CROSS-CPP

Ecosystem for Services based on integrated Cross-sectorial Data Streams from multiple Cyber Physical Products and Open Data Sources

CIDM

Common Industrial Data Model

CIDM
Common Industrial Data Model

White Paper

November 2020

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 780167

Foreword

Welcome to our Cross-CPP CIDM white paper. From the very first, we were absolutely convinced that Data Markets have to become more attractive for its key stakeholders to overcome existing obstacles, as e.g. the limited access to multiple data streams, or privacy concerns. Thus, we build up an Open Ecosystem to empower Data Owners to exploit and control their most valuable data assets from smart products and to give Data Customers access to this great spectrum of sensor data. All along, we have followed the maxim to think about the needs of Data Owners and Data Customers, but also to win smart product manufacturer (e.g. car makers) to open up their products, by designing a convincing trustworthy Ecosystem.

In this white paper, we will present you the Common Industrial Data Model (CIDM), which - being the central standardization concept - is one of the most crucial parts of our Cross-CPP ecosystem. It features the essential link between Manufacturers of cyber-physical mass products (CPP) and Service Providers, making a huge step in simplifying the processes of data exchange and data provision, at the same time allowing data owners full control over their data at any time.

If you got curious about how all that is made possible, just continue on the following pages, enjoy the reading, and please contact us with your feedback or questions!

Cross-CPP consortium partners

Executive Summary

With the increasing number of connected sensors and actuators within cyber physical mass products (CPP) there is an enormous amount of data continuously generated, representing on the one side a new information resource to create new value, allowing the improvement of existing services or the establishment of diverse new cross-sectorial services, on the other side a major big data-driven business potential - not only for the manufacturers of CPP, but in particular also for cross-sectorial industries and various organisations with interdisciplinary application services.

We currently see, however, that these business potentials are currently still locked since manufacturing industry producing CPP are driven by CPP specific business approaches. This situation is mainly characterised by today's sporadic proprietary CPP data access restricted to CPP manufacturer specific products and services and limited access to CPP data caused by missing or distributed access to CPP data as well as by diverse brand specific data formats.

The CIDM Standard presents the crucial link between CPP manufacturers and Service Providers, featuring a flexible and transparent cross industrial CPP data model, enabling a brand independent data exchange between the Data Providers and Data Consumers for data coming from multiple industrial sources, applicable for various industrial sectors. It is an open and highly scalable big data format, flexible enough to harmonize proprietary data harvested by the CPP Manufacturer into datasets tailored to the needs of the data consumers coming from different industrial sectors. It is satisfying the large range of data and data representations needed by the data customer, at the same time enabling the data owners to have full control about their data.

Thus, the CIDM standard is not an immutable model, but rather represents a *living standard*. Whenever it is required, with respect to the market needs of the service provider community the CIDM standard can be adapted to match these new requirements.

In this White Paper, we present to you the following key points that show up the high potential of the CIDM for all Cross-CPP stakeholders:

Having the CIDM in place, both sides - service providers as well as data providers - are in a win-win situation and profit from the standardisation.

1 Challenges on the Way to a Brand Independent Data Exchange

The Key ***motivation to establish the Common Industrial Data Model (CIDM)*** *standard* is to give cross-sectorial industries access to the great spectrum of sensor data coming from high volume products from various industrial sectors (vehicles, buildings, etc.). With the increasing number of connected sensors and actuators within such mass products there is an enormous amount of data continuously generated by mass products representing

- ***a NEW information resource to create new value***, allowing the improvement of existing services or the establishment of diverse new cross-sectorial services, by combining data streams from various sources and
- ***a major big data-driven business potential***, not only for the manufacturers of Cyber Physical Products (CPP), but in particular also for cross-sectorial industries and various organisations with interdisciplinary application services.

A Missing Interface Standard Constraints the CPP Data Access

However, these business potentials are currently still locked since manufacturing industry producing Cyber Physical Products (CPP) are driven by CPP specific business approaches. This situation is mainly characterised by the following major difficulties for both, the CPP manufactures/owners as data providers and the Service Providers as data consumers:

- Today's sporadic proprietary CPP data access is in most cases restricted to CPP manufacture specific products and services and not open for third parties interested in these CPP data.
- No or limited access to CPP data caused by missing or distributed access to CPP data as well as by diverse brand specific data formats. This situation is forcing Service Providers interested in CPP data to build up and maintain interfaces to diverse CPP manufacturers with different data models, causing high efforts and costs for data collection and processing.

These constraints of brand-specific solutions, not providing brand-independent CPP data to the outside world, are hindering on the long-term the value creation by service providers. Furthermore, such situation is characterized by far too complex and individual value chains resulting in economic inefficiency.

The Cross-CPP CIDM Approach to Overcome the CPP Data Access Constraints

The Common Industrial Data Model (CIDM) represents a unique standard data model, removing brand proprietaries and dependencies and creating a common data format. Figure 1 demonstrates the CIDM as essential link between CPP Manufacturers and Service Providers. On the left side CPP Manufacturers transfer their proprietary data into the CIDM data format. By this, brand-dependent information is removed and aligned to fit the CIDM format. Service Providers, on the right side, can access these data in a standard, brand independent format and can receive generic data sets of various brands. Both sides are in a win-win situation and profit from the standardisation.

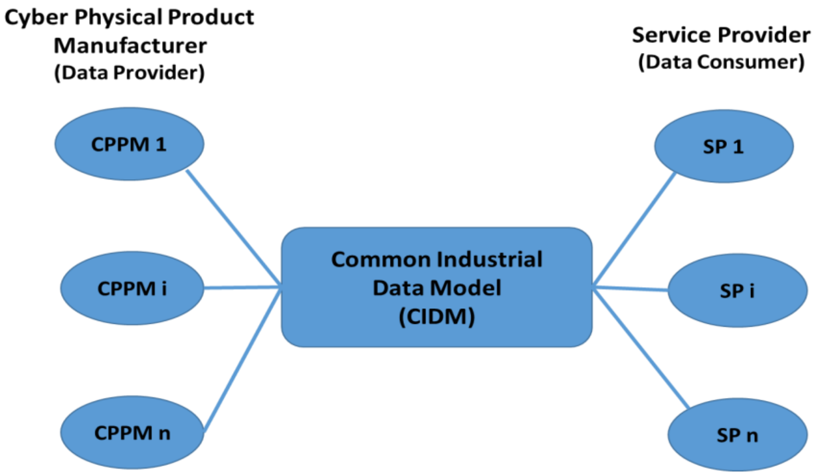

Figure 1: CIDM Standard link between CPP Manufacturers and Service Providers

The CIDM Standard, presenting the crucial link between CPP manufacturer and Service Provider, has to overcome several obstacles and facing the following basic requirements:

- To be a flexible and transparent cross industrial CPP data model, enabling a brand independent data exchange between the Data Providers and Data Consumers for data coming from multiple industrial sources, applicable for various industrial sectors.
- To be an open and highly scalable big data format flexible enough to harmonize proprietary data harvested by the CPP Manufacturer into datasets tailored to the needs of the data consumers coming from different industrial sectors.
- To satisfy the large range of data and data representations needed by the data customer, therefore the model has to support of different signal and measurement types
- To provide Data Ownership Information to enable the data owner to have full control about his data.

Not to invent the wheel twice, for the CIDM approach various exiting data models were taken as reference. There are several existing data formats fulfilling some of the mentioned requirements, among others the W3C – “Vehicle Data” format, which is only meeting some of the requirements, the ERTICO – “SensorIS” data format, which meets most of the requirements but only supports as data types time series signals. Proprietary and closed source models exist but are out of scope as openness is mandatory. None of the approaches fulfil the key CDIM requirements. Any extension of these standards would be very complex and therefore not recommended.

As most suitable starting point for a data model for standardizing data formats from different industrial sectors the Common Vehicle Information Model (CVIM)[1] was identified. This Model fulfils all stated basic requirements providing a brand-independent and transparent data model. However, the CVIM was originally developed for the brand-independent access of vehicle data by Service Providers.

Therefore, the key extension for the transition from CVIM to CIDM data model is related to the required extensions to support besides vehicle data also CPP data sources from other cyber physical products of various industrial sectors (e.g. CPP type ‘Smart Building’). The extended CIDM Model pushes the Win-Win value chain for all ecosystem partners, due to the fact that the costs for data exchange can be shared by a large amount of data customers, which will make a single service much more economical.

[1] CVIM was developed in the scope of the AutoMat project “Automotive Big Data Marketplace for Innovative Cross-sectorial Vehicle Data Services”, EU funded AutoMat – 644657

2 The CIDM Standard Structure to Meet the Challenges

A data model architecture was selected convenient to satisfy the stated challenges and which fulfil the stated features for the provision of the data. Therebye, the CIDM architecture consists of 3 layers as presented in Figure 2. The CIDM format is unbiased and treats all data in the same way. It defines Signals as information providers and Measurement Channels as the recording of the measurement configuration of those signals. Hereby, CIDM does not differentiate between the data origin within the CPP, the CPP type or its manufacturer.

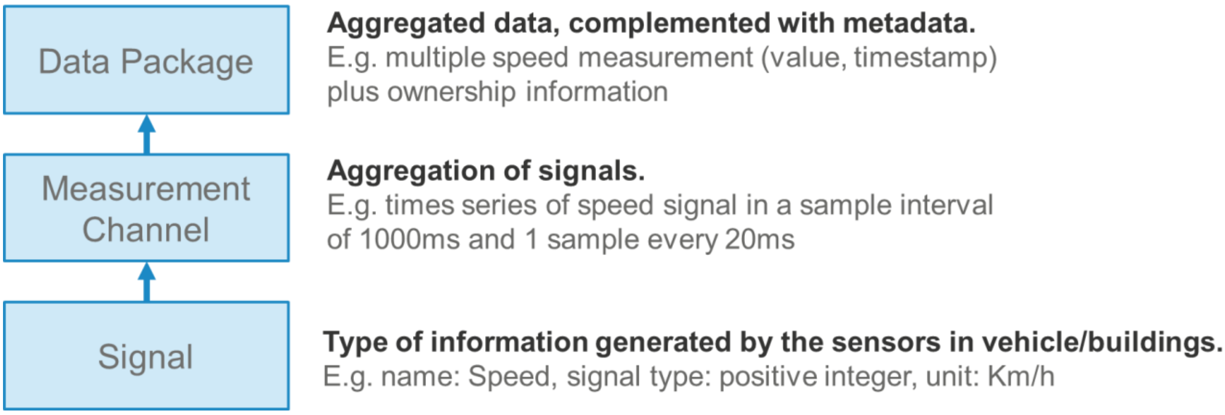

Figure 2: CPP Data Model Main Structure

Signal Layer Specification

Sensors are the perception organs of CPP devices like vehicles and buildings. It is their main duty to detect physical phenomenon and chemical quantities by transferring them into electrical signals. The signal layers describe different types of signals and formats represented in the system. A new property is needed to group signals regarding the signal source type, CPP-type etc.

Measurement Layer Specification

The measurement layer defines how sensor signals are captured and processed. As data contained in signals may far exceed the available transmission bandwidth and as the full resolution may not be required in most applications, signals are pre-processed and aggregated into measurement channels. One Measurement Channel describes how samples from one or more sensor signal are aggregated and measured. Various types of data channels are supported (Time Series, Histogram etc.).

Data Package Layer Specification

Data Packages provide a structure for storing and retrieving data. Data Packages contain the actual data of Signal measurements. They cover the Signals as information providers and the Measurement Channels defining the process of the data acquisition of these Signals. In addition, Data Package provide meta/header information containing time of recording, data ownership information, etc. Data Packages contain data from exactly one Measurement Channel.

3 CIDM Data Provision Tailored to the Service Needs

CPPs can provide a huge amount of data to the outside world. E.g. inside today's vehicles ~4000 bus-signals at a sampling interval of about 10ms are theoretical accessible, corresponds to 2.68 GB/hour. Forwarding that amount of data for thousands of vehicles from various brands would overstress the communication links for the data exchange as well as requiring huge amount of storage areas. Even when technical feasible, such an approach is not reasonable from an economic point of view. Therefore, there is the essential need to reduce amount of data to be transferred and stored.

On the other hand data model has to enable the Service Providers to access CPP data optimal tailored to the needs of their services, in respect to the required signals, sampling rates, type of representations etc. The CIDM offers a wide spectrum of data representations on the various model layers, tailored to the needs of the Service Providers and on the other hand enabling to constrain data volumes to be communicated and stored.

Accessible Signals

Signals represent the in-CPP provider of observations, raw sensors measurements and CPP status information. Signals can vary between different type of CPPs, different brands, models or even CPPs of the same model may differ by value range, sampling rate or resolution etc. Different types of signals both continuous and discrete data (like wipers on/off, outside temperature, room temperature, sun intensity etc.) as well as textual (e.g. Vehicle model) and enumeration (e.g. fuel type) are supported. The CIDM Model enables that the list of signals handled by the CIDM can be modified/extended.

Confirmed by the CPP Manufacturer, each product provides a specific set of signal accessible in the CIDM format. For each type of CPP the accessible signals are specified. However, some manufacturer might not provide all specified data due to missing sensors in the product model or company confidentiality etc.

Types of Data Representation (Measurement Channels)

To minimise the amount of data exchanged and stored the data contained in signals are aggregated and pre-processed, represented by so called Measurement Channels. By several Measurement Channels the CIDM supports various data representations tailored to the needs of the Service Provider community. Examples of CIDM Measurement Channels are presented in Figure x.

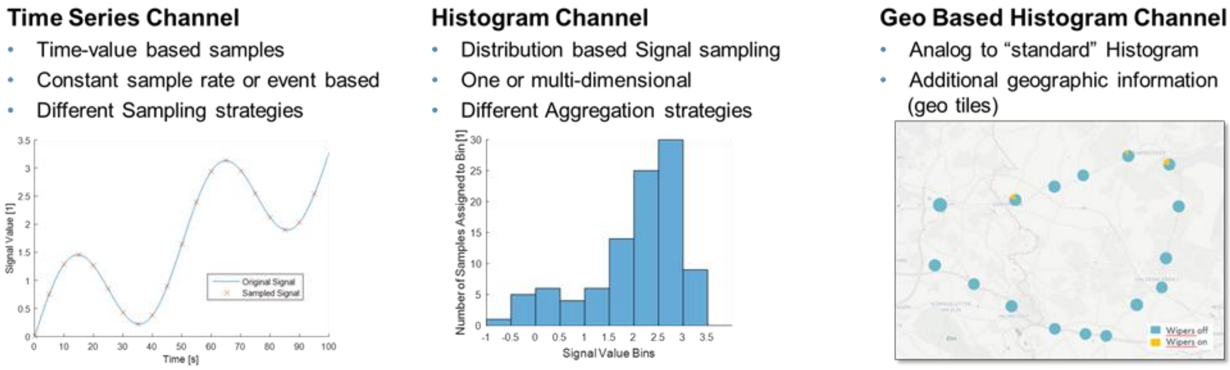

Figure 3: Examples of measurement channels

Typical measurement channels of the CIDM address:

- **Time Series:** Sequences of data recordings (samples) from a signal in a time order with a constant sampling rate or event based. Each data sample is assigned to a specific point in time, respective geo location may be recorded, too. Furthermore, min/max and averaging processing features are supported. Events might represent switching status or parameter thresholds etc. By a service tailored sampling rate, as well as the min/max and event based features the data volumes can be reduced.

- **Histogram:** Distribution of sampled sensor data, providing a rough sense of the density of the underlying distribution of the sampled sensor data. Thereby, the entire value range is divided into a series of intervals and it is counted how many values fall into each interval. CIDM also support multidimensional histograms. Histograms may be created over different time ranges so called “Capture Interval”. They may have dimension of days, months etc. Histograms enable a very high sampling rate but an enormous reduction of the data to be stored and retrieved. However, there is a loss of reference in respect to time and location of the samples.

- **Geo-based Histogram:** The concept of geo-histograms enriches the histogram approach by a geospatial dimension. To the classic histogram information (how often a certain value occur in a certain time period) additionally the information of where a certain histogram was observed during the observation period is recorded. The additional geographic information is represented by geo-tiles, embodying a certain map tile based on the Mercator projection.

The CIDM architecture is open to add additional Measurement Channels if required for the analysis of available signals by the Service Providers for their services.

Data Package the Transport Format

On the Data Package layer, several measurement channels are aggregated in Data Packages to be stored and retrieved. In addition to the data themselves, Data Packages also contain metadata, with support information like ownership and quality assessment etc. Therefore, the CVIM Data Package is divided in two parts: (i) descriptive metadata and (ii) the aggregated data. (see Figure xx). Metadata support indexing and sorting the data by leveraging the statistic properties, timestamps and geographic estimate in form of a bounding box. In addition, it provides ownership information, privacy levels and data stakeholders. For the sake of completeness and validity, CPP manufactures can sign data packages using sequence numbers, checksums and signature. The second part of Data Packages stores aggregated data of Measurement Channels. This may be either a time series recording or pre-processed value distribution in form of a histogram etc.

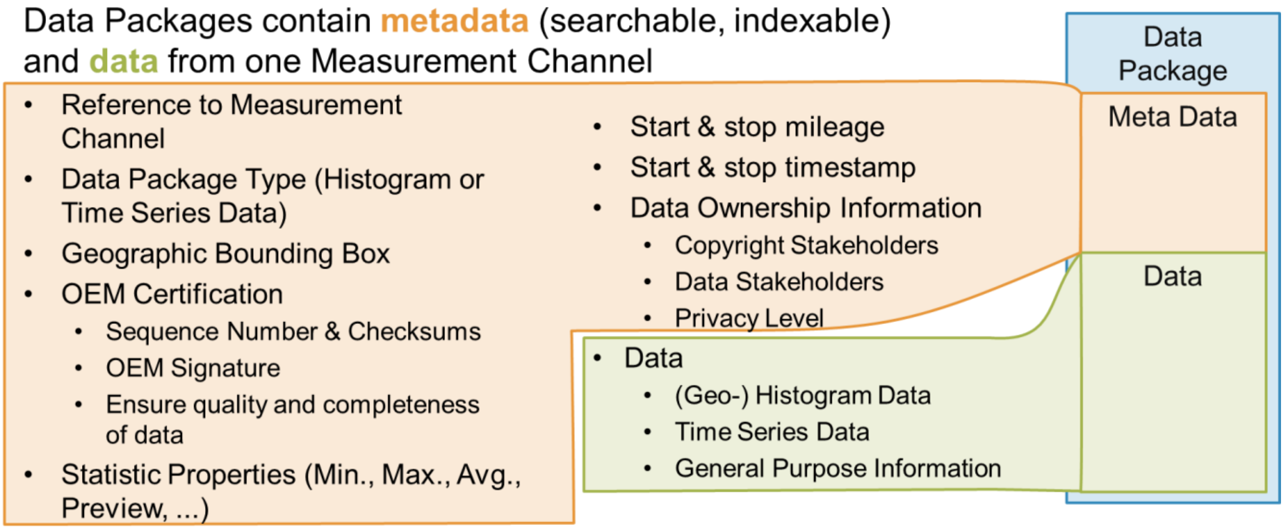

Figure 4: Data Package structure

Finally, it is important to mention that each CPP needs to transmit not just sensor information, but also information about the CPP itself – i.e. basic CPP information. To illustrate, it may be relevant to know the colour of a vehicle, the materials or number of windows in a room as this may condition the readings of temperature sensors: e.g. a black vehicle under the sun will probably yield much hotter estimations. The same happens for a room in which too many windows are located. Note that such basic CPP information could be attached as additional parameters to the existing measurement channels; yet, basic CPP information is probably not going to change frequently, and therefore they do not need to be sent in every reading. Therefore, the Data Package passes through the creation of a specific basic CPP information channel, in which one basic CPP information measurement is transmitted at the beginning– or whenever an update is made necessary.

In fact, the CPP model explained was conceived for the Cross-CPP project in which the main entities exchanging information where vehicles and buildings. However, the widely use of this model as the data model in other domains such as aeronautics, smart manufacturing, oceanographic studies is almost straight forward as only the new channels will have to be defined. The model could even be shared among competitors.

4 CIDM Represents a Living Standard

The brand-independent, open and transparent CIDM standard is not rigid, but rather representing a living standard. Whenever, it is required in reference to the market needs of the service provider community to extent the amount of signals to be recorded or to modify/extent the type of measurement channels the CIDM standard can be adapted to match these requirements. This feature of the CIDM standard to permit respective extension/modifications represents a key measure to enlarge significantly the number of data customers.

A standardisation board composed by the key shareholders of Cross CPP Ecosystem will decide about the CPP data model reconfiguration requests of the Service Providers and an updated version of the CIDM will be agreed.

About Cross-CPP

The objective was to establish an IT environment for the integration and analytics of data streams coming from high volume (mass) products with cyber physical features, as well from Open Data Sources, aiming to offer new cross sectorial services and focusing on the commercial confidentiality, privacy and IPR and ethical issues using a context sensitive approach. The project addresses cross-stream analysis of large data volumes from mass cyber physical products (CPP) from various industrial sectors such as automotive, and home automation. The business objective of the research was to allow for analyses of such data streams in combination to other (non-industrial, open) data streams and for the establishment of diverse enhanced sectorial and cross-sectorial services. The project developed: (i) New models for integration and analytics of data streams coming from multi-sectorial CPP, including shared systems of entity identifiers applicable to multi-sectorial CPP (as well as the definition of agreed data models for data streams from multiple CPP aiming at defacto standard; (ii) Ecosystem, including a common Marketplace, and methodology to use such models to build multi-sectorial cloud based services, (iii) Toolbox for real-time and predictive cross-stream analytics, context modelling and extraction, and dynamically changing security policy, privacy and IPR conditions/rules and (iv) set of services such as services based on a combination of data streams from home automation and (electrical) vehicles to provide enhanced local weather forecast and predict and optimise energy consumptions in households. The project has built upon the results from past and current projects, where results from the project AutoMat, addressing services developed based on data streams from vehicles, were used as a basis for Cross-CPP development extend it to integrated, cross-sectorial data streams analytics. More information is available at https://cross-cpp.eu

Funded by the Horizon 2020 Framework Programme of the European Union