IoT4CPS – Trustworthy IoT for CPS

FFG - ICT of the Future
Project No. 863129

Deliverable D5.3
Cross-Platform Interoperation Model

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: Cross-Platform Interoperation Model
Type: public
Editor(s): Violeta Damjanovic-Behrendt
E-mail: violeta.damjanovic@salzburgresearch.at
Author(s): Violeta Damjanovic-Behrendt
Doc ID: D5.3

Amendment History

Version	Date	Author	Description/Comments
V0.1	01.04.2019	Violeta Damjanovic-Behrendt	Initial version prepared
V0.2	15.06.2019	Violeta Damjanovic-Behrendt	Document structure and finalization of the first section
V0.3	15.09.2019	Violeta Damjanovic-Behrendt	Data interoperation standards
V0.4	20.11.2019	Violeta Damjanovic-Behrendt	Introduction of the IDS and the European Common Data Spaces initiatives
V0.5	26.11.2019	Violeta Damjanovic-Behrendt	Document sent for the QA
V0.6	27.11.2019		Review by Mario Drobics, AIT
V0.7	29.11.2019		Review by Christos Thomos, Infineon
V0.8	10.12.2019	Violeta Damjanovic-Behrendt	Improvements according to the comments received from the reviewers
V1.0	12.12.2019	Violeta Damjanovic-Behrendt	Final document

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

Abbreviations

ADS	Automated Driving Systems
AI	Artificial Intelligence
Auto-ISAC	Automotive Information Sharing and Analysis Center
CAD	Computer Aided Design
CAM	Connected and Automated Mobility
CCAM	Cooperative, Connected and Automated Mobility
CIM	Computer Integrated Manufacturing
CIM	Cloud Information Model
CISDA	Computational Intelligence for Security and Defence Applications
CPS	Cyber Physical Systems
CTI	Cyber Threat Information
DEX	Data Exchange Specification
DG CONNECT	Directorate-General for Communication Networks, Content and Technology
DG GROW	Directorate-General for Internal Market, Industry, Entrepreneurship and SMEs
DG MOVE	Directorate-General for Mobility and Transport
ENISA	European Union Agency for Cybersecurity
GDPR	General Data Protection Regulation
HIDS	Host-based Intrusion Detection System
IDM	IDentity Management
IDP	IDentity Provider
IDS	International Data Spaces
IDSA	IDS Association
IoT	Internet of Things
MSP	Multi-Sided Platform
NHTSA	National Highway Traffic Safety Administration
ODI	Open Data Initiative
ODP	Open Data Portal
OSLC	Open Services for Lifecycle Collaboration
OWL	Web Ontology Language
PLCDM	Product LifeCycle Data Management
PLCS	Product Life Cycle Support
RDF	Resource Definition Framework
RDFS	RDF Schema
RDL	Reference Data Library
SFP	Software Fault Prediction
SHACL	Shape Constraint Language
SME	Small and Medium-sized Enterprises
SOSA	Sensor-Observation-Sampling-Actuation
SSNO	Semantic Sensor Networks Ontology
STEP	Standard for the Exchange of Product Model Data
UNECE	United Nations Economic Commission for Europe
VSSO	Vehicle Signal and Attribute Ontology
WDI	Wireless Data Interface

Executive Summary

The objectives of this report are to investigate IoT platform interoperability landscape and its relevance to the field of Connected and Automated Mobility (CAM) and to model cross-platform interoperation at the data level. CAM applications bring along unique opportunities, but also risk exposures. For example, sharing data with many parties in the vehicle ecosystem can improve a variety of driving features, e.g. it can increase driving comfort and safety, improve driving experience, optimize lifecycle processes related to manufacturing and the use of vehicles, contribute to societal targets related to sustainable manufacturing, reduction of fuel consumption, road safety, and more. At the same time, sharing vehicle data can compromise the privacy of stakeholders, e.g. through exposure of their behavioural patterns. Thus, the objective of this report is to raise awareness of good data practices in the CAM industry, and to identify challenges and possible solutions for ensuring the quality of data interoperation among future vehicle infrastructures.

1. Introduction

Enabling connectivity of vehicles requires many challenges in the Connected and Automated Mobility (CAM) sector to be solved (FNC-2018, 2018); for example:

- Safe navigation of connected vehicles requires that road infrastructure is able to transmit information about traffic, weather and road works to the cloud. It further requires holistic infrastructure policies to be put in place to enable transparency of data sharing and decision making (e.g. driving and navigational decisions);
- Data sharing and data transparency are both required to ensure continuous safety conditions in the ecosystem of connected vehicles;
- Network connectivity needs to be faster and wider, and some major service providers are expected to provide mainstream 5G services by 2020;
- International standards for vehicle communication and connectivity need to be harmonized. For example, international bodies such as the UNECE (United Nations Economic Commission for Europe) and ITU, are currently working with global automotive players towards harmonization of vehicle standards (i.e., 802.11p, 5G, DSRC);
- Cybersecurity challenges for connected vehicles need to be addressed by international bodies. For example, UNECE and ENISA are working on cybersecurity challenges in this sector (see ENISA, 2016). In November 2019, ENISA published the "Good Practices for the Security of Smart Cars" report that points at the prospective challenges and risks of cyberattacks which may face owners and users of connected and smart' vehicles. With the growth of the connected, autonomous and smart vehicle technologies, the development of comprehensive system regulations to guard against hacks is becoming increasingly important (see ENISA-SC, 2019). The latest ENISA's report aims to serve as a reference point for promoting cybersecurity for smart cars across Europe and raising awareness on relevant threats and risks that can result in remote mobilisation of affected vehicles, road accidents, the loss or theft of sensitive user data, financial losses and potential danger to the safety of other drivers and road users.

The objective of this report is to capture cross-platform interoperation models for connected vehicles, which may co-exist with many external IoT infrastructures and stakeholders. The envisioned cooperation among those infrastructures can include:

- Connected vehicle being observed by external infrastructure (e.g. CCTV, car park systems, smart city, etc.);
- Connected vehicle interpreting information received from external infrastructures (e.g. smart city traffic management system);
- Connected vehicle sending information to other connected vehicles (which may directly affect the actions/ decisions of the other connected vehicle);
- Connected vehicle interpreting information received from other connected vehicles;
- Third party use of data trails (e.g. braking/acceleration data brought by insurance companies), and many more.

1.1 Relation to Other Tasks in IoT4CPS

This report is based on the IoT4CPS **D5.2 "Product Lifecycle Data Management (PLCDM) Stakeholder Perspectives"** that captures multi-tenancy aspects related to connected vehicles and emerging standards for data and information exchange in the Automotive industry. The D5.2 is based on the IoT4CPS **D2.2 "Business needs consolidation – competitive intelligence"** and a specific project's use case that is described in D2.2 **Section 3.2 "AVL: Security Verification Along the Full Life Cycle of IoT-based Industrial Instrumentation Systems"**. The D5.2 creates a basis for further investigation of the concept of "digital twinning" in WP5, considering real-world situations and processes related to various lifecycle phases in the Automotive industry (see the IoT4CPS **D5.5.1 "Lifecycle Data Management Prototype I"** and **D5.5.2 "Lifecycle Data Management Prototype II"**).

The data model presented in D5.2 is further improved by adding identity, security, privacy and safety aspects, which is afterwards described in the **D5.4.1 "Identity, Security and Safety in Product Lifecycle Data Management"** report. The D5.4.1 report also points the readers at the task T7.2 on traceability through lifecycle phases, task T7.3 that is about a taken security by isolation approach, and task T6.1 is about secure and safe platform design for CAM applications.

1.2 Document Organisation

Section 2 identifies assets, stakeholders and privacy concerns relevant to the CAM applications. Section 2 emphasizes a wide scope of the landscape of the pervasive digital tracking and user's profiling through many industries, including the Automotive industry. The overall understanding of the scope of possible security risks and a privacy leakage is important when modelling security and privacy features of data models of a Digital Twin prototype. Section 3 presents the method applied to design a multi-stakeholder- and cybersecurity-centered data model in IoT4CPS. In Section 4, we discuss cross-platform interoperability of the proposed data model, by considering the following two approaches: (1) knowledge and data models and initiatives based on the Semantic Web (e.g. OASIS Open Service for Lifecycle Collaboration (OSLC), ISO 15926, ISO 10330-239, the Cloud Information Model (CIM), the European Union Open data Portal (EU ODP) and several relevant ontologies), and (2) other European initiatives for data sharing and data interoperation (e.g. International Data Spaces (IDS), the upcoming Gaia-X project, the Data Market Austria (DMA) and European Common Data Spaces). Section 5 concludes the report.

2. Assets and Stakeholders in Connected and Automated Mobility (CAM) Applications

The benefits of IoT data lifecycle observations and management are many, e.g. increase in productivity and quality of manufacturing outputs through configurations, optimizations and control automation; improved throughput, sustainability and yield rates of heavy machinery; self-healing features that automatically enable necessary maintenance and repair activities, and allow for modification of processes along a product lifecycle; improved security against threats, etc. CAM applications are designed with multi-tenancy in mind from the get-go, in order to provide benefits for various stakeholders, e.g. smart vehicle owners, vehicle drivers, passengers sharing the vehicle, applications of smart cities and smart roads, insurance companies, environmental and climate change organisations, car sharing dealers, and more. At the same time, vehicles are picking up data about our driving style, patterns and habits, detailed driving histories, etc. Sharing such data can compromise the privacy of stakeholders involved in the vehicle's ecosystem. Thus, in this section we discuss relevant assets and stakeholders included in CAM applications, and look at what kind of data associated to both assets and stakeholders is being collected and how it's being used. We also look in this section, at some emerging regulations on smart and connected vehicle's data privacy and steps to mitigate possible risks.

Figure 1 illustrates our research approach taken in WP5 "IoT Lifecycle Management" of the IoT4CPS project, that illustrates the IoT product lifecycle phases of a Smart Automotive System. The process starts with the collaborative Computer Aided Design (CAD) applications that enable machines to access the original CAD drawings to get the design and assembling information electronically and further initiate the manufacturing processes through Computer Integrated Manufacturing (CIM). After the vehicle is manufactured, it become available at the market and sold to the user (a connected car owner, in Figure 1). The car owner parks a new connected car in a garage of a smart home, and through Bluetooth, it connects with other smart home gadgets. The connected car owner drives the vehicle to the smart office' garage and connects with other available office and garage gadgets and other connected cars (e.g. through V2I (Vehicle to Infrastructure)). The car owner visits friends at their smart homes, without being aware of their installed smart home devices that listen to all conversations and record audio and visual observations, e.g. Amazon Alexa, Amazon Echo, Apple HomePod, Google Home, Facebook's Portal, smart refrigerators and many other connected smart toys that typically contains sensors, microphones, cameras, speech recognition, face recognition, data storage components, etc. The connected car owner further drives the vehicle through smart cities, passing many ALPRs (Automatic Licence Plate Readers) cameras that can automatically identify and record licence plate numbers on passing cars. ALPRs can be operated not only by smart cities and installed on police cars. They can also be operated by private companies to amass vast quantities of location data and sell this data to third parties (Cyphers & Gebhart, 2019). The connected car owner performs periodic maintenance services and at some point, decides to sell the vehicle. The new connected car owner experiences the similar connectivity path, from driving the car to his/her own smart home, to smart office(s), smart homes of friends, smart cities and infrastructures (roads and motorways information, traffic lights, speed limits), etc. Finally, the connected car comes to its "end of life" and needs to be disposed of properly. The data of all car owners, passengers, and many smart and connected gadgets related to the vehicle's lifecycle are stored and available in the cloud. The data could be removed from devices, but cloud data recovery and backup procedures promise effective and automated data replications, even across clouds. Hence, our research approach in IoT4CPS WP5 is to consider two perspectives of IoT lifecycle data management: multi-stakeholder and cybersecurity (see Figure 1) in order to provide guidelines on what need to be accounted for to enable "digital twinning" of cybersecurity and privacy features for future CAM applications.

In order to enable data acquisition, exchange and processing in IoT4CPS, we adopt the concept of a Digital Twin to model the data management architecture and employ data analytics to proactively address data privacy, security and safety of CAM applications (see IoT4CPS D5.5.1 for details). Figure 1 illustrates a conceptual view of the adopted concepts of a Digital Twin, its authorization boundaries and CPSs composed of subsystems (systems of systems) along the entire lifecycle. CPSs need to operate in highly dynamic cloud environments, and when one

component changes its behaviour or breaks down (e.g. due network problems or security attacks), the system should expose "smart" behaviour by recognizing the faulty situations and returning to its normal processing with minimum damage (aka "graceful degradation"). Thus, the Digital Twin in IoT4CPS need to provide both monitoring capabilities of the connected vehicle's ecosystem, and reasoning over a diverse body of knowledge and data, in order to enable adaptive behaviour to overcome security and safety critical situations.

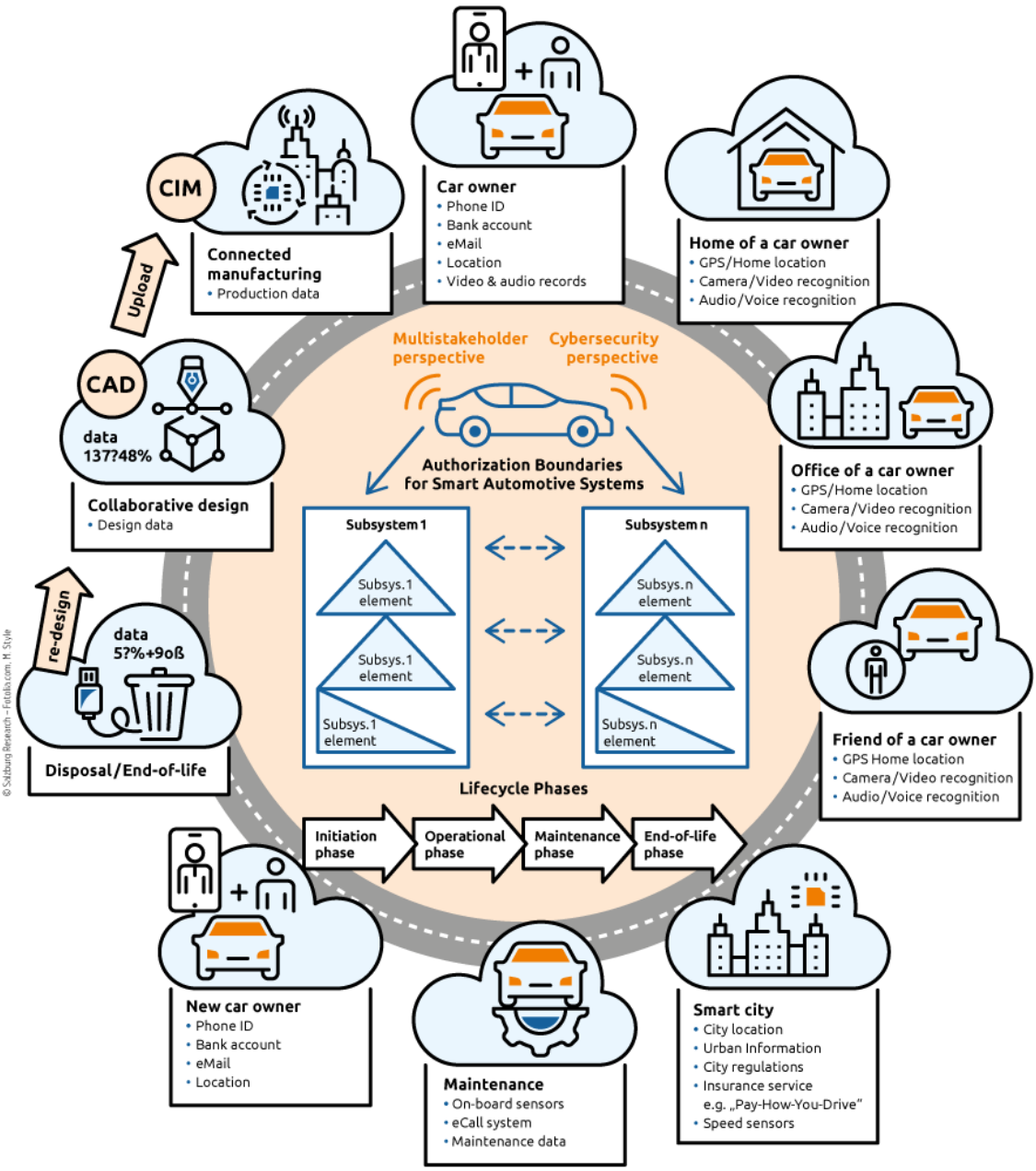

Figure 1 – Multi-Stakeholders along PLCDM of a Connected Vehicle

The **D5.2 "Product Lifecycle Data Management (PLCDM) Stakeholder Perspectives"** presents two use cases that are motivated by use cases described in "Austrian Action Programme on Automated Mobility" (BMVIT, 2019)

and extended by adding the AVL's Device.CONNECT™ that collects data about the road and environmental conditions, e.g. air pollution, temperature near the surface of the road, humidity. The following two subsections briefly present these use cases that we have been used as basis for our analysis in this report.

2.1 Use Case 1: Safety & Cybersecurity+ through the Lifecycle Phases

The use case "Safety+ through an all-round view" (see BMVIT (2019)) describes 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 vehicle's surrounding.

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 2). These data are neither about the initiation phase (design, engineering, production) nor maintenance of the connected car nor its end-of-life.

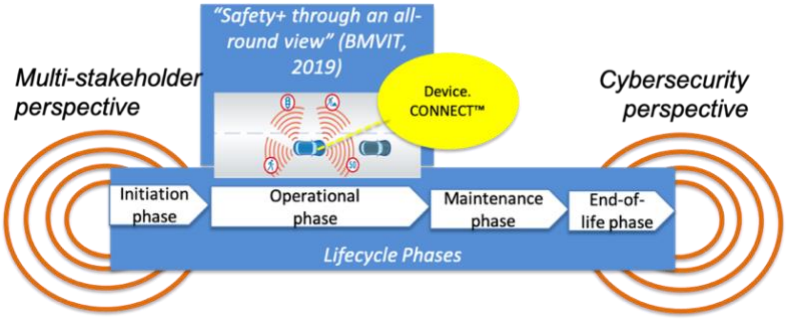

Figure 2 – The enrichment of the "Safety+ Through an All-Round View" use case in IoT4CPS

2.2 Use Case 2: Assistive Intelligence+ through the Lifecycle Phases

The use case "New flexibility" (see BMVIT (2019)) is about automated vehicles that offer 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 positively affect 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™ 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. By observing the light signals, the connected car can infer which car will turn and which other cars will continue straight. This

type of scenarios shows a strong potential to eliminate traffic fatalities in the future. A case of privacy misuse of combined data coming 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 phases and verifies the system overall safety and cybersecurity conditions (see Figure 3).

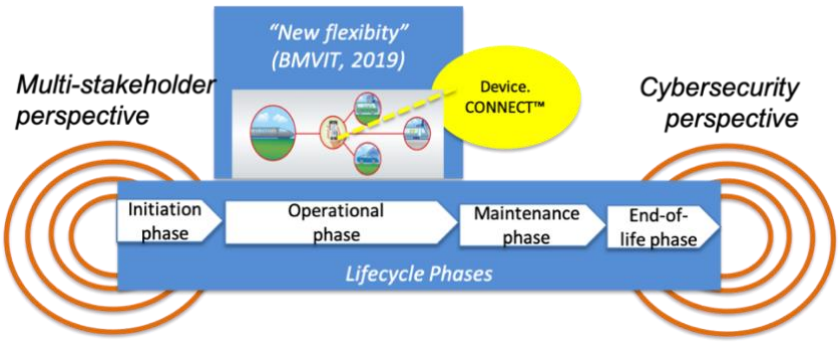

Figure 3 – The enrichment of the "New Flexibity+" use case in IoT4CPS

2.3 Identification of Assets, Stakeholders and Privacy Concerns

In D5.4.1, for each of the two use cases (see sections 2.1 and 2.2), we identified relevant and potential devices (assets) and users (stakeholders). Furthermore, based on the threat model presented in **D4.1 "Automotive Ethernet Protection Profile"**, we assigned relevant identity, security, safety and privacy risks to both assets and stakeholders of the connected vehicles (see IoT4CPS D5.4.1 for more details). The identified assets include: CAD Uploader, Collaborative Analysis Checker, CAD/CIM Initiator, Robotic Assembly Checker, Supply Chain Status Control, AVL's Device.CONNECT™, Powertrain Control, Chasses Control, Infotainment Control, External Media, CIM Remote Monitoring Service, Integrity Monitoring Service, Privacy Data Monitoring Service, Other Data Monitoring Service. The identified stakeholders include the following list: Manufacturer, Supplier, Aftermarket Supplier, Driver, Passenger, Cross-Collaborative Services, Powertrain Services, Smart City Services, Road Services, Insurance Services, Energy/Fuel Services, Testing & Certification Services, Network Connectivity Providers & Services, Marketing Services, and more. Note that a similar methodology is presented in the latest ENISA's report "ENISA Threat Landscape for 5G Networks. Threat assessment for the fifth generation of mobile telecommunications networks (5G)" published in November 2019 (ENISA-5G, 2019). ENISA's methodology includes the following steps: (i) assets identification, (ii) identification of threats in a form of threat taxonomy, (iii) mapping of risk scenarios to cybersecurity threats and (iv) mapping of stakeholders to previously identified assets.

Many stakeholders within a CAM ecosystem collect and process data for different purpose, e.g. car manufacturers use data to improve manufacturing processes and provide better functionality of their services; the infotainment manufacturers use customer data to personalize their recommendation service, e.g. music, GPS, etc.; third parties collect data through mobile phones and device connections (dongles) in order to e.g. resale data to advertisers and marketers, or to intercept financial details being transferred; hackers steal data to make a profit or to leak data to the public, etc. The most valuable types of personal data that are observed online include the following:

- Payment information – of interest primarily for someone's financial gain;
- Identification information – of interest for e.g. hackers to access to one or many accounts of a stakeholder;

- Personal medical records – of interest for hackers to gain access to healthcare services for themselves, or to resale information to other stakeholders, e.g. insurance companies;
- Classified information – for blackmailing purposes and spam; etc.

ALPRs (Automated License Plate Readers) cameras detect and read license plates, but they can also use other characteristics of cars, like make, model, colour, and wear, in order to help identify them. ALPRs are often used by law enforcement, but also by private companies that re-sell collected vehicle data local police, federal immigration enforcement agencies, private data aggregators, insurance companies, lenders, or bounty hunters (Cyphers & Gebhart, 2019)(McKinsey&Company, 2016). For example, the start-up Flock Safety offers ALPR-based "neighbourhood watch" services. Another example is the company Vigilant Solutions that uses ALPR technology to collect data from thousands of sources in the US into a single database, which it calls "PlateSearch." Many third parties, e.g. law enforcement agencies pay for access to PlateSearch. Apart cameras in CAM applications and smart city infrastructures, **passive MAC address tracking** is also used to track vehicle movement. Phones inside of vehicles, and the vehicles themselves, broadcast probe requests (short radio signals constantly sent from the phone in order to find nearby Bluetooth devices and WiFi networks). These probe requests include device's unique MAC address. Wireless "beacons" can be placed strategically around roads in order to listen for those signals and MAC addresses for tracking the movement of devices over time (see Figure 4).

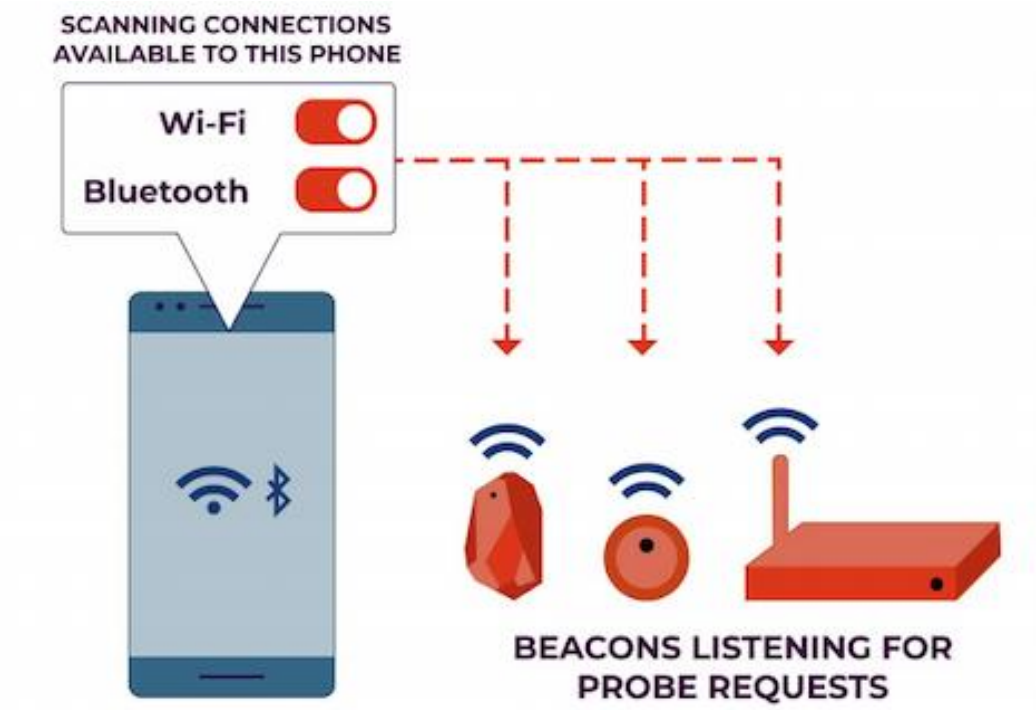

Figure 4 – Passive MAC Address Tracking Through Constant Scanning of WiFi and Bluetooth connections; source (Cyphers & Gebhart, 2019)

Figure 5 illustrates the landscape of the commercial pervasive digital tracking and user's profiling through business of many industries (CrackedLabs, 2017).

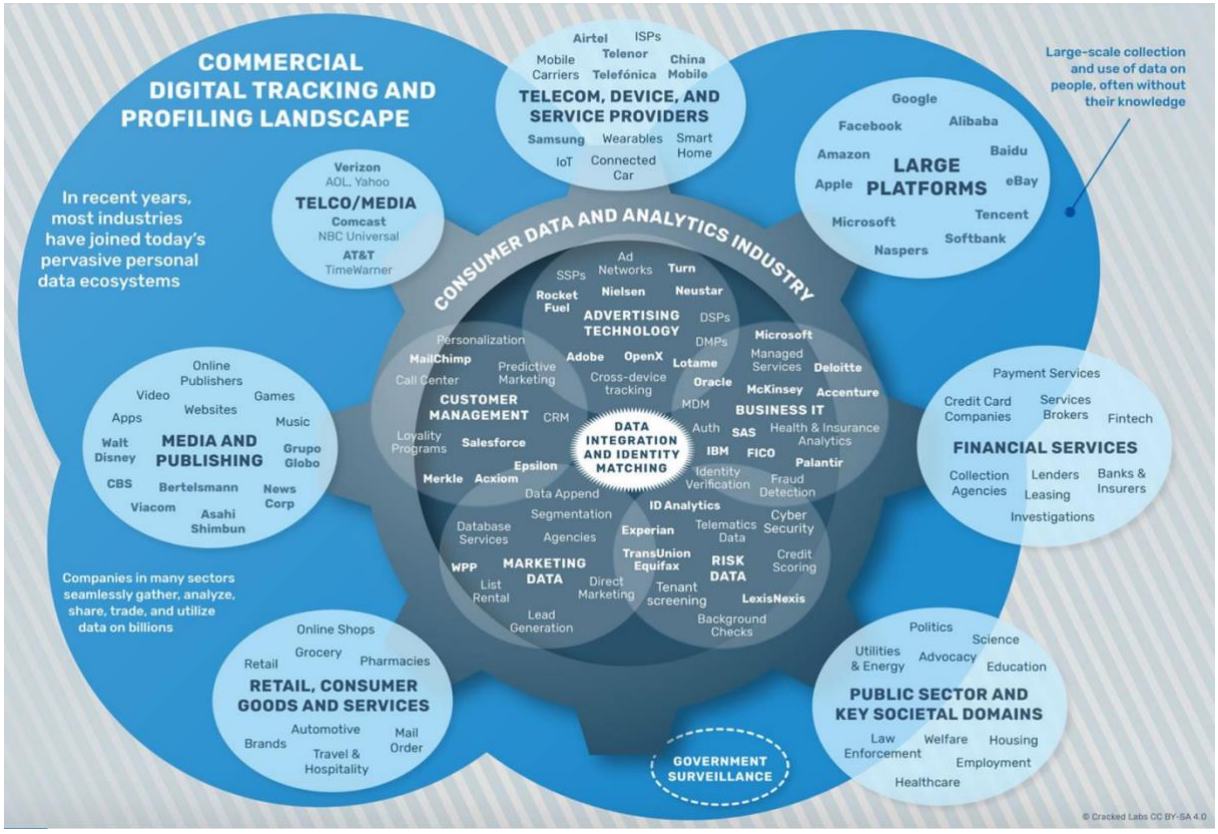

Figure 5 – Landscape of Surveillance Throughout Business Applications; source (CrackedLabs, 2017)

Today, many online fraud detection companies "evaluate" online transactions for clients in government, insurance, healthcare, etc. based on the analysis of digital behaviours, identities, and a range of IoT devices. Trustev is one of such companies. Trustev is based in Ireland, and uses a wide range of privacy data to assess users, e.g. device fingerprints, credit checks, social media content, and many more (see Figure 6).

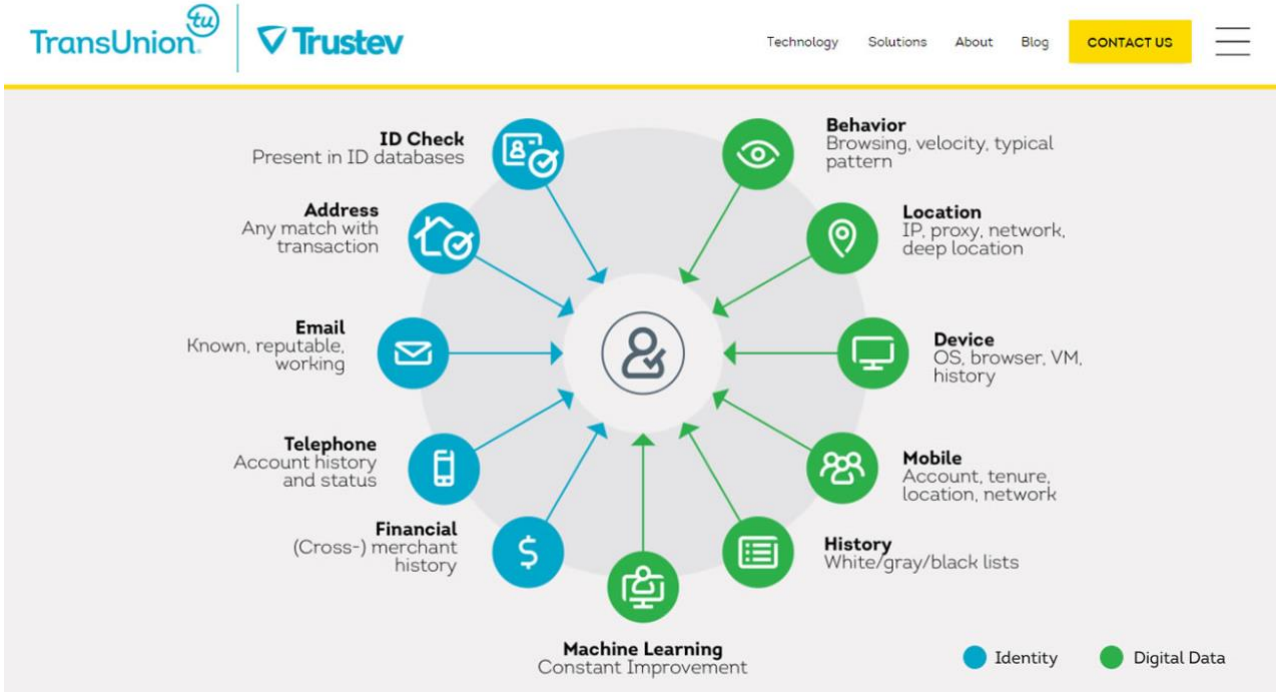

Figure 6 – Screenshot of Trustev's website, June 2, 2017 © Trustev; source (CrackedLabs, 2017)

Another example is a data broker based in San Francisco called People Data Labs (PDL). PDL claims on their website to have data on over 1.5 billion people for sale, e.g. 1+ billion personal email addresses, 420M+ LinkedIn URLs, 1+ billion Facebook URLs and IDs, 400M+ phone numbers, etc. The process of collecting data by PDL services is unclear, and is internally called “data enrichment”. The recent incident with PDL from November 2019, has been ironically titled "Data Enrichment Exposure From PDL Customer"; here, personal data for 1.2 billion people was discovered in an open-for-everyone Elasticsearch server (EOS, 2019). The PDL case is just the latest on a list of data breach discoveries. At the beginning of 2019, 2.2 billion records were found distributed on hacker forums across several tranches known as Collections #1-5, as reported in Wired (WIRED-data leak, 2019). In March 2019, an email marketing firm called Verifications.io left 809M records publicly accessible. In 2018, a breach of the sales intelligence firm Apollo exposed billions of data points.

3. Multi Stakeholder- and Cybersecurity-centered Data Model in IoT4CPS

The identified assets, stakeholders and their potentially associated identity, security, safety and privacy issues (risks) create a basis for the definition of multi stakeholder- and cybersecurity-centered data model in IoT4CPS. The data model presented in D5.4.1 assimilates diverse publicly available datasets and is used to enable cybersecurity and safety decision-making support for Digital Twin testbeds (see Figure 7).

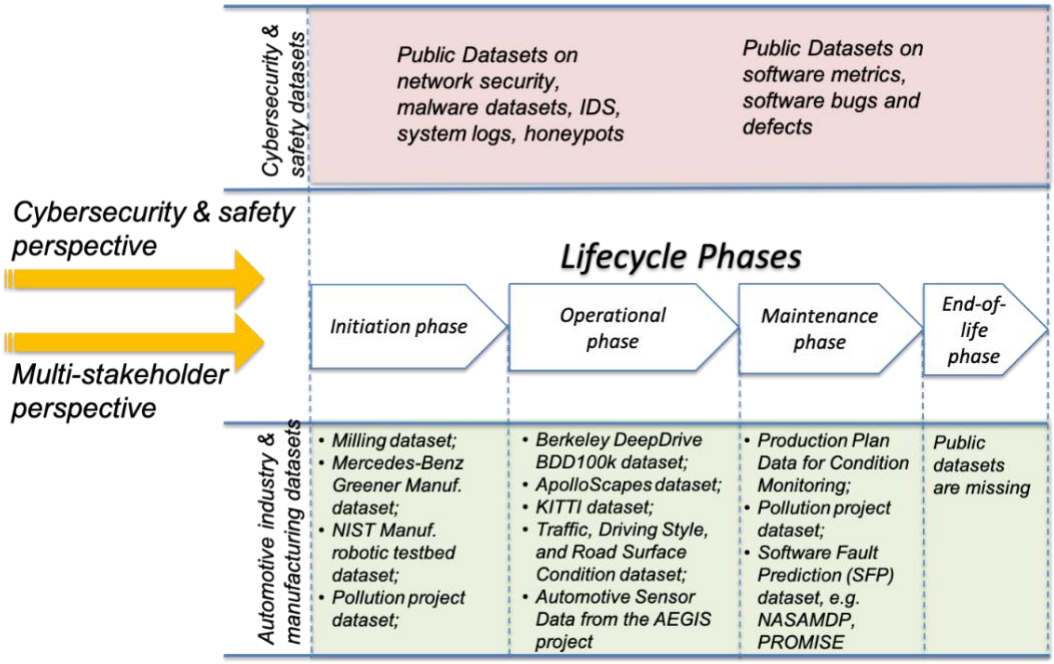

Figure 7 - Design Method for Multi-Stakeholder- and Cybersecurity-centered Data Model in IoT4CPS

The selected public datasets include the following:

- For the initiation phase, we selected: Milling dataset acquired from the experiments on a milling machine with different speeds, feeds, and depth of cut; Mercedes-Benz Greener Manufacturing dataset with an anonymized set of variables, each representing a custom feature in a Mercedes car; the NIST manufacturing robotics test bed that includes robotics data for advanced manufacturing and material handling; and Data for pollution project with value measurements of various pollutants: lead, carbon, nitrogen- dioxide, ozone, etc.;
- For the operational phase, we selected: Berkeley DeepDrive BDD100k that 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; the ApolloScapes that is a large dataset collected from 26 types of stakeholders, e.g. cars, bicycles, pedestrians, buildings, etc.; the KITTI dataset with online benchmarks for visual odometry, image tracking, and semantic segmentation; the Traffic, Driving Style and Road Surface Condition dataset that includes attributes for predicting road surface, traffic and driving style; and the Automotive Sensor Data collected during 35 trips conducted by one driver driving one vehicle.
- For the maintenance phase, we selected: Production Plan Data for Condition Monitoring that observes features of several components for which the predictions need to be performed; Data that contains measurement of various pollutants: lead, carbon, nitrogen- dioxide, ozone, etc.; and the industrial available dataset for Software Fault Prediction (SFP), e.g. NASAMDP (http://mpd.ivv.nasa.gov) and PROMISE (http://promise.site.uottawa.ca/SERepository/).
- For the end-of-life phase, public datasets are missing and will be generated for the purpose of research experimentations.

- For cybersecurity and safety validation, we select several datasets: ADFA Intrusion Detection Datasets collect data related to the Host-based Intrusion Detection System (HIDS) evaluation (Creech and Hu, 2013) (Creech, 2014); the Cyber Research Center Datasets (ITOC CDX) provides a comprehensive set of log data collected during attacks; the NSL-KDD Benchmark Dataset helps researchers to compare different intrusion detection methods (Tavallaee et al., 2009), and the DARPA Intrusion Detection Datasets include an offline evaluation and a real-time evaluation, based on a larger sample of training data about network-based attacks.

As discussed in D5.4.1, the above presented data model needs to be normalized in order to identify relationships, trends and anomalies related to security and safety vulnerabilities within the connected vehicle's ecosystem. At the same time, implementing the Digital Twin-based prototype in IoT4CPS, requires even more data to be included, e.g. data from external organizations and individuals would be required in order to properly address multi-tenancy aspects of the vehicle ecosystem. The focus of the following section is to look at the cross-platform interoperation aspects to support the IoT4CPS data model for more interoperation and wider use of data coming from different sources in the ecosystem.

4. Relevant Cross-Platform Data Interoperation Models

The IoT4CPS **D5.2 "Product Lifecycle Data Management (PLCDM) Stakeholder Perspectives"**, Section 5 presents several ontological data models for semantic enrichments of the IoT4CPS data model, e.g. SSNO (Semantic Sensor Networks Ontology), SOSA (Sensor-Observation-Sampling-Actuation Ontology) as a subset of the SSNO ontology, and a newly proposed ontology for formal description of automotive attributes, vehicle signals and sensors, e.g. VSSo (Vehicle Signal and Attribute Ontology) (Klotz et al., 2018). VSSo is based on the Vehicle Signal Specification (GENIVI) and uses SSN/SOSA pattern for the description of signals. In parallel, we consider the CTI (Cyber Threat Information) ontology and the recent V2I (Vehicle to Infrastructure) WDI (Wireless Data Interface) Ontology (Venkata et al., 2019).

Adopting the Semantic Web approaches in IoT4CPS is the one way to contribute to the cross-interaction of the connected vehicle system with the external stakeholders in the cloud. In parallel to Semantic Web ontologies, taxonomies, knowledge base and knowledge graphs, there is currently a strong push to different forms of distributed knowledge in Europe, e.g. through the use of the International Data Spaces (IDS) initiative from 2015, the Gaia-X Project, which is expected to develop the next generation of data infrastructure for Europe and will start in late 2020 (Gaia-X-pub1, 2019) (Gaia-X-pub2, 2019), and the European Common Data Spaces initiative that is created in April 2018 (EC Common European data Spaces, 2018).

The following two subsections addresses the most important principles and open data initiatives related to anticipated approaches to support cross-platform data interoperation models in IoT4CPS.

4.1 Data Interoperation Based on Semantic Web Standards and Knowledge Base

There exist several manufacturing data standards and initiatives for open data that are related to the use of the Semantic Web technologies.

4.1.1 OASIS Open Services for Lifecycle Collaboration (OSLC) Version 3.0

OSLC[1] is defined by IBM in 2008, and today is hosted by OASIS. OSLC aims to set specifications for the integration of software development, in order to make easier for tools to work together. The OSLC specification builds on top of RDF (Resource Definition Framework), Linked Data and REST, supporting integration at the data level through links created among related data resources. OSLC resources are defined as RDF properties, while OSLC operations on resources are performed using HTTP. OSLC specifies user interface techniques to enable preview, creation and selection of links.
The latest version of OSLC (Core 3.0) is based on W3C semantic standards, e.g. SHACL (Shape Constraint Language) (Knublauch and Ryman, 2019).

Relevance to IoT4CPS: OSLC provides standard adapters and connectors to a wide range of systems. It provides standard mechanisms to support data federation and tools- and application-integration workflows (for lifecycle collaborative applications) in the cloud. It is of interest to IoT4CPS to enable the data exchange for all RDF-based tools and applications.

4.1.2 ISO 15926: Industrial Automation Systems and Integration – Integration of Lifecycle Data for Process Plants Including Oil and Gas Production Facilities

The ISO 15926 is a standard for data integration, data sharing and data exchange between different computer systems. Its data model, initially formalized in EXPRESS and mapped in XML, is today formatted in OWL (Web Ontology Language). The community around the ISO 15926 provides a set of data libraries in OWL, through the Post Caeser Association (PostCaesar Association, 2019).

[1] OSLC Core 3.0 Specification: https://docs.oasis-open.org/oslc-core/oslc-core/v3.0/csprd03/part1-overview/oslc-core-v3.0-csprd03-part1-overview.html

Relevance to IoT4CPS: ISO 15926 is of interest to IoT4CPS as a data enabler and data integrator for OWL-based lifecycle data.

4.1.3 ISO 10303-239: Standard for the Exchange of Product Model Data (STEP)

The ISO 10303-239 PLCS (Product Life Cycle Support), known as STEP for short, specifies the information required to support a product throughout its lifecycle. The community around ISO 10303-239 PLCS identified the following elements (see Figure 8):

- A number of subsets of the ISO 10303-239 PLCS created to support the data exchange, called Data Exchange Specifications (DEXs). DEXs are built from reusable components (called Templates) created to ensure interoperability between DEXs.
- A set of Reference Data for the semantic enhancement of context of the ISO 10303-239 PLCS model, which may be further tailored to specific business needs. It is stored in a Reference Data Library (RDL) using OWL format (Reference Data Overview, 2019).
- A set of Model Usage Guide documents, which ensure the consistency in use of the standard.

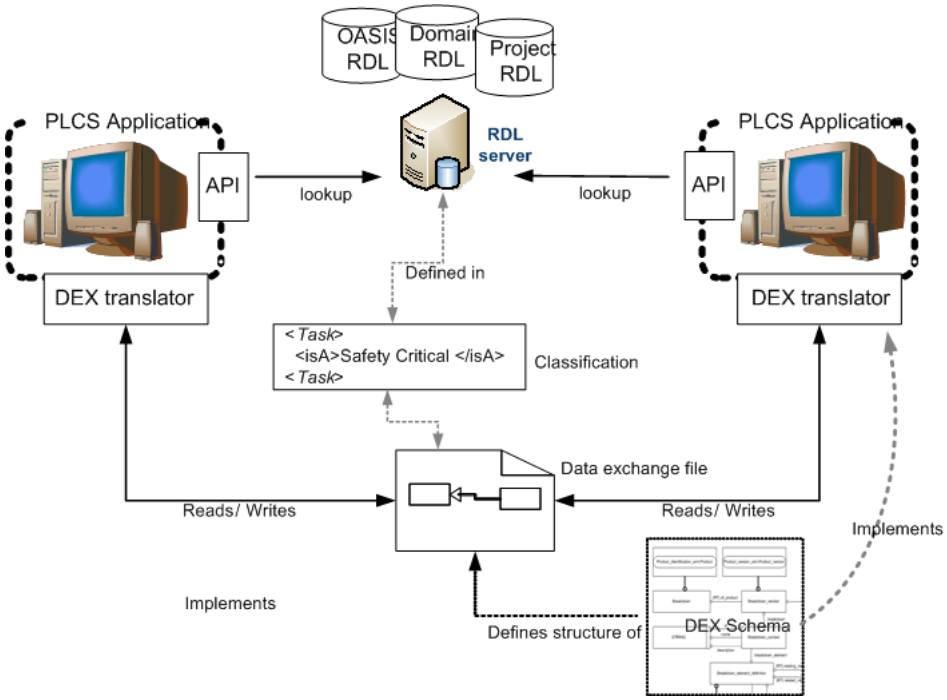

Figure 8 - ISO 10303-239 PLCS Data Exchange

Relevance to IoT4CPS: ISO 10303-239 (STEP) is an old industrial data exchange standard which is recently redesigned to support OWL data libraries. This allows further extensions of the existing industrial product data models towards cloud and IoT models.

4.1.4 Cloud Information Model (CIM)

The CIM data model supports connecting enterprise products through multiple cloud and on-premise applications based on own data models. The CIM is open sourced as part of the Joint Development Foundation (under the Linux Foundation). It is available through different standard languages: RDFS (RDF Schema), SHACL, R2RML, JSON-LD-based serialization, JSON Schema, etc. It is created as a response of Salesforce and Amazon to the Open Data Initiative (ODI) of Microsoft, SAP and Adobe.

Relevance to IoT4CPS: CIM is also relevant to IoT4CPS for its standardized and cloud-based data models used to connect enterprise products.

4.1.5 The European Union Open Data Portal (EU ODP)

The EU ODP is a catalogue of free to use and reuse data, for various purposes including commercial and non-commercial. The access to the free data is supported using RDF, SparQL and REST APIs (EU Open Data Portal, 2019). The data available through the EU ODP include: geographic, geopolitical and financial data, statistics data, election results, legal acts, data on crime, health, the environment, transport and scientific research.

Relevance to IoT4CPS: EU ODP is relevant to IoT4CPS for its offer of free-to-use data, contributing to the variety of datasets for data analytics of the Digital Twin prototype.

4.1.6 Relevant Ontologies – SSNO, SOSA, VSSo, V2I WDI

SSNO[2] (Semantic Sensor Networks Ontology) describs sensors and their observations, the involved procedures, the studied features of interest, the samples and the observed properties, as well as actuators. SOSA (Sensor-Observation-Sampling-Actuation Ontology) as a subset of the SSNO and presents a conceptualization of all entities, activities and properties that typically constitute a CPS. SOSA encompasses all of the three modelling perspectives: the activities of observing, sampling, and actuating [8].
VSSo (Vehicle Signal and Attribute Ontology) is a recent ontology for formal description of automotive attributes, vehicle signals and sensors. It is based on the Vehicle Signal Specification (GENIVI) and uses SSN/SOSA pattern for the description of signals (Klotz et al., 2018). We also consider recently designed V2I (Vehicle to Infrastructure) WDI (Wireless Data Interface) Ontology (Venkata et al., 2019) which is designed to support reasoning about new threats upon the addition of an unknown component to the system.

Relevance to IoT4CPS: Together, SSN and SOSA are able to support a wide range of applications and use cases, including satellite imagery, monitoring, industrial and household infrastructures, social sensing, and the Web of Things, as fields of a wide interest to vehicle ecosystems. The reasoning features over threats and vulnerabilities of V2I applications, as it has been recently discussed in (Venkata et al., 2019) are of the ultimate interest to IoT4CPS.

4.2 Other European Initiatives for Data Interoperation

4.2.1 International Data Spaces (IDS)

The IDS Association (IDSA) (see: http://www.industrialdataspace.org/en/) consists of more than 90 organizations from more than 15 countries, representing different groups of IDS users and stakeholders.
Figure 9 illustrates the concept of IDS and connected Industry 4.0 (IDSA, 2019). The IDSA envisions data exchange and communication through IDS Connectors and their associated IDS Data Usage Constraints. IDS Connectors are software components that annotate data to be exchanged, with defined usage policies. IDS Connectors need to be established between data providers and data consumers, in order to guarantee data sovereignty based on monitored data usage agreements. Data providers define the data usage constraints within usage agreements.

The IDS is established as a Multi-Sided Platform (MSP) for secure and trusted data exchange, governed by different stakeholders, e.g. data provider, data consumer, research organization, software/service provider, accounting and auditing firm, etc. (Otto & Jarke, 2019). The IDS Reference Architecture Model (Otto et al. 2018) facilitates interaction between the various actors using the IDS platform. The adoption and the use of the IDS platform are guided by the EU regulatory instruments.

[2] https://www.w3.org/TR/vocab-ssn/

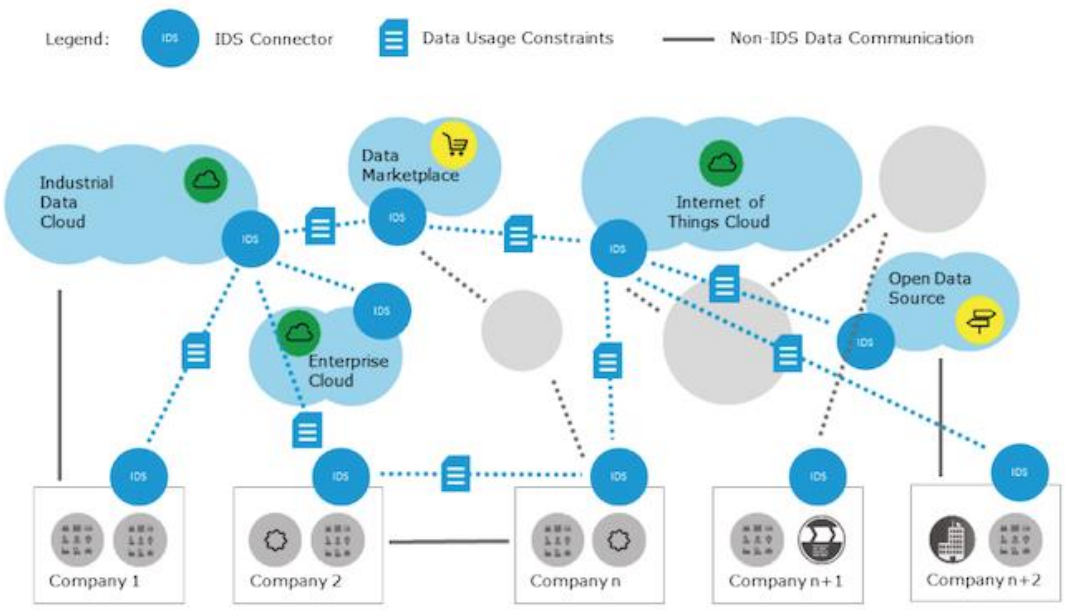

Figure 9 – International Data Spaces (IDS): Concept of Connected Industry 4.0 World

The IDentity Management (IDM) in IDS is illustrated in Figure 10. Here, the Clearing House acts as an intermediary that provides clearing and settlement services for all financial and data exchange transactions within the IDS (Otto et al. 2018). The IDentity Provider (IDP) offers services to create, maintain, manage and validate identity information of and for IDS participants.

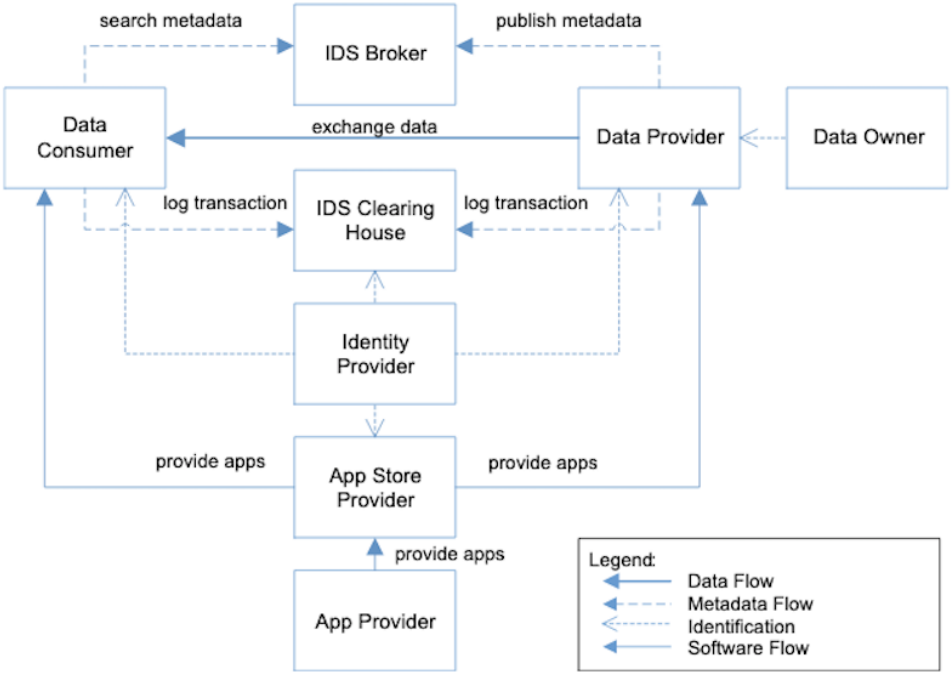

Figure 10 – Identity Management in IDS

4.2.2 Data Market Austria (DMA)

The Data Market Austria[3] (DMA) is the Austrian project funded by FFG and BMVIT (grant no. 855404) with the aim to deliver a platform for commercialization of data and services (DMA D6.1, 2017). Besides the core DMA services (e.g. matchmaking, search), the DMA platform is also about interoperation with the 3rd party services, which requires a unified framework (middleware) for processing all services and data. DMA also consider the development of mechanisms for data quality assessment and enrichments, which is also of interest to Austrian the Digital Twin prototype in IoT4CPS.

4.2.3 European Common Data Spaces

In 2018, the European Commission launched several initiatives and policies related to data, data sharing, data privacy, data processing using a variety of Artificial Intelligence (AI) methods, etc. Some of those initiatives and policies are already in place as the General Data Protection Regulation (GDPR); others prepare specific measures related to different types of data, e.g. scientific information, public sector data and private sector data, or suggest the free flow of non-personal data, encourage high-performance computing, etc. The focus of the Digital Europe Programme in coming years is on developing an infrastructure which offers businesses and the public sector access to AI tools and components and data resources, as well as reference testing and experimentation facilities in some prioritised applications (EC Common European Data Spaces, 2018). For example, in the area of Connected and Automated Mobility (CAM) in Europe, the recent proposal **"5G Strategic Deployment Agenda for Connected and Automated Mobility in Europe"** (5G CAM, 2019) considers the European Commission's principles to access and reuse non-personal data generated by IoT objects, as defined in (EC Common European Data Spaces, 2018).

The concept of European Common Data Spaces will improve the development of a competitive data market in the Automotive sector, and set agreed principles for a data governance related to (5G CAM, 2019):

- Principles for data sharing that are necessary for public interest (e.g. safety reasons);
- Principles for ensuring non-discriminatory access and innovation in relation to services relying on access to data in the respect of investment in data management; and
- Principles for data sharing that are necessary for other public interest purposes (e.g. environmental purposes, sustainable manufacturing, sustainable urbanization, etc.).

[3] https://datamarket.at/

5. Relevant Policies, Initiatives and Regulations

The design of a Digital Twin research prototype in IoT4CPS requires an effective data strategy to be put in place, and in parallel, requires knowledge and understanding of relevant policies, initiatives and regulations to be addressed at national and international levels. ENISA's latest report on good practices for security of Smart Cars (ENISA-SC, 2019) provides concrete and actionable good practices to improve the cybersecurity posture of connected and autonomous vehicles, and maps these practices to current legislative and policy initiatives and cybersecurity regulations. For example, in 2014, the European Commission's Directorate-General for Mobility and Transport (DG MOVE) set up the Cooperative Intelligent Transport Systems (CITS) platform with the aim to ensure interoperability of connected cars across borders and along the whole value chain. In 2017, the Directorate-General for Internal Market, Industry, Entrepreneurship and Small and Medium-sized Enterprises (SMEs) (DG GROW) launched an initiative on safety regulations (ENISA-SC, 2019). In 2018, the Directorate-General for Communications Networks, Content and Technology (DG CONNECT) launched an initiative on Cooperative, Connected and Automated Mobility (CCAM) with the aim to (i) provide a guidance on a governance framework for access and sharing of data generated by connected vehicles, (ii) clarify cybersecurity requirements for the connected car environment, and (iii) provide guidance on the use of 5G connectivity for large scale testing and experimentation for connected vehicles. In 2019, the EC set up a group of a hundred experts named "the Single Platform for open road testing and pre-deployment of cooperative, connected, automated and autonomous mobility" in order to provide advice and support regarding testing and pre-deployment activities for CCAM.

On the international scene, the National Highway Traffic Safety Administration (NHTSA) from the US government issued a document introducing several cybersecurity best practices for smart cars (NHTSA, 2016). The US Automotive Information Sharing and Analysis Center (Auto-ISAC) maintains a series of Automotive Cybersecurity Best Practices, which provide guidance on the implementation of automotive cybersecurity principles. Several cybersecurity standards and recommendation documents are also under development, e.g. UNECE is drafting a proposal for a recommendation on cybersecurity with a focus on key cyber threats and vulnerabilities against.

Recent privacy regulations are expected to have a strong influence on the Smart Automotive sector and its CAM applications, e.g.:

- eCall system (from the 1 of May 2018, all new cars sold in the EU need to be equipped with the eCall alarm systems that can automatically call the emergency services and, in case of an accident, send the location of the car);
- the General Data Protection Regulation (GDPR) (from the 25 of May 2018, GDPR regulatory requirements that are related to privacy data protection within the EU, entered into force),
- the NIS Directive (from the 09 of May 2018, the Directive on Security of Network and Information Systems (NIS), affected search engines, cloud providers and online marketplaces, and set cybersecurity regulations, incident response procedures, and more to be adopted);
- SELF DRIVE act in USA (The Safely Ensuring Lives Future Development and Research in Vehicle Development), from 2017, requires from car manufacturers to develop a cybersecurity plan to regulate access to CAM system and applications;
- "Automated Driving Systems (ADS): A Vision for Safety 2.0" by the US Department of Transportation and the NHTSA, from 2017, encourages best practices (e.g. not to allow for "the development of systems that guard against cyberattacks and protect consumer privacy") and prioritizing safety.

6. Conclusion

Connected vehicles are equipped with sensors, IoT and CPSs components that collects and share a variety of data. Connected vehicles can also be seen as AI devices that process data in real time, offering new business models and opportunities to improve the efficiency and widen the functionality of vehicle's services related to e.g. safe and secure transportation. Identity, security and safety in the Automotive Industry have always been a concern for authorities, governance bodies, manufacturers and the public alike. However, a common standard allowing a complete integration of safety and security measures in the car's PLCDM is still missing [1]. To this end, several frameworks and best practices in the Automotive Industry are being developed to ensure smart car security-by-design (ENISA, 2016), for example:

- Policy and standards, i.e. adherence to regulations, liability, traceability;
- Organizational measures, i.e. general measures, secure development and security until the end-of-life;
- Technical good practices, i.e. communication protection;
- Identification, authentication, authorization;
- Security audit;
- Self-protection, cryptography, user data protection, and more.

Current challenges of data-driven research projects relate to the acquisition of quality datasets. Hence, in WP5 of IoT4CPS, a comprehensive analysis of public datasets is performed to provide an insight into datasets and data repositories of interest to the Automotive sector. In this report, in order to widen the re-usability of publicly shared automotive data, we look at the existing cross-platform interoperation models, including those suggested by knowledge-base and data technologists (e.g. Semantic Web and ontologies) and those being developed through Austrian and European data management initiatives (e.g. IDS, Austrian DMA, the European Common Data Spaces, etc.).

References

5G CAM (2019). 5G Strategic Deployment Agenda for Connected and Automated Mobility in Europe, Initial Proposal published on 31 October 2019. Online available: https://5g-ppp.eu/wp-content/uploads/2019/10/20191031-Initial-Proposal-5G-SDA-for-CAM-in-Europe.pdf Last accessed: November 25, 2019

B. Cyphers & G. Gebhart (2019). Behind the One-Way-Mirror: A Deep Dive into the Technology of Corporate Surveillance. EFF (Electronic Frontier Foundation). CC BY 4.0. Available online from: https://bit.ly/2Ec2bei , Last accessed: December 11, 2019.

BMVIT (2019). "Austrian Action Programme on Automated Mobility". Online available from: https://www.bmvit.gv.at/dam/jcr:56570b3f-9b2a-42b7-838c-4a201a501ef3/action_automated_mobility_2019-2022_ua.pdf

CrackedLabs (2017). Wolfie Christl: "Corporate Surveillance in Everyday Life". Online available: https://crackedlabs.org/en/corporate-surveillance Last accessed: November 25, 2019.

Creech, G. (2014). Developing a high-accuracy cross platform Host-Based Intrusion Detection System capable of reliably detecting zero-day attacks. PhD Thesis. Engineering & Information Technologies, UNSW Canberra. Online: https://bit.ly/2PEcheT Last accessed: November 25, 2019.

Creech, G. and Hu, J. (2013). A Semantic Approach to Host-based Intrusion Detection Systems Using Contiguous and Discontiguous System Call Patterns. Computers.

DMA D6.1 (2017). D6.1 Service Technology Specification and Developemnt Roadmap. Online available from: https://datamarket.at/wp-content/uploads/2017/10/DMA_Deliverable_D6.1_FINAL_v01.pdf, Last accessed: December 10, 2019

EC Common European Data Spaces (2018). Towards a Common European Data Spaces, SWD(2018) 125 final. Online available: https://ec.europa.eu/transparency/regdoc/rep/1/2018/EN/COM-2018-232-F1-EN-MAIN-PART-1.PDF Last accessed: November 25, 2019.

ENISA (2016). Cyber Security and Resilience of Smart Cars: Good Practices and Recommendations. Online available: https://bit.ly/2C3PLEb, Last accessed: November 02, 2019.

ENISA-5G (2019). ENISA Threat Landscape for 5G Networks. Threat assessment for the fifth generation of mobile telecommunications networks (5G)". Online available: https://www.enisa.europa.eu/publications/enisa-threat-landscape-for-5g-networks, Last accessed: December 11, 2019.

ENISA-SC (2019). Good Practices for the Security of Smart Cars. Online available: https://www.enisa.europa.eu/publications/enisa-good-practices-for-security-of-smart-cars, Last accessed: December 11, 2019

EOS (2019). Data Enrichment Exposure From PDL Customer. Online available from: https://eossystems.co.uk/2019/11/28/data-enrichment-exposure-from-pdl-customer/, Last accessed: December 11, 2019.

EU Open Data Portal (2019). Online available: http://data.europa.eu/euodp/en/data/ Last accessed: November 25, 2019.

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

Gaia-X-pub1 (2019). Was ist das Projekt Gaia-X? Online available from: https://www.bmwi.de/Redaktion/DE/FAQ/Dateninfrastruktur/faq-projekt-gaia-x-01.html Last accessed: December 10, 2019.

Gaia-X-pub2 (2019). Eine vernetzte Dateninfrastruktur als Wiege eines vitalen, europäischen Ökosystems. Online available from: https://www.bmwi.de/Redaktion/DE/Artikel/Digitale-Welt/dateninfrastruktur.html, Last accessed: December 10, 2019.

IDSA (2019). International Data Spaces Association: "Reference Architecture Model", IDSA, Version 3.0 (2019). Online available: https://www.internationaldataspaces.org/ Last accessed: November 25, 2019.

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.

Knublauch, H. and Ryman, A. (2019). Shapes Constraint Language (SHACL). Draft. Online available: https://w3c.github.io/data-shapes/shacl/ Last accessed: November 25, 2019.

McKinsey&Company (2016). Monetizing Car Data. New Service Business Opportunities to Create New Customer Benefits. In Advanced Industries, September 2016. Online available from: https://mck.co/36pxI8L Last accessed: December 11, 2019

NHTSA (2016). "Cybesecurity best practices for modern vehicles". Online available from: https://www.google.com/url?sa=t&rct=j&q=&esrc=s&source=web&cd=1&ved=2ahUKEwji2PqR0cDjAhXq2e

AKHcnrAnUQFjAAegQIAxAC&url=https%3A%2F%2Fwww.nhtsa.gov%2Fstaticfiles%2Fnvs%2Fpdf%2F812333_CybersecurityForModernVehicles.pdf&usg=AOvVaw33nVAk2UWXpL3tzDmpRBjl Last accessed: December 11, 2019.

Otto & Jarke (2019). Otto, Boris and Jarke, Matthias, 2019. "Designing a multi-sided data platform: findings from the International Data Spaces case". Electronic Markets, 2019. doi="10.1007/s12525-019-00362-x

Otto, B., Lohmann, S., Steinbuß, S., & Teuscher, A. (2018). IDS Reference Architecture Model (Version 2.0). Berlin, Munich:International Data Spaces Association, Fraunhofer. Online available: https://www.fraunhofer.de/content/dam/zv/de/Forschungsfelder/industrial-data-space/IDS_Referenz_Architecture.pdf Last accessed: November 25, 2019.

PostCaesar Association (2019). ISO 15926 in OWL. Online available: https://www.posccaesar.org/wiki/ISO15926inOWL Last accessed: November 25, 2019.

Reference Data Overview (2019). Online available: http://docs.oasis-open.org/plcs/plcslib/v1.0/cs01/help/techdes_rdl_content.html Last accessed: November 25, 2019

Tavallaee, M., Bagheri, E., Lu, W. and Ghorbani, A. (2009). A Detailed Analysis of the KDD CUP 99 Data Set. 2nd IEEE Symposium on Computational Intelligence for Security and Defence Applications (CISDA).

Venkata et al., (2019). SIMON: Semantic Interface Model for Security in Cyber Physical Systems Using Ontologies. In proceedings of the ICSEA 2019: The 14th International Conference on Software Engineering Advances. Online available from: https://csrl.cse.unt.edu/kavi/Research/ICSEA-2019.pdf, Last accessed: December 10, 2019.

WIRED-data leak (2019). 1.2 Billion Records Found Exposed Online in a Single Server. Online available from: https://www.wired.com/story/billion-records-exposed-online/, Last accessed: December 11, 2019.