heterogeneous set of sensors comprising colour cameras, several laser scanners, precise GPS devices and an inertial measurement unit (Blanco et al., 2009). It also includes a unified and extensive testbed for comparing and validating different solutions to common robotics applications, such as Simultaneous Localization and Mapping (SLAM), video tracking or the processing of large 3D point clouds. All the datasets, calibration information and associated software tools are available for download.

The format of the vehicle path file is plain text, each row containing the 6D pose at a given instant and its uncertainty (its covariance matrix). Each data column contains the following details:

<<Timestamp, X_coordinate (meters), Y_coordinate (meters), Z_coordinate (meters), yaw_angle (radians), pitch_angle (radians), roll_angle (radians), C_{ij} (the covariance matrix)>>

4.1.5 The KAIST Urban Datasets

The dataset is available from: http://irap.kaist.ac.kr/dataset/

This data set provides Light Detection and Ranging (LiDAR) data and stereo image with various position sensors collected from a highly complex urban environment. The presented dataset captures a variety of features in urban environments (e.g. metropolis areas, complex buildings and residential areas) (Jeong et al., (2019-1)).

The dataset formats include the following details[1]:

2D LiDAR Dataset

This is a simplified version of datasets, emphasizing two dimensions for the data:

- *r* is the range value of each point, and
- *R* is the reflectance value.

The sensor's Field of View (FOV) is 190°. The start angle of the first data is −5°, and the end angle is 185°. The angle difference between each sequential data is 0.666° (Jeong et al., 2019). The timestamps of all 2D LiDAR data are stored sequentially in 'SICK_back_stamp.csv' and 'SICK_middle_stamp.csv'.

```
<<r, R>>
```

3D LiDAR Dataset

The timestamp of the last packet is used as the timestamp of the data at the end of one rotation. The timestamps of all 3D LiDAR data are stored sequentially in 'VLP_left_stamp.csv' and 'VLP_right_stamp.csv'.

```
<<x, y, z, R>>
```

- x, y and z denote the local 3D Cartesian coordinate of each LiDAR sensor;
- R is the reflectance value.

Stereo Camera Image Datasets

The stereo images were acquired at 10 Hz and stored in the lossless PNG format in unrectified 8-bit Bayer pattern images. The Bayer pattern of the images is RGGB. Each image is named using timestamps.

GPS (Global Positioning System) Datasets

```
<<timestamp, latitude, longitude, altitude, 9-tuple vector (position covariance)>>
```

VRS (Virtual Reference Station) GPS Datasets

Version 1:

```
<<timestamp, latitude, longitude, x_coordinate, y_coordinate, altitude,
fix_state, number_of_satellite, horizontal_precision, latitude_std,
longitude_std, altitude_std, heading_validate_flag,
magnetic_global_heading, speed_in_knot, speed_in_km, GNVTG_mode>>
```

Version 2:

```
<<timestamp, latitude, longitude, x_coordinate, y_coordinate, altitude,
fix_state, number_of_satellite, horizontal_precision, latitude_std,
longitude_std, altitude_std, heading_validate_flag,
magnetic_global_heading, speed_in_knot, speed_in_km, GNVTG_mode,
ortometric_altitude>>
```

The `x` and `y` coordinates are in the UTM coordinate system of the meter unit. The `fix_state` is a number indicating the state of the VRS GPS.

[1] http://irap.kaist.ac.kr/dataset/dataformat.html

IMU (Inertial Measurement Unit) Datasets

The IMU datasets contain details such as: rotational pose, gyro, acceleration, magnet filed data measured by AHRS IMU sensor.

Version 1:

```
<<timestamp, quaternion_x, quaternion_y, quaternion_z, quaternion_w,
Euler_x, Euler_y, Euler_z>>
```

Version 2:

```
<<timestamp, quaternion_x, quaternion_y, quaternion_z, quaternion_w,
Euler_x, Euler_y, Euler_z, Gyro_x, Gyro_y, Gyro_z, Acceleration_x,
Acceleration_y, Acceleration_z, MagnetField_x, MagnetField_y,
MagnetField>>
```

FOG (Fiber Optic Gyro) Datasets

The FOG dataset stores the relative rotational motion between consecutive sensor data.

```
<<timestamp, delta_roll, delta_pitch, delta_yaw>>
```

Encoder Datasets

The Encoder dataset stores the incremental pulse count values of wheel encoder.

```
<<timestamp, left_count, right_count>>
```

Altimeter Datasets

The Altimeter dataset stores the altitude values measured by the altimeter sensor.

```
<<timestamp, altitude>>
```

Calibration Datasets

Calibration is performed whenever data is acquired. The calibration data is provided in both the Euler format and SE(3) format.

Sequence Dataset

This dataset stores the names and timestamps of all sensor data in order.

```
<<timestamp, sensor_name>>
```

Baseline

A baseline is the trajectory of the vehicle generated by various algorithms.

4.1.6 Electric Vehicle Trial Data

The datasets are available from: https://data.gov.au/dataset/87f276c3-5fba-4f31-9032-199793d6f4a7/resource/5092c15c-a5c6-4a6f-bb40-492028470fc0/download/sgsc-ev-electric-vehicles-data.csv
Other useful links: https://data.gov.au/dataset/ds-dga-87f276c3-5fba-4f31-9032-199793d6f4a7/details

This trial dataset involves a fleet of 20 Mitsubishi iMiEV cars (2010 model) used by households and businesses. The data details Electric Vehicle trips during the trial, including car appliances in use (air-conditioning, lights, etc), odometer readings, speed, trip length and the battery state of charge at the beginning and end of each trip. Each data column contains the following details:

```
<< TEL_TRIPSTART, TEL_STARTODO, TEL_TRIPEND, TEL_ENDODO, TEL_SOC_START,
TEL_SOC_END, TEL_AVG_VELOCITY, AC_ON_DURATION, AC_ON_PCT,
HEADLAMP_ON_DURATION, HEADLAMP_ON_PCT, TRIPSTART_LOCN, TRIPEND_LOCN,
TEMPERATURE, CHARGE_TYPE, CONNECT_START, CHARGE_START, CONNECT_END,
CHARGE_END, CHARGE_AMOUNT, FULL_CHARGE_FLAG, CHARGING_OCCURRENCES,
UNACCOUNTED_KM, FULL_CHARGE_KMSPAN,
FULL_CHARGE_EFFICIENCY, FULL_CHARGE_AMOUNT, TRIP_ID, CALENDAR_KEY,
VEHICLE_ID, REC_START_DATE, REC_END_DATE, START_DATE, START_ODO_ACTUAL,
START_BAT_EST, START_BAT_ACTUAL, END_DATE, END_ODO_EST, END_ODO_ACTUAL,
END_BAT_EST, END_BAT_ACTUAL, TRIP_DISTANCE, DISTANCE_GPS, TEL_DISTANCE,
CALC_DISTANCE, NO_PASSENGERS, ACCESSORIES, LOG_AVG_VELOCITY,
GPS_AVG_VELOCITY, START_ALTITUDE, END_ALTITUDE>>
```

4.1.7 Other Automotive Driving Datasets

The list of other useful datasets related to Automotive Driving is given in Table 6 below.

Table 6: Other public datasets for the Automotive Driving sector

Dataset name, URL, reference	Short description	Target env.	Sensors/image	Sensors/LiDAR	Sensors/GPS	Sensors/IMU	Sensors/Encoder
Training Dataset for Self-Driving Car Available from: https://www.kaggle.com/roydatascience/training-car	Large dataset comprising of all images used in training the model for Self-Driving Car applications	Urban	+	+	+		
ApolloScapes Available from: https://github.com/ApolloScapeAuto/dataset-api	Large dataset that defines 26 different items e.g. cars, bicycles, pedestrians, buildings, streetlights, etc. Huang et al. (2018)	Urban	+	+	+	+	
KITTI Available from: https://www.kaggle.com/kerneler/starter-kitti-vehicles-96b4f3f8-8	The KITTI dataset contains online benchmarks for visual odometry, image tracking, and semantic segmentation. This is the most widely used dataset in autonomous vehicle and computer vision research. Geiger et al. (2013)	Urban	+	+	+	+	
Cityscape Available from: https://www.cityscapes-dataset.com/dataset-overview/	A large dataset that records urban street scenes in 50 different cities.	Urban	+	+	+	+	
Oxford's Robotic Car Available from: https://robotcar-dataset.robots.ox.ac.uk/	Over 100 repetitions of the same route through Oxford, UK. It captures weather, traffic, and pedestrian data, along with data on construction and roadworks. Maddern et al. (2016)	Urban	+	+	+	+	+
Ford campus vision and LiDAR dataset	The data collected from sensors mounted on the vehicle (e.g. Applanix	Campus	+	+	+	+	+

Available from: http://robots.engin.umich.edu/SoftwareData/Ford Pandey et al. (2011)	POS LV and consumer Xsens MTI-G IMU, a Velodyne 3D-LiDAR scanner, two push-broom forward looking Riegl LiDARs, and a Point Grey Ladybug3 camera system. The data is collected while driving the vehicle around the Ford Research campus and downtown Dearborn, Michigan in 2009.						
WPI datasets Available from: http://computing.wpi.edu/dataset.html	Datasets for traffic lights, pedestrian and lane detection	Urban	+	+	+	+	
Bosh Small Traffic Light Dataset Available from: https://hci.iwr.uni-heidelberg.de/node/6132 https://github.com/bosch-ros-pkg/bstld	Dataset for small traffic lights for deep learning. Behrendt & Novak (2017)	Urban	+		+		
MIT AGE Lab Available from: https://lexfridman.com/carsync/#Dataset Fridman et al., (2016)	A sample of the 1,000+ hours of multi-sensor driving datasets collected at AgeLab. It contains: • Individual Sensor Data related to each individual sensor: video, audio, IMU, GPS and steering wheel position. • Video Data with timestamps for each frame in the video. • Synchronized Data with the individual sensor data fused and sampled at the frame rate (fps) specified in the filename.	Campus	+	+	+	+	+
LISA (Laboratory for Intelligent & Safe Automobiles) UC San Diego Datasets Available from: http://cvrr.ucsd.edu/LISA/vehicledetection.html	This dataset includes traffic signs, vehicles detection, traffic lights, and trajectory patterns.	Urban	+		+		
New College Dataset Smith et al. (2009)	It provides data from a campus and a park that were obtained using a Segway robotic platform. A LiDAR was mounted on the side of the Segway to obtain distance measurements providing both both stereo and omnidirectional images.	Campus	+	+	+	+	+
Traffic, Driving Style and Road Surface Condition	Low-level parameters acquired by the car via OBD-II and through the micro-devices embedded in the user smartphone, with the goal of accurately characterizing the	Urban	+		+	+	+

(Ruta et al., 2018) Available from: https://www.kaggle.com/gloseto/traffic-driving-style-road-surface-condition	overall system composed by driver, vehicle and environment Predicted attribute: road surface, traffic and driving style The dataset includes the following attributes: altitude change (over 10 seconds); current speed value; average speed in the last 60 sec; speed variance in the last 60 sec; speed variation for every second of detection; longitudinal acceleration; engine load (in %); engine coolant temperatures (in C degree); Manifold Air Pressure (MAP) (to compute the optimal air/fuel ratio); Revolutions Per Minute (RPM) of the engine; Mass Air Flow (MAF) Rate measured in g/s, (to set fuel delivery and spark timing); Intake Air Temperature (IAT) at the engine entrance; vertical acceleration, measured by the smartphone accelerometer and pre-processed with a low-pass filter.						
Automotive Sensor Data. An Example Dataset from the AEGIS Big Data Project Available from: https://zenodo.org/record/820576/files/Automotive-ResearchDataSet-VIF-AEGIS.zip?download=1 Stocker et al. (2017)	This is a research dataset for the automotive demonstrator within the "AEGIS - Advanced Big Data Value Chain for Public Safety and Personal Security" project, which has received funding from the EU H2020 research and innovation programme (grant agreement No 732189). The time series data has been collected by using a BeagleBone single plate computer which collects data for driving analytics. The BeagleBoard can be connected to the OBD2 interface of a vehicle to capture data from CAN bus and has been additionally equipped with further sensors (GPS, gyroscope, acceleration). The data in this research dataset was collected during 35 different trips conducted by one driver driving one vehicle in the Graz area, Austria.	Urban	+	+	+	+	+

4.2 Public Datasets for the Automotive Manufacturing Sector

4.2.1 Public Datasets on Software Metrics

Public datasets on software metrics are mainly created from open source projects and often provide static code metrics (Lines of Code (LOC), Cyclomatic Complexity by McCabe (CC), etc.) and bug information (Altinger, 2016). Most of the bug commit information has been extracted using the SZZ algorithm, which is introduced by Śliwerski et al. (2015) as an approach to identify bugs in a software repository (including GitHub repositories). The name of the SZZ algorithm is given after the initials of the three authors.

One of the first public available datasets has been released by the **NASA metric data program (NASAMDP)** (NASAMDP, 2004) (online available from: http://mpd.ivv.nasa.gov). The NASAMDP datasets contain software metrics collected at ten different projects within NASA flight software. Another dataset containing software engineering data is called **PROMISE** (online available from: http://promise.site.uottawa.ca/SERepository/). PROMISE is founded and administrated by Sayyad et al., (2005). and Menzies et al., (2005). PROMISE includes 60 projects usable for Software Fault Prediction (SFP). Similarly, **Software-artifact Infrastructure Repository (SIR)** published by Do et al., (2005) can be considered to be the first database on software bugs, containing 81 projects with a rather small code size ranging from 24 LOC to 8.570 LOC.

Practically, the only industrial available dataset for SFP has been released by NASAMDP and the PROMISE repository (Altinger, 2016). Menzies et al., (2005) shows how ML approaches can be used to build up defect prediction models. For example, the following ML algorithms: OneR, J48, and NB, can be used to predict error prone software modules. The NASAMDP and the PROMISE repository of software engineering data can be used for the evaluation of the predictions.

Finally, Altinger (2015) contains datasets on automotive software repository that is publicly available from: http://www.ist.tugraz.at/_attach/Publish/AltingerHarald/MSR_2015_dataset_automotive.zip

A collection of other bug datasets is given in Table 7.

Table 7: Other public datasets on software metrics (based on (Altinger, 2016))

Reference to dataset	Created for	Hosting
Zimmerman et al., 2007	Eclipse 2.0, 2.1 and 3.0.	25.210 files with 25.585 defects
Kamei et al., 2008	Eclipse 3.0 and 3.1	9.726 Java files of whom 16,98% are marked as faulty
Herraiz et al., 2009	5000 open source projects	N/A
Mockus et al., 2009	GoogleCode and SourceForge	1398 projects with 207.904.557 files in total
D'Ambros et al., 2010	Eclipse JDT Core, Eclipse PDE UI, Equinox framework, Mylyn and Apache Lucene projects	It contains software consisting of 2.131 classes and containing 1.923 bug commits.
Jus et al., 2014	Software testing research	The initial commit contains 357 bugs on five open source Java projects ranging between 22.000 and 96.000 LOC.

4.2.2 Other Automotive Manufacturing Datasets

Public datasets related to the automotive manufacturing sector are listed in Table 8. Note that getting manufacturing datasets is almost not possible due to their commercially sensitive nature. NASA has made available some datasets from large civil aircraft with associated faults (see Table 7).

Table 8: Public datasets for the Automotive Manufacturing sector

Dataset name, URL, reference	Short description of the dataset
Mercedes-Benz Greener Manufacturing Available from: https://www.kaggle.com/c/mercedes-benz-greener-manufacturing/data	This dataset contains an anonymized set of variables, each representing a custom feature in a Mercedes car. For example, a variable could be 4WD, added air suspension, or a head-up display. The ground truth is labelled 'y' and represents the time (in seconds) that the car took to pass testing for each variable.
Production Plan Data for Condition Monitoring Available from: https://bit.ly/2Nkkohi	The dataset includes eight run-to-failure experiments and eight features of the component for which the predictions need to be performed, within production lines. The condition of this component is important for the function of the plant and the resulting product quality. The degradation of the component under test was calculated and visualized. This procedure was repeated for all eight datasets to get

	a prediction of the degradation for all components (Birgelen et al., 2018).
The NIST manufacturing robotics test bed Available from: https://bit.ly/2JdaMzI https://bit.ly/2LnYGXb	It consists of several labs located in three buildings on the main NIST campus. Combined, these serve as a resource for research in robotics for advanced manufacturing and material handling. The test bed contains representative state-of-the-art manufacturing robots, including ones that have been designed specifically for safe interactions with human workers in shared environments. The testbed also includes advanced multi-fingered grippers, sensors, conveyors, and an industrial robot arm that can be mounted on a linear rail or on a pedestal. A custom-configured Automatic Guided Vehicle (AGV) is used for research in industrial vehicle safety and performance standards, including mobile manipulation. The robot systems also include vision and force-torque sensing capability. Research in the testbed labs focuses on human-robot collaboration, rapid re-tasking of robot systems, improvements to robot safety standards, and performance evaluation of robots, industrial vehicle systems, sensor systems, and dexterous manipulation for industrial applications.
Data for pollution project Available from: https://bit.ly/2Xg3foQ	Pollution data collected in Singapore. Measurement of various pollutants: lead, carbon, nitrogen- dioxide, ozone, etc.
Milling datasets Available from: https://ti.arc.nasa.gov/c/4/b Agogino and Goebel (2007)	Experiments on a milling machine for different speeds, feeds, and depth of cut. Records the wear of the milling insert, VB. The data set was provided by the BEST lab at UC Berkeley.

5. Ontological Data Representation Models to Support PDLCM

Knowledge engineering mostly relates to the construction of shared conceptual frameworks, which could be designated as ontologies or knowledge graphs. In the following, we explore currently existing ontologies related to the Smart (Automotive) Manufacturing, the Automotive Driving and PLCDM.

5.1 Ontologies for the Smart Manufacturing Sector

Domain knowledge about the factory floor processes and manufacturing equipment and assets, needs to be modelled and integrated into the Digital Twins' applications and services. Here, the ontologies can be seen as natural candidates for implementing a variety of Knowledge Based Systems (KBSs) (Giovannini et al. 2012)). For example, the ontologies can be used to capture a formal and shared representation of a particular domain of disclosure, e.g. in the Smart Manufacturing sector.

Table 9 summarizes some of well-known ontologies in the manufacturing domain.

Table 9: Manufacturing ontologies

Reference	Short description of the ontology
Cai, Zhang and Zhang (2001)	An ontology-based solution to demonstrate the interoperability between manufacturing services.
Diep, Alexakos and Wagner (2007)	The P2 Ontology, to enable interoperability between components and applications throughout the manufacturing process lifecycle.
Lemaignan et al. (2006)	MASON, an upper ontology of manufacturing systems
Menzel and Grüninger (2001), Schlenoff et al. (2000)	The PSL (Process Specification Language) Ontology was designed to facilitate the exchange of process information among manufacturing systems and has been published as ISO 18629.
Chang, Rai and Terpenny (2010)	The DFM (Design For Manufacturing) Ontology captures relevant domain manufacturing knowledge, enhances the knowledge exchange and retrieval of manufacturing design alternatives from heterogenous data sources, and supports designers in making design decisions.
Alsafi and Vyatkin (2010)	A reconfiguration agent that is based on the MASON Ontology and designed to infer knowledge about the manufacturing environment and its requirements.
Ameri, Urbanovsky and McArthur (2012)	A systematic approach for the development of manufacturing ontologies based on MSDL (Manufacturing Service Description Language) and the MSDL Ontology that enhances formal representation of manufacturing services in mechanical machining domain.
Kiritsis et al. (2013)	The LinkedDesign Ontology that can be adjusted and adopted for different manufacturing systems.
Chungoora et al. (2013)	The core ontological concepts encoded in the ECLIF (Extended Common Logic Interchange Format) (ECLIF (2010)) format for application configurations of products and information platforms in manufacturing domains.
Garetti and Fumagalli (2012), Garetti and Fumagalli (2012a), Garetti, Fumagalli and Negri (2015)	The P-PSO (Politecnico di Milano – Production Systems Ontology) provides a metamodel of various manufacturing system domains and applications.
Negri et al. 2015a; Negri et al. 2017	The P-PSO ontology has evolved into the MSO (Manufacturing Systems Ontology) for logistics, discrete and production manufacturing systems and processes.
Garetti et al. (2013)	The integration of ontologies and Web Services within the control architecture of automated manufacturing systems.
Negri et al. (2015)	The core requirements for the use of manufacturing domain ontologies in a Web Service architecture for the control of manufacturing systems.

Mohammed et al. (2017)	To design and implement a flexible architecture for event-driven manufacturing systems that can be deployed in multiple industrial cases, the authors combine the flexibility of knowledge-driven systems with the vendor-independent properties of RESTful Web Services.
Zhang et al. (2012)	The Cloud Recommender System based on an OWL (Web Ontology Language) ontology.
Afify et al. (2013)	A Software-as-a-System discovery and selection system based on the WordNet ontology.
Liu et al. (2014)	Ontology-based service matching in Cloud computing.
Rodríguez-García et al. (2014)	An ontology-based annotation and service retrieval system for the Cloud environment.

In addition, semantic integration of sensor data has been explored through many efforts to create sensor taxonomies, ontologies and standards, as summarized in Table 10.

Table 10: Other ontologies that can be applied to sensor data applications

Reference	Short description of the ontology
vCard https://www.w3.org/TR/vcard-rdf/	vCard describes contacts (people and organizations). It is an IETF RFC6350 standard. The ontology was created in 2014.
SKOS (Simple Knowledge Organization System) https://www.w3.org/TR/2008/WD-skos-reference-20080829/skos.html	It is W3C recommendation from August 2009.
GEO https://www.w3.org/2005/Incubator/geo/XGR-geo-ont-20071023/	Geo vocabulary
GR (Good Relations) https://www.w3.org/wiki/GoodRelations	It is an ontology for e-commerce applications, used by Google, Amazon, etc.

PROV-O (Provenance Ontology)
https://www.w3.org/TR/prov-o/
	It describes the provenance of data using:
	`prov:Agent (organization, person, software, sensor, actuator…)`
	`prov:Activity (translate a document, predict, measure…)`
	`prov:Entity (observation, prediction, aggregated or obfuscated…)`

OM http://www.semantic-web-journal.net/content/ontology-units-measure-and-related-concepts	Ontology of Units of Measure, precision, etc.
CDT https://ci.mines-stetienne.fr/lindt/v2/custom_datatypes.html	Custom Datatypes
OWL-Time https://www.w3.org/TR/owl-time/	Time ontology in OWL
SOSA/SSN https://www.w3.org/TR/vocab-ssn/	Semantic Sensor Networks (OGC, W3C)
SAREF http://ontology.tno.nl/saref/	Smart Appliances REFerence Ontology (ETSI)
Dogont http://iot-ontologies.github.io/dogont/	An ontology for intelligent environments

5.2 Ontologies for the Automotive Driving Sector

Some of the most prominent automotive ontologies are summarized in Table 11.

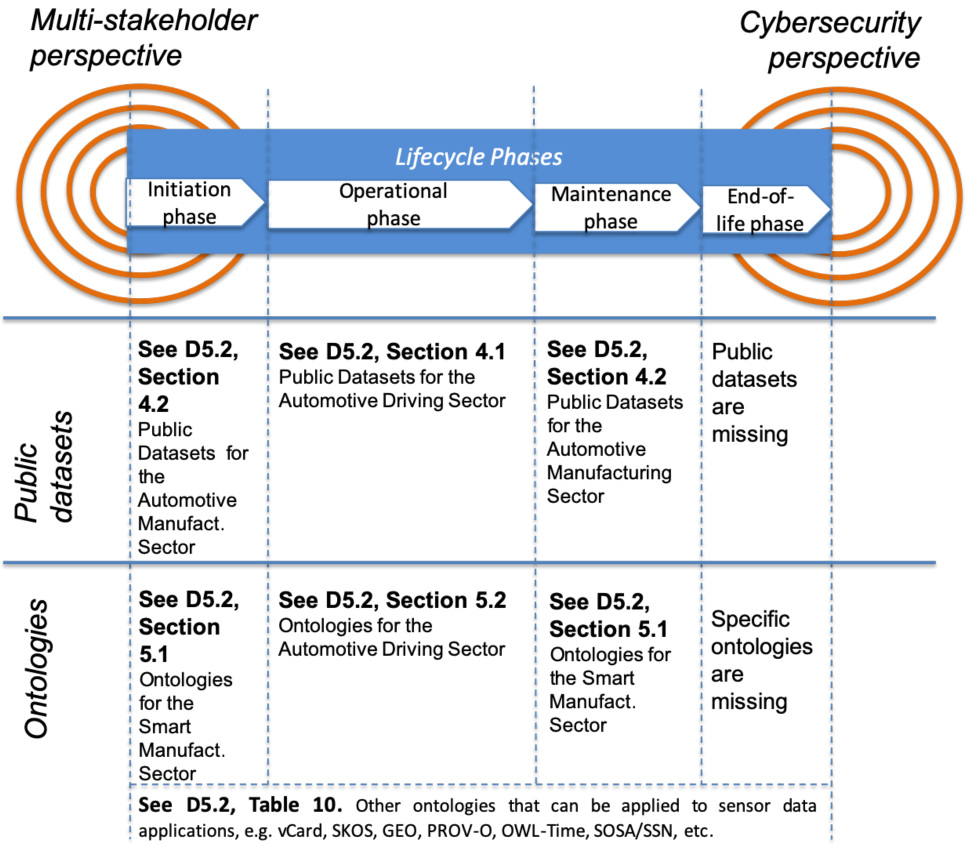
Table 11: Other ontologies that can be applied to sensor data applications

Reference	Short description of the ontology
Automotive Ontology http://www.automotive-ontology.org/ https://www.w3.org/community/gao/	OWL ontology for the automotive industry.
Volkswagen Ontology http://www.volkswagen.co.uk/vocabularies/vvo/ns	It contains specific domain vocabulary and an approach that describes an ontology for automotive HMI (Feld and Müller, 2011).
ISO 26262 Standard for Automotive Functional Safety	This is not an ontology, but Standard for Automotive Functional Safety. It encompasses standard definitions for specific terms related to the deployment of the functional safety process along the entire engineering environment chain. It contains some ontological aspects described in its *Vocabulary*.
OSLC (Open Services Lifecycle Collaboration)	OSLC defines some general vocabularies and several specifications, which are related to ontologies.
VSSo (A Vehicle Signal and Attribute Ontology) https://klotzbenjamin.github.io/vss-ontology/	An ontology for describing automotive attributes, branches and signals. It is based on the Vehicle Signal Specification (GENIVI) and reuses the SSN/SOSA pattern for signals.

6. Design Methods for the PLCDM- and Multi-Stakeholder-Centered Data Models in IoT4CPS

The PLCDM- and multi stakeholders-centered data models are crucial for modelling of the Digital Twin demonstrator performances in IoT4CPS, through its task T5.2. Specifically, these data models will be extended by adding the security and safety lifecycle models and threat intelligence models in task T5.4, and cross-collaboration aspects with external stakeholders in task T5.3. The final data models will underline the targeted security and safety engineering and verification capabilities, through the Digital Twin demonstrators.

In this report, we present our methodology to create relevant datasets, based on public datasets that address the major lifecycle stages of the Automotive Driving. Figure 8 illustrates the intended scope of the data models in IoT4CPS, in its final stage of the project's progress.



	Initiation phase	Operational phase	Maintenance phase	End-of-life phase
Public datasets	**See D5.2, Section 4.2** Public Datasets for the Automotive Manufact. Sector	**See D5.2, Section 4.1** Public Datasets for the Automotive Driving Sector	**See D5.2, Section 4.2** Public Datasets for the Automotive Manufacturing Sector	Public datasets are missing
Ontologies	**See D5.2, Section 5.1** Ontologies for the Smart Manufact. Sector	**See D5.2, Section 5.2** Ontologies for the Automotive Driving Sector	**See D5.2, Section 5.1** Ontologies for the Smart Manufact. Sector	Specific ontologies are missing

See D5.2, Table 10. Other ontologies that can be applied to sensor data applications, e.g. vCard, SKOS, GEO, PROV-O, OWL-Time, SOSA/SSN, etc.

Figure 8 – Underlying data model for the Digital Twin demonstrator in IoT4CPS, based on public data sources

The selected candidate public datasets to be aggregated and assimilated for the purpose of PLCDM analyses in the Digital Twin demonstrator are the following ones:

- Initiation phase:
  - **Milling datasets** (available from https://ti.arc.nasa.gov/c/4/b) that include experiments on a milling machine for different speeds, feeds, and depth of cut.
  - **Mercedes-Benz Greener Manufacturing** dataset (available from https://www.kaggle.com/c/mercedes-benz-greener-manufacturing/data) contains an anonymized set of variables, each representing a custom feature in a Mercedes car.
  - **The NIST manufacturing robotics test bed** (available from https://bit.ly/2JdaMzI) consists of datasets on robotics for advanced manufacturing and material handling.

  - **Data for pollution project** (available from https://bit.ly/2Xg3foQ) containing measurement of various pollutants: lead, carbon, nitrogen- dioxide, ozone, etc.
- Operational phase:
  - **Berkeley DeepDrive BDD100k** is one of the largest and most diverse datasets for self-driving cars, containing over 100K videos of driving experiences enhanced by geographic, environmental, and weather diversity.
  - **ApolloScapes** (available from https://github.com/ApolloScapeAuto/dataset-api) is a large dataset describing 26 types of stakeholders, e.g. cars, bicycles, pedestrians, buildings, streetlights, etc.
  - **KITTI** (https://www.kaggle.com/kerneler/starter-kitti-vehicles-96b4f3f8-8) with online benchmarks for visual odometry, image tracking, and semantic segmentation.
  - **Traffic, Driving Style and Road Surface Condition** (available from: https://www.kaggle.com/gloseto/traffic-driving-style-road-surface-condition) includes many attributes for predicting road surface, traffic and driving style.
  - **Automotive Sensor Data. An Example Dataset from the AEGIS Big Data Project** (https://zenodo.org/record/820576/files/Automotive-ResearchDataSet-VIF-AEGIS.zip?download=1) with data collected during 35 different trips conducted by one driver driving one vehicle in the Graz area, Austria.
- Maintenance phase:
  - **Production Plan Data for Condition Monitoring** (https://bit.ly/2Nkkohi) that observes features of several components for which the predictions need to be performed.
  - **Data for pollution project** (available from https://bit.ly/2Xg3foQ) containing measurement of various pollutants: lead, carbon, nitrogen- dioxide, ozone, etc.
  - Industrial available dataset for SFP i.e. **NASAMDP** and **PROMISE**.
- End-of-Life phase:
  - Public datasets for this phase are missing and will be generated for the purpose of the research experimentations in IoT4CPS.

7. Conclusion

Current challenges in data centric projects still relate to the creation and analysis of relevant data. The Automotive industry typically do not publicise data due to its commercially sensitive nature, that can possibly allow for data to be copied, which increases the company's risk through data theft, loss or exposures. Hence, in IoT4CPS, we look at relevant open-source systems and public datasets that can be reused or used as a basis to construct specific data repository of interest to the project's business cases. We identified some gaps with currently missing datasets related to particular PLCDM stages in the Automotive sector, e.g. end-of-life phase. The end-of-life phase of IoT and CPSs is generally not addressed in the current research, and we see the potential for IoT4CPS to address the current gaps in two ways: firstly, we plan to generate missing data that will be further assimilated with the IoT4CPS data model for the experimentation through the Digital Twin demonstrator, and secondly, we will contribute to the creation of roadmaps to better address missing PLCDM stages in the Automotive industry.

In addition, the IoT4CPS data models will be semantically enriched by using standard ontologies and some newly proposed ontologies for formal description of car signals and sensors, e.g. VSSo (Vehicle Signal and Attribute Ontology) (Klotz et al., 2018). Such approach will contribute to the cross-interaction of the connected car system with the external stakeholders in the cloud (based on semantically annotated interaction between the connected cars and web services (microservices)).

8. References

Adolphs, P. (2015). RAMI 4.0: An Architectural Model for Industrie 4.0. Available online from: https://www.omg.org/news/meetings/tc/berlin-15/special-events/mfg-presentations/adolphs.pdf (last accessed: May 2019)

Afify, Yasmine, Moawad, Ibrahim Fathy, Badr, Nagwa, Tolba, Mohamed. 2013. "A Semantic-Based Software-As-A-Service (SAAS) Discovery and Selection System." In Proceedings of the 8th International Conference on Computer Engineering & Systems (IC-CES). IEEE. pp. 57- 63.

Agogino, A. and Goebel, K. (2007). BEST lab, UC Berkeley. "Milling Data Set ", NASA Ames Prognostics Data Repository (http://ti.arc.nasa.gov/project/prognostic-data-repository), NASA Ames Research Center, Moffett Field, CA

Alsafi, Yazen, and Vyatkin, Valeriy. 2010. "Ontology-based Reconfiguration Agent for Intelligent Mechatronic Systems in Flexible Manufacturing." In Robotics and Computer-Integrated Manufacturing, 26(4), 381-391.

Altinger, H., 2015. Dataset on automotive software repository, Feb. 26, 2015. Online available from: http://www.ist.tugraz.at/_attach/Publish/AltingerHarald/MSR_2015_dataset_automotive.zip

Altinger, H., 2016. State of the Art Software Development in the Automotive Industry and Analysis upon Applicability of Software Fault Prediction. Doctoral Thesis. Graz University of Technology. 2016. Online: http://www.ist.tugraz.at/_attach/Publish/AltingerHarald/PHD_Altinger_automotive_SW_analysis.pdf

Ameri, Farhad, Urbanovsky, Colin, and McArthur, Christian. 2012. "A Systematic Approach to Developing Ontologies for Manufacturing Service Modeling." In Proc. of the 7th Inter. Conf. on Formal Ontology in Information Systems (FOIS).

Behrendt,. K. and Novak, L. (2017). A Deep Learning Approach to Traffic Lights: Detection, Tracking, and Classification. In Proceedings of the 2017 IEEE International Conference on Robotics and Automation (ICRA).

Binder, C., Neureiter, C., Lastro, G., Uslar, M., and Lieber, P. (2018). "Towards a Standards-Based Domain Specific Language for Industry 4.0 Architectures". In Proceedings of the Ninth International Conference on Complex Systems Design & Management, CSD&M Paris 2018, pp. 44-55.

Bitkom, VDMA (2015) ZVEI: Umsetzungsstrategie Industrie 4.0, Ergebnisbericht der Plattform In-dustrie 4.0.

Blanco, J.L., Moreno, F.A., and Gonzalez, J. (2009). A Collection of Outdoor Robotic Datasets with centimeter-accuracy Ground Truth. Autonomous Robots, Vol. 27, No. 4, pp. 327—351. DOI: 10.1007/s10514-009-9138-7

BMVIT (2016). "Automated – Connected – Mobile. Action Plan Automotive Driving – Executive Summary". Online available : https://www.bmvit.gv.at/en/service/publications/downloads/action_automated_driving_2016-2018.pdf

BMVIT (2019). "Austrian Action Programme on Automated Mobility". Online available from: https://www.bmvit.gv.at/en/service/publications/downloads/action_automated_mobility_2019-2022_ua.pdf

Cai, M., Zhang, W.Y., Zhang, K. 2001. "ManuHub: A Semantic Web System for Ontology-Based Service Management, Distributed Manufacturing Environments." In IEEE Transactions on Systems, Man and Cybernetics, Part A: Systems and Humans, 41(3): 574– 582.

Chang, Xiaomeng, Rai, Rahul, Terpenny, Janis. 2010. "Development and Utilization of Ontologies in Design for Manufacturing", Journal on Mechanical Design, 132(2):021009-021009-12, 1-12.

Chungoora, Nitishal, Young, Robert I., Gunendran, Georg, Palmer, Claire, Usman, Zahid, Anjum, Najam A., Cutting-Decelle, Anne-Francoise, Harding, Jennifer A., Case, Keith. 2013. "A Model-Driven Ontology Approach for Manufacturing System Interoperability and Knowledge Sharing." Journal on Computers in Industry, 64 (4), 392-401.

Cordts, M., Omran, M., Ramos, S., Rehfeld, T., Enzweiler, M., Benenson, R., Franke, U., Roth, S., and Schiele, B. (2016). "The Cityscapes Dataset for Semantic Urban Scene Understanding," in Proc. of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR).

Cordts, M., Omran, M., Ramos, S., Rehfeld, T., Enzweiler, M., Benenson, R., Franke, U., Roth, S., and Schiele, B. (2015). "The Cityscapes Dataset," in CVPR Workshop on The Future of Datasets in Vision.

D'Ambros, M., Lanza, M. and Robbes, R. (2010). "An extensive comparison of bug prediction approaches," In Proceedings of the 7th IEEE Working Conference on Mining Software Repositories, pp. 31–41.

Damjanovic-Behrendt, V. (2018). "A Digital Twin-based Privacy Enhancement Mechanism for the Automotive Industry", 2018 International Conference on Intelligent Systems (IS): Theory, Research and Innovation in Applications. Funchal, Madeira, Portugal, 2018, pp. 272 – 279. DOI: 10.1109/IS.2018.8710526. Online available from: https://ieeexplore.ieee.org/document/8710526/

Diep, Daniel, Alexakos, Christos, Wagner Thomas. 2007. "An Ontology-Based Interoperability Framework for Distributed Manufacturing Control." In Proceedings of the IEEE Conference on Emerging Technologies and Factory Automation (ETFA), Patras, Greece, 855–862.

Do, H., Elbaum, S. and Rothermel, G., 2005. "Supporting controlled experimentation with testing techniques: An infrastructure and its potential impact," Empirical Software Engineering, vol. 10, no. 4, pp. 405–435, 2005.

Dollar, P., Wojek, C., Schiele, B., Perona, P. (2009). Pedestrian detection: A benchmark. In: Computer Vision and Pattern Recognition, 2009. IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 304–311.

ENISA (2016). Cyber Security and Resilience of Smart Cars: Good Practices and Recommendations.

Feld M., and Müller C.: The automotive ontology: managing knowledge inside the vehicle and sharing it between cars, 2011.

FNC-2018 (2018). Symposium on the Future Networked Car (FNC-2018), within the 88th Geneva International Motor Show, March 2018.

FPF (2018). Personal Data in Your Car. National Automobile Dealers Association and the Future of Privacy Forum. Online available: https://www.nada.org/personaldatainyourcar/

Fridman, L., Brown, D.E., Angell, W., Abdic, I., Reimer, B., Noh, H.Y. (2016). "Automated Synchronization of Driving Data Using Vibration and Steering Events". Pattern Recognition Letters. Online: https://arxiv.org/pdf/1510.06113.pdf

Garetti, Marco, and Fumagalli, Luca. 2012. "P-PSO Ontology for Manufacturing Systems." INCOM 2012 – Information Control in Manufacturing conference, Vol 14, No. 1, 449-456. 10.3182/20120523-3-RO-2023.00222.

Garetti, Marco, and Fumagalli, Luca. 2012a. "Role of Ontologies in Open Automation of Manufacturing Systems." In Proceedings of the 17th Summer School in Industrial Mechanical Plants, Venice, Italy.

Garetti, Marco, Fumagalli, Luca, Lobov, Andrei, Martinez Lastra, Jose. 2013. "Open Automation of Manufacturing Systems Through Integration of Ontology and Web Services." In Proceedings of the 7th IFAC Conference on Manufacturing Modelling, Management, and Control. 10.3182/20130619-3-RU-3018.00169

Garetti, Marco, Fumagalli, Luca, Negri, Elisa. 2015. "Role of Ontologies for CPS Implementation in Manufacturing." In Management and Production Engineering Review, 6(4). doi: 10.1515/mper-2015-0033

Geiger, A., Lenz, P., Stiller, C. and Urtasun, R. (2013) Vision meets robotics: The kitti dataset. The International Journal of Robotics Research 32(11): 1231–1237.

Herraiz, I., Izquierdo-Cortazar, D. and Rivas-Hernandez, F. (2009). "Floss-metrics: Free/libre/open source software metrics," In Proceedings of the 13th IEEE European Conference on Software Maintenance and Reengineering (CSMR'09), pp. 281–284.

Huang, X., Cheng, X., Geng, Q., Cao, B., Zhou, D., Wang, P., Lin, Y. and Yang, R. (2018) The apolloscape dataset for autonomous driving. Arxiv.

IEC 62890 (2016). International Electrotechnical Commission: IEC 62890: Life-cycle management for systems and products used in industrial-process measurement, control and automation.

Jeong, J., Cho, Y. and Kim, A. (2019-1). The Road Is Enough! Extrinsic Calibration of Non-Overlapping Stereo Camera and LiDAR Using Road Information. IEEE Robotics and Automation Letters (RA-L).

Jeong, J., Cho, Y., Shin, Y., Roh, H., and Kim, A. (2019). Complex Urban Dataset with Multi-level Sensors from Highly Diverse Urban Environments, IJRR 2019.

Jus, R. Jalali, D. and Ernst, M.D. (2014). "Defects4j: A database of existing faults to enable controlled testing studies for java programs," In Defects4J: A database of existing faults to enable controlled testing studies for Java programs, San Jose: ACM, Jun. 2014, pp. 437–440, isbn: 978- 1-4503-2645-2. doi: http://dx.doi.org/10.1145/2610384.2628055. Online available from: http://defects4j.org.

Kamei, Y. Monden, A. Morisaki, S. and Matsumoto, K.-i. (2008). "A hybrid faulty module prediction using association rule mining and logistic regression analysis," In Proceedings of the Second ACM-IEEE International Symposium on Empirical Software Engineering and Measuremen (ESEM '08), New York, NY, USA: ACM, pp. 279–281, Doi: 10.1145/1414004.1414051. Online available from: http://doi.acm.org/10.1145/1414004.1414051

Kiritsis, Dimitris, Kadiri, Soumaya, Perdikakis, Apostolos, Milicic, Ana, Alexandrou, Dimitris, Pardalis, Kostas, 2013. "Design of Fundamental Ontology for Manufacturing Product Lifecycle Applications." In: Emmanouilidis C., Taisch M., Kiritsis D. (Eds) Advances in Production Management Systems. Competitive Manufacturing for Innovative Products and Services. APMS 2012. IFIP Advances in Information and Communication Technology, Vol. 397. Springer, Berlin, Heidelberg.

Klotz, B., Troncy, R., Wilms, D., Bonnet, C. (2018). VSSo: A Vehicle Signal and Attribute Ontology. In Proceedings of the 9th International Semantic Sensor Networks Workshop, Monterey, CA, USA.

Lemaignan, Severin, Siadat, Ali, Jean-Yves, Dantan, and Semenenko, Anatoli. 2006. "MASON: A Proposal for An Ontology Of Manufacturing Domain." IEEE Workshop on Distributed Intelligent Systems: Collective Intelligence and Its Applications (DIS 2006), Prague, Czech Republic, 195–200.

Liu, Li, Yao, Xiaofen, Qin, Liangjuan, Zhang, Miao. 2014. "Ontology-Based Service Matching in Cloud Computing." In Proceedings of the IEEE International Conference on Fuzzy Systems (FUZZ- IEEE), pp. 2544-2550. 10.1109/FUZZ-IEEE.2014.6891698

Maddern, W., Pascoe, G., Linegar, C., and Newman, P. (2016). "1 Year, 1000km: The Oxford RobotCar Dataset", *The International Journal of Robotics Research (IJRR)*,.

Menzel, Christopher, and Grüninger, Michael. 2001. "A Formal Foundation for Process Modeling." In Proceedings of the Second International Conference on Formal Ontologies in Information Systems, Welty and Smith (Eds.), 256-269.

Menzies, T.J., Krishna, R. and Pryor, D., 2015. The promise repository of empirical software engineering data. North Carolina State University. Online available from: http://openscience.us/repo

Mockus, A. (2009). "Amassing and indexing a large sample of version control systems: Towards the census of public source code history.," In MSR, Vol. 9, pp. 11–20.

Mohammed, Wael M., Ramis Ferrer, Borja, Iarovyi, Sergii, Negri, Elisa, Fumagalli, Luca, Lobov, Andrei, Martinez Lastra, Jose. 2017. "Generic Platform for Manufacturing Execution System Functions in Knowledge-Driven Manufacturing Systems." International Journal of Computer Integrated Manufacturing, 262-274.

NASA, 2004. Metrics data program data repository. Online available from: http://mdp.ivv.nasa.gov.

Negri, Elisa, Fumagalli, Luca, and Macchi, Marco. 2017a. "A Review of the Roles of Digital Twin in CPS-based Production Systems." In Proceedings of the 27th Int. Conf. on Flexible Automation and Intelligent Manufacturing (FAIM2017), Italy. Vol 11, 939-948.

Negri, Elisa, Fumagalli, Luca, Garetti, Marco, Tanca, Letizia, 2015. "Requirements and Languages for the Semantic Representation of Manufacturing Systems." Computers in Industry. Vol. 81, Issue C, 55-66.

Negri, Elisa, Fumagalli, Luca, Macchi, Marco, Garetti, Marco. 2015a. "Ontology for Service-Based Control of Production Systems." In S. Umeda, M. Nakano, H. Mizuyama, H. Hibino, D. Kiritsis, G. von Cieminski (Eds.) IFIP Int. Conf. on Advances in Production Management Systems (APMS), Japan, 484-492.

Negri, Elisa, Perotti, Sara, Fumagalli, Luca, Marchet, Gino, Garetti, Marco. 2017. "Modelling Internal Logistics Systems Through Ontologies." Journal on Computers in Industry, Vol. 88, Issue C, 19-34.

Pandey, G., McBride, J.R. and Eustice, R.M. (2011) Ford campus vision and LiDAR data set. International Journal of Robotics Research 30(13): 1543–1552

Peynot, T., Scheding, S., and Terho, S. (2010). The Marulan Datasets: Multi-sensor Perception for Unmanned Ground Vehicles (UGV)*International Journal of Robotics Research (IJRR)*, November 2010, Vol. 29, No. 13, pp. 1602-1607.

Peynot, T., Terho, S., and Scheding, S. (2009). *Sensor Data Integrity: Multi-Sensor Perception for Unmanned Ground Vehicles.* Australian Centre for Field Robotics (ACFR), The University of Sydney, 2009. Technical Report ACFR-TR-2009-002 Online available from: http://sdi.acfr.usyd.edu.au/ACFR-TR-2009-002.pdf

RAMI (2016). DIN SPEC: 91345: 2016-04. Reference Architecture Model Industrie 4.0

Rodríguez-García, Miguel Ángel, Valencia-García, Rafael, García-Sánchez, Francisco, Samper-Zapater, J. Javier. 2014. “Ontology-Based Annotation and Retrieval of Services in the Cloud.” Knowledge-Based Systems. 56, C, pp. 15-25. DOI=http://dx.doi.org/10.1016/j.knosys.2013.10.006

Ruta, M., Scioscia, F., Loseto, G., Pinto, A., Di Sciascio, E. (2018) Machine learning in the Internet of Things: A semantic-enhanced approach. Semantic Web Journal, Volume 10, No. 1, pp. 183--204.

Sayyad S.,J. and Menzies, T.J. (2005) The PROMISE Repository of Software Engineering Databases. School of Information Technology and Engineering, University of Ottawa, Canada. Available: http://promise.site.uottawa.ca/SERepository

Schlenoff, Craig, Grüninger, Michael, Ciocoiu, Mihai, Lee, Jintae. 2000. “The Essence of the Process Specification Language.” Transactions of the Society for Computer Simulation, 16(4), Feb 2000, 204-216.

Śliwerski, J., Zimmermann, T., and Zeller, A. (2005). When Do Changes Induce Fixes?. In Proc. of the 2005 International Workshop on Mining Software Repositories, Vol. 30. 1–5.

Smith, M., Baldwin, I., Churchill, W., Paul, R. and Newman, P. (2009). The new college vision and laser data set. International Journal of Robotics Research 28(5): 595–599.

Stocker, A., Kaiser, C., & Festl, A. (2017). Automotive Sensor Data. An Example Dataset from the AEGIS Big Data Project [Data set]. Zenodo. http://doi.org/10.5281/zenodo.820576

von Birgelen, A., Buratti, D., Mager, J., Niggemann, O. (2018) Self-Organizing Maps for Anomaly Localization and Predictive Maintenance in Cyber-Physical Production Systems. In: 51st CIRP Conference on Manufacturing Systems (CIRP CMS 2018) CIRP-CMS. Paper available open access: https://authors.elsevier.com/sd/article/S221282711830307X

Warren, M., McKinnon, D., He, H., and Upcroft, B. (2010). Unaided stereo vision based pose estimation. In Proceedings of the Australasian Conference on Robotics and Automation (Eds. Wyeth, Gordon and Upcroft, Ben). http://eprints.qut.edu.au/39881/

Xia, Y., Zhang, D., Kim, J., Nakayama, K., Zipser, K., and Whitney, D. (2018). "Predicting Driver Attention in Critical Situations" ACCV

Xu, H., Gao, Y., Yu, F., and Darrell, T. (2017). "End-to-end learning of driving models from large-scale video datasets." CVPR 2017

Yu, F., Xian, W., Chen, Y., Liu, F., Liao, M., Madhavan, V., & Darrell, T. (2018). BDD100K: A Diverse Driving Video Database with Scalable Annotation Tooling. CoRR, abs/1805.04687.

Zhang, Miranda, Ranjan, Rajiv, Haller, Armin, Georgakopoulos, Dimitrios, Menzel, Michael, Nepal, Surya. 2012. “An Ontology-Based System for Cloud Infrastructure Services Discovery.” In Proc. of the 8th Int. Conf. on Collaborative Computing: Networking, Applications and Worksharing, pp. 524-530.

Zhang, S., Benenson, R., Schiele, B.: Citypersons, 2017. A diverse dataset for pedestrian detection. In The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) 2017.

Zimmermann, T. Premraj, R. and Zeller, A. (2007). “Predicting defects for Eclipse,” In Proceedings of the IEEE International Workshop on Predictor Models in Software Engineering, 2007. PROMISE’07, pp. 9–9.