

Cross-CPP – An Ecosystem for Provisioning, Consolidating, and Analysing Big Data from Cyber-Physical Products

Ana Correia
Christian Wolff
Institute for Applied Systems
Technology Bremen GmbH
Bremen, Germany
[correia, wolff]@atb-bremen.de

Massimiliano Zanin
Ernestina Menasalvas
Centro de Tecnología Biomédica
and Escuela Técnica Superior de
Ingenieros Informáticos
Universidad Politécnica de
Madrid Madrid Spain
[massimiliano.zanin,
ernestina.menasalvas]@upm.es

Elisa A. Herrmann
Victor Corral
ARI
Atos IT Solutions and Services
Iberia
Madrid, Spain
[elisa.herrmann,
victor.corral]@atos.net

Miriam Kachelmann
Meteologix AG
Sattel (SZ), Switzerland
miriam@kachelmann.com

Rance DeLong
The Open Group
Reading, Berkshire, UK
r.delong@opengroup.org

Pavel Smrz
Brno University of Technology
Brno, Czech Republic
smrz@fit.vutbr.cz

ABSTRACT

It is expected that with the increasing number of connected sensors and actuators within mass products, the large spectrum of sensor data coming from high volume products in various industrial sectors (vehicles, smart home devices, etc.) will rise in short-term. This enormous amount of data continuously generated by CPPs will represent (1) *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 (2) *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 as well as various organisations with interdisciplinary applications.

In spite of major advances in the field, several challenges still hinder the use of these data, like the lack of, or only few, CPP ecosystems that are in the best-case manufacturer specific and not open for external companies interested in using such data.

We present here a solution that envisions to establish a CPP Big Data Ecosystem to bring to the outside world CPP data from various industrial sectors, brand independent, allowing for external service providers that use CPP data from this unique CPP data access point (as well as from other sources) to develop cross-sectorial services.

KEYWORDS

Cyber-Physical Products, Cross-sectorial services, Big Data Marketplace

1 Introduction

In the present world where mass products have an increasing number of connected sensors and actuators, it is expected that a large spectrum of sensor data coming from high volume products in various industrial sectors (vehicles, smart home devices, etc.) will rise in short-term. This CPP enormous amount of data has in today's landscape only sporadic proprietary CPP ecosystems, which are restricted to manufacturer-specific services and not open for third parties interested in the CPP data. The Cross-CPP project will tackle these issues by focusing on what CPP and their sensor data can bring to the outside world. Therefore, as key challenges, Cross-CPP aims to overcome several obstacles by establishing a CPP Big Data Ecosystem, with the following main characteristics:

- Brand independent concept, open for integration of diverse CPP data providers coming from different industrial areas, also providing a standardized cross industrial CPP data model which needs to be flexible enough to incorporate data coming from various industrial sectors.

- A CPP Big Data marketplace providing a single CPP data access point with just one interface (one-stop-shop) to service providers, as well as support functionalities for easy data mining/analytics. By these means, data customers (Service Providers) just need to set-up and maintain one interface to gather diverse CPP data from different CPP providers.
- Controlled access to diverse CPP data streams and optimal management of data ownership and data rights, applicable to various cross CPP data streams.

In general, as seen in Figure 1, the ecosystem can be separated into three pillars:

1. **Left pillar:** Data Providers (CPP Producers / Owners) -> Comprising data harvesting and making CPP data from various industrial sectors available, transfer brand specific data streams into the common CPP data model.
2. **Middle pillar:** Cross-CPP Cloud Storage & Big Data Marketplace (MP) -> Comprising a cloud based concept for CPP Cloud Storage. Enabling controlled access to CPP data from different sources, offering support to Service Providers in the form of an easy access and detection of needed data, as well as of flexible cross data stream analysis tools.
3. **Right pillar:** Data Customer (Service Provider) -> Cross-sectorial industries or manufacturers of CPP using CPP data from various products to create new value out of that data ("CPP-data" has no value in itself), by improving services or the establishment of diverse new cross-sectorial services.

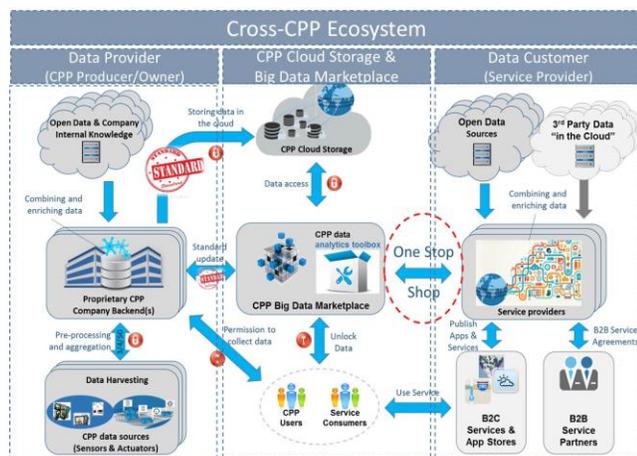


Figure 1: Cross-CPP Ecosystem

In Section 2 we describe the state of the art, continue with the Cross-CPP Ecosystem concept, and use the Cross-CPP architecture to explain the different modules of the Ecosystem as well as their purpose (section 3). We subsequently explain how, in the scope of the project, it is being applied in industry (Section 4). Section 5 concludes with the main learnt lessons and steps ahead.

¹ The Common Vehicle Information Model (CVIM), represents a brand-independent, open and transparent data model for vehicle data. The CVIM is representing a living data structure, where in reference to the needs of the service provider community the

2 State of the Art

2.1 Data Model

The proliferation of intelligent devices together with modern computing paradigms such as cloud, fog service-oriented computing is exponentially growing the amount of data recorded and stored [1]. The massive amount of data available within a company represents the key to competitiveness. However, data and more data are useless without methods, methodologies and tools to manage them. From a technical point of view, the progression in the amount of data is handled by a new breed of technologies and techniques such as NoSQL databases, MapReduce computation framework, machine learning algorithms etc. Nonetheless, the usage of these technologies is impractical if data are not structured i.e. data are not framed into a predictable and regularly occurring data format in order to be managed by computational components/modules for data analytics tasks as confirmed in [2]. Therefore, it is necessary to standardize and homogenize the way data are represented and structured (agreed data model) to cope with the problem of integrating data from multiple vendor-based systems for the sake of knowledge generation and information distribution to upper managerial decision-making tools.

In order to allow for integrated data access the project has started the development of the Agreed Data Model from the so-called CVIM¹ developed in the scope of H2020 AutoMat project (GA no. 644657) for vehicles representing an agreed format for storing data in the cloud [3], and extend it to other CPP. From application point of view a combination of cooperative storage clouds and traditional storage clouds were addressed. This data cloud also represents the regulating interface for the data exchange between the CPP and the various service providers. This approach provides a breakthrough regarding an open data exchange that overcomes the drawbacks of current restricted products ICT services concepts. Currently there exists no standard information model for keeping, maintaining and aggregating data from and for CPPs. The development of agreed data models is not just a technical issue. The key is to not only define a data model, but to reach consensus among industrial players and make sure the models are used and shared i.e. to become "de-facto standards".

2.2 Data Marketplace

As it is well known, there exist a wide scope of marketplaces in the internet handling B2B, B2C and C2C relations. However, the marketplace for trading data streams are rare, especially in industrial domains. For example, in smart manufacturing domain [4] shared, secure, open-access infrastructures rich in functionality for easier system integration and composability and a marketplace that can drive technological capability beyond just products by integrating services on standards, uncertainty quantification, benchmarking, performance-use metrics, systems modelling, etc. are still missing, but many initiatives are currently active (e.g. in US Leadership Coalition [5]). In AutoMat this concept was for the

number of signals to be recorded as well as the type of measurement channels can be modified or extended.

first time applied for the exchange of customer owned vehicle data and service providers [6]. Therefore, the AutoMat Vehicle Big Data Marketplace being the comprehensive platform to manage the sales and provisioning of all type of vehicle related data from all OEMs (the project brought VW, Renault and Fiat together) and the various service providers [7], will serve as a basis for Cross-CPP Marketplace by extending it to cover data streams from various CPP. This large amount of continuously gathered CPP heterogenous data represents major economic big data business potentials, not only for specific industry verticals (as Automotive) but for cross-sectorial industries with interdisciplinary applications. Today's proprietary approaches focus on bringing company services into Vehicles (e.g. [8]), Home Systems Entertainment without open-up to specialized cross-industries companies.

Despite of that fact, it is still a major business potential locked because the automotive and related industries were not yet able to establish an open service ecosystem equivalent to existing market applications such as e.g. Smartphone Apps. Such approaches fit quite well, and could easily be adapted to the requirements of the data handling in a cloud environment and the management of the information exchange between companies, vendors of CPP and service providers.

The above approach is not cost-effectiveness for companies because the associated costs and time-consuming process by each sign-off agreement makes not feasible to scale up into production environments the monetization of our CPP data lakes. Despite of this fact, it is how companies have been doing since the early stages of data-driven services conceptualization. They have been trying to create ad-hoc applications and services for their customers which must be sustained and supported by periodical subscription fees of connected services. Data Providers and Data Consumers charge all the cost of the product development lifecycle into the final cost of the services offered to end-users. Finally, they realized about the fact that rather than working on isolated restricted (closed) interfaces with specific Services Providers, it doesn't render into economically feasible scenarios. Cross-CPP seeks for open-up and democratize the access to CPP data among cross-sectorial Service Providers under a standardized CPP Data Model trying to maximize the monetization of their data users by cross-sectorial Data Consumers which are experts and clearly knows the needs of our customers for accessing to digital services.

2.3 Big Data Analytics

The challenges of the CPP data stream mining is to analyse how evolving data in the different sectors (e.g. home, automotive) behave. In this context, the algorithms have to be designed and adapted to deal with resource aware learning, change detection, novelty detection, multi-horizons analysis, and reasoning about the learning process in the different domains [9] [10] [11]. The CPP real-time and predictive analytics toolbox will be an extension of the software approach developed within the JUNIPER project [12]. This approach was based on Java 8, which introduced Streams and Lambda expressions to support the efficient processing of in-memory stream sources. One of the primary aims was to accelerate

data processing (including analytics) for a parallel (e.g. cloud) platform, i.e. to process the data as fast as possible using all of the available processors. It is therefore an ideal basis for data analytics on parallel (i.e. cloud platforms), including pipelined and data-parallel approaches.

2.4 Context sensitivity

The acceptance and usability of complex cross-sectorial services can be considerably improved by making them context sensitive. With the recent advance of context sensitivity, an increasing need arises for developing formal context modelling and reasoning techniques.

The basis for context-aware applications is a well-designed Context Model (CM). As context integrates different data and knowledge sources and binds knowledge to the user to guarantee that the understanding is consistent, context modelling is extensively investigated. A CM enables applications to understand the user's activities in relation to situational conditions.

Typical context modelling techniques include key-value models, object-oriented models, and ontological methods [13]. The problem to be solved is how to extract context from the CPP use. Since it is planned to model context with ontology, context extraction mainly is issue of context reasoning and context provisioning: how to inference high level context information from low level raw context data [14] and [15]. Application of context awareness for cross sectorial services has not yet been sufficiently researched. In the case of such services the notion of context refers to process preferences of CPP and process skills of devices, physical capabilities of the CPP and environment conditions.

The modelling of context in this case presents an additional challenge, as the mentioned services are highly dynamic and reside in distributed environments. Up to now there were no industrial driven attempts to provide harmonised modelling of context under which CPP are used or under which data streams from CPP are generated. The key innovation issues to be solved are: how to allow building common, re-usable context models for cross sectorial services; and how to provide a generic solution adaptable to different scenarios.

3 Cross-CPP concept

This section presents the results of the developed concept for the general Cross-CPP Ecosystem Architecture. Figure 2 shows the overview of all modules that were planned to reuse and significantly enhance the results of the past projects w.r.t. Cross-CPP project needs and objectives. This figure was derived from the Cross-CPP Ecosystem (see Figure 1) and it summarises the Cross-CPP modules and its key software components and how they correlate in the system.

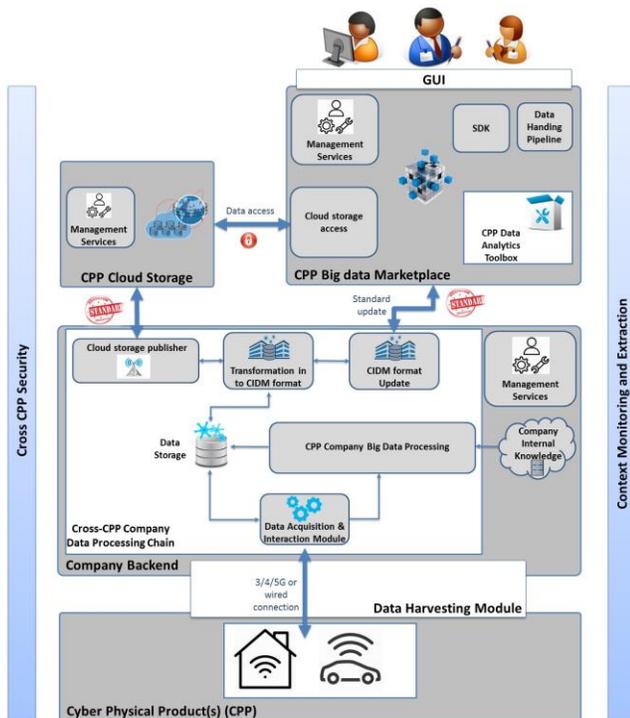


Figure 2: Conceptual Cross-CPP system architecture [16]

3.1 Data harvesting

The Data Harvesting module acts as intermediate layer between the CPPs and the Company Backend module. The connection will be realised using 3/4/5G mobile web technologies in the case of vehicles and wired connection in the case of smart infrastructure devices.

The main functionalities of the module are the:

- Set-up CPP system and data acquisition configuration
- Continuous data acquisition and transmission

Where the first will offer functions to allow for the signals coming from the CPPs to be configured in terms for instance of retrieval and transmission rate, etc. and the second is the actual data acquisition and transmission of said signal data from CPPs to the Company Backend (see section 3.2).

The main components in this module are:

- The data logger in the CPP continuously measures data during CPP usage, according to the deployed measurement configuration.
- The measured data are continuously aggregated and stored in CPP data packages. According to the deployed measurement configuration, the aggregation and storage of data is done for each defined measurement channel.
- The stored CPP data packages are sent to the CPP Company Backend at the agreed frequency.
- This component receives the request for management configuration initialisation / update from the CPP

Company Backend and deploys the new configuration in case the data owner agrees.

- Performs the basic authorization and authentication security services

Both the Data Harvesting module as well as the Company Backend are conceptually generic for any CPP producer/Data Provider but have a company specific implementation due to the diversity of the companies and their internal systems.

3.2 Company Backend

The Company Backend module holds the Cross-CPP company data processing chain, which receives data from the Data Harvesting Module and after processing, enrichment with company internal knowledge, formatting and transformation, stores the processed data into CIDM format in the CPP Cloud Storage (see section 3.3). The Company Backend module holds bidirectional connections to the CPP Cloud Storage as well as to the CPP Big data Marketplace for transferring the data and updates of the CIDM format.

The main functionalities of the module are to:

- interpret and transform proprietary CPP manufacturer specific CPP data into physical information in reference to agreed owner permissions,
- validate the information and, if need be, mask it to enforce privacy.
- convert the information into the required quasi standard information representation, the Common Industrial Data Model (CIDM) format and publish it to the owner’s CPP Cloud Storage, and
- manage the configuration procedures for the data mining at CPP level, by providing CPP specific data logger configurations (e.g. manages for instance the case that a certain CPP, even if from the same sector, often has other models or configurations and therefore has some signals that other does not).

3.3 Common Industrial Data Model and CPP Cloud Storage

The Common Industrial Data Model (CIDM) is an open and highly scalable big data format, designed to harmonize IoT proprietary data into generic datasets.

The structure of the data model consists of three layers depicted in Figure 3.

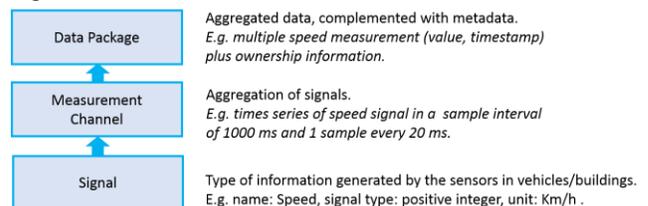


Figure 3 CPP Data Model main structure

- Starting from the bottom part, Signals describe the type of physical phenomena and chemical quantities of vehicles and buildings, including the name of the signal, the format and unit.
- As measurements of the phenomena may far exceed the available transmission bandwidth or the full resolution may not be required in most applications, data from the CPP need to be pre-processed and aggregated conforming a “measurement channel” that include the signals to aggregate (1 or more), the aggregation type (time series, histograms, etc.) and the configuration of the aggregation.
- Finally, at the highest level, data packages provide the actual data coming from the CPP, aggregated according to a measurement channel selected. The data packages are stored in and retrieved from the Cloud Storage. In addition to the data, data packages also contain metadata with support information like ownership and quality assessment.

The CPP Cloud Storage is a cloud-based data storage infrastructure that offers secure and private “data vaults” to data providers, to store their devices data packages in the CIDM format. The storage infrastructure provides an Application Programming Interface (API) to enable data collection from the Company Backend as well as data access by the CPP Big Data Marketplace. The module includes a web application that allows users to have the control of their data by managing read or write access permissions granted to the Marketplace and the Company Backend.

3.4 CPP Big Data Marketplace and Data Analytics toolbox

The CPP Big Data Marketplace connects Data Providers and Data Consumers for selling and acquiring Connected Vehicle and Smart Building data under the standardized data model (CIDM), assuring security and privacy of the data. The Marketplace main purpose is to allow Services Providers to create new B2B and B2C data-based products and services.

The architecture of the Marketplace module consists of a set of components with different responsibilities:

- **Indexing:** This component indexes the metadata of the CPP data stored in the different Cloud Storages modules to provide data discovery services and to locate and retrieve CPP data from Cloud Storages when needed.
- **Discovery:** This interface allows to check the types, amount and quality of the CPP data stored in the Cloud Storage spaces considering data consumer constrains and requirements.
- **Cloud Storage Access:** This component has two responsibilities, to handler data change notifications from the Clouds Storages and to request CPP data packages to the proper Cloud Storage when a Service Provider request data.
- **Data Broker:** This component provides an interface to retrieve the CPP data in subscription (streaming) mode or in a pull mode (REST request).

- **Catalogue:** This component manages the set of signals and measurement channels available for the current version of the CIDM.
- **Data Providers and Service Provider Manager:** To manage data provider users, the services provider users and to manage the sharing process between them.

The module comprises a backend application with a RESTful API to provide the functionality, an index storage, and a frontend application (web application) to provide the different actors the visualization of the data and the visual management of all the resources. The module includes a Software Development Kit for retrieving datasets in an easy way.

An added value to the marketplace is the Data Analytics Toolbox that extends the marketplace functionality to provide the Service Providers with Analytics Capabilities. It includes, besides the CPP Data Analytics toolbox, the Software Development Kit. The former is composed of a set of modules facilitating the analysis of the collected data. It is based on a modular structure, in which new analytics services can be added to fulfil new user requirements; and it is aimed at supporting both fast prototyping of new ideas, and efficient implementation of data synthesis and analysis techniques. Each module, devoted to a specific analysis, communicate with the Cross-CPP Marketplace to get the data and return the results. The communication between each module and the central system can take two forms:

- **Pull mode:** analyses are performed over the data provided by the system (or the final user), without taking into account the evolution of the system up to that point. The user/service provider requests for an analysis and once the results are yielded, the whole computation is deleted.
- **Push, or stream, mode:** internal models are updated in an asynchronous way, using any new data available, such that the request for an analysis just implies retrieving the result. In this case, the availability of new data triggers the update in the analysis and the user accesses results as if they were any other data stream in the system.

Within the plethora of data analytics techniques that have been developed in the last decades, we have here selected some of them for being relevant in a large range of applications. In short, these include:

- **Basic statistics:** module that aims at providing with some very simple statistical functions, calculated over a subset of the stored data, and with the objective of minimizing communication overheads.
- **Time series:** module providing the service provider with a set of tools for detecting when a time series, that represents the evolution of a measurement, suffers from a sudden change.
- **Trajectories:** aims at making a set of basic tools available to the service provider, in order to simplify the handling and manipulation of trajectories.
- **Machine learning:** module supporting incremental learning algorithms by means of existing libraries and

frameworks, that proved to be applicable in high velocity settings.

Both the raw data and the analytics results can be accessed in two ways. On one hand, the system provides an SDK for programmatic access to the data. Alternatively, these can be explored through a GUI, mainly used by Service Providers to select and configure the access to the cross-sectorial CPP big data pool offered by the Cross-CPP data providers via the CPP Cloud Storage.

3.5 Cross-CPP Security

The Cross-CPP security approach applies and extends an implementation of the NGAC Standard [17] that provides fine-grained attribute-based access control for access to the CPP Cloud Storage.

The distinguishing characteristic of the Cross-CPP implementation derive from the objective to provide dynamically changing security policies that depend upon the context under which data streams are used or generated, and to adapt services to the current needs of the user based on the current context.

To achieve this objective the implementation is extended with an enhanced declarative policy language that enables changing policy modes based on the current values of context variables, and a new Event Processing Point that enables the currently active policy to be dynamically changed on the basis of context change and the occurrence of events generated within the access control system or elsewhere in the Cross-CPP system.

The primary components of the access control implementation are a Policy Server that provides interfaces for policy decisions and for policy administration, an Event Processing Point that provides an interface to the Context System and a mechanism to execute changes to the access control policy as the result of specified events, and a Policy Tool to assist in the development and testing of Declarative Policies and Event-Response packages. These artefacts are expressed in two distinct languages that are used to configure the behaviour of the Policy Server and the Event Processing Point respectively.

Client applications are modified to operate on protected resources through a simple Policy Enforcement Point which consults the Policy Server for grant/deny decisions based on the current active policy.

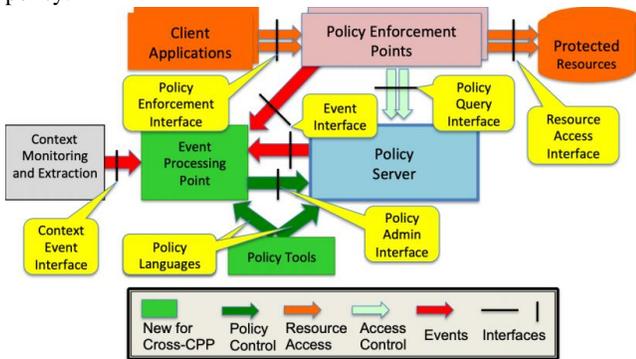


Figure 4: Cross-CPP Functional Architecture of NGAC extended for context sensitivity

3.6 Context Monitoring and Extraction

The monitoring and extraction module extracts context from the CPP use to support security and improve services.

The context sensitivity of the Cross-CPP ecosystem can be supported by monitoring and extracting information about the use of CPP, and can support the adaptation of services (by allowing the services to retrieve only the information that matches the context extracted) as well of the Cross-CPP Security module (adapting the access to resources by individuals based on the CPP use). Although the implementation of the Context Monitoring & Extraction is expected to be different depending on the source of the data, the approach followed will be the same and is explained below.

For each CPP, one has to define which concepts are relevant for the description of the situations (context), under which the CPP signals are generated and measured. Once of the concepts relevant to the description of context of CPP data streams generation are defined, the next step is to define the concepts which are relevant for the cross-sectorial service where it will be used. As a first approach, some general situations are considered (situations that could be interesting for a wide range of cross-sectorial services as well as for the Cross-CPP context sensitive security enforcement). This will translate in a context model that is in a first phase generic for all CPPs and on a second step specialised for each CPP type.

With the Context Model defined, the Context Monitoring and Extraction module will

- monitor the data that are needed to extract the context
- pre-process monitored raw data
- extract context by identifying the current context, based on monitored raw data, the current context model and historic con-text information stored in a context repository
- Based on the identified context, situations can be compared to previous ones and stored/passed on to other modules that may further consume this information (e.g. Cross-CPP Security)

The approach is best seen in Figure 5.

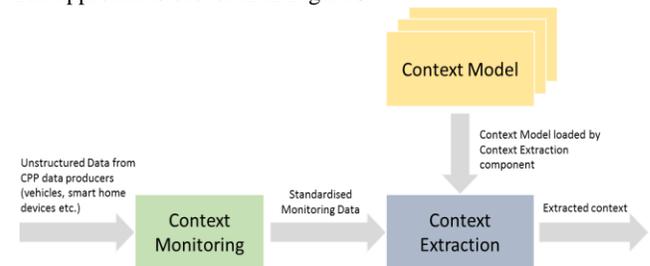


Figure 5: High level view of the functioning of the Context Monitoring and Extraction module

4 Industrial Application

The described ecosystem is applied by two data providers (CPP producers):

- Vehicles
- Smart infrastructure

and three data consumers (service providers).

As described within the previous sections, the easy and secure access to diverse data streams via a platform like the CPP Big Data Marketplace enables service providers to significantly enhance existing service solutions, and in some cases also create innovative new services, that have not been possible before. Within the Cross-CPP project, three main service areas are targeted to prototype this new approach: general weather forecasting services, weather-based navigation/warnings, and an e-charging service.

Weather forecasting service

Cross-sectorial data streams can considerably improve the forecasting quality of weather forecasting models as they can provide an unprecedented density of data points necessary for weather model initialization. Even in the case of only moderate sensor quality compared to the common meteorological sensors, new plausibility checks and analytics developed within Cross-CPP can process these data streams appropriately so they can be successfully ingested into a weather model to help its initialization and thus improve its resulting output. The conditions and means of successful data assimilation of a diverse range of sensors into an existing service state a challenge, that is addressed intensively by service providers within Cross-CPP and assisted by the Analytics Toolbox functionality. One approach to facilitate data ingestion is the development of a new high-resolution 100x100m weather model, which allows for a smoother incorporation of these new data points. Routine operation of such models is not known, nor is the inclusion of CPP data into such models. Despite that, a new *Plausibility Check* has been developed to test neighbouring data points from one source against each other (homogeneous check) as well as against data points from conventional sources (heterogeneous check). Despite the enhancement of weather models, the easy and secure access to new and diverse sensor data like wind shield rain intensity sensors and wiper data will be used to develop a virtual rain radar, an innovative new prototype service, that will mimic the main features of a radar map by using the live sensor data from vehicles. This service is especially useful in regions and countries that lack expensive radar stations (see e.g. European radar coverage in [18], and can also help fill gaps in times of radar outages.

The access to live vehicle safety related sensors like slippery data (road slickness data), to enhance slippery road detection can also be used to enhance weather-related warnings to vehicle owners nearby, and also enables the cross-check of other available weather data as especially conditions like freezing rain are still weather events hard to predict.

Furthermore, data from smart building weather stations are used within Cross-CPP to compute individually tailored weather forecasts for them in order to enhance automatic building operations like e.g. facade systems, window blind or support energy efficiency. For that purpose, each building receives their own weather model, that corrects the conventional forecast for usually unknown on-site specifics, which the model learns through the received CPP-signal. The improvement of such individual models compared to a forecast without this correction is measured by Mean

Absolute Error (MAE) between the DMO (direct model output) and the observation data. First tests have already revealed a significant error reduction. Hence, access to smart building data again helps to improve conventional forecasts.

Weather warnings

Another service enhancement made possible by a Big Data Marketplace is a weather-based navigation and the provision of on-route live weather warnings. Within Cross-CPP, a prototype for a weather-based navigation service will be developed which uses vehicle data as well as data from a meteorological service provider to offer an enhanced navigation which takes current and future weather conditions on the route of the vehicle driver into account. It will also provide a live weather warning mode as well as initial re-routing to avoid bad weather on the trip. These navigational service enhancements are not only of use for the private consumer, but especially for logistic industries and automated driving.

E-mobility charging service

Main idea of the service is to exchange information among data providers related to “E-Charging”, meaning vehicles will be providing information about their battery status and other information relevant during the charging process, and buildings about their e-chargers infrastructure – free charging locations and constraints.

The service is to send information about the presence of charging station inside of the building or located outside (public parking lot, airport, hospital) to the vehicle. The service is using real-time data in the communication online with the car / building about occupancy of e-charges placed outside of the building or inside (in the garages), as well as the vehicle’s own information about its capacity of the battery. This together with its current position and speed could possibly calculate time of arrival and to reserve an e-charger for this specific car. A future idea to extend this would be to use weather stations that may also possibly provide relevant data, that could be used for expected electricity generation calculation as well.

Service applications like these demonstrate the necessity and value of a secure and easy interface between data providers (CPP producers) and service providers to foster Big data related growth within the service sector and unlock currently merely touched business potentials.

5 Conclusions

In this contribution we have described the conceptual view of the work done in the Cross-CPP project.

The developed CPP Ecosystem Architecture was outlined starting from a first draft of the CPP Ecosystem Workflow, which defines the information flow between key stakeholders and Cross-CPP system modules. The pictured architecture concept summarises the Cross-CPP modules and its key software components and how they correlate in Cross-CPP. Especially the defined Ecosystem Architecture, with its detailed representation of modules broken

down into a first overview of needed software components, presenting a blueprint for the system development within the project implementation phase.

The project analyses the reuse of work from previous projects to analyses the hypothesis that data from diverse CPPs in different sectors may be made available and reused by different Service Providers to produce cross-sectorial services. Cross-CPP is overcoming the identified challenges by establishing a CPP Big Data Ecosystem, which will develop the following main characteristics:

- Brand independent concept, open for integration of diverse CPP data providers coming from different industrial areas, also providing a standardized cross industrial CPP data model, setting the basic structure of Cloud Storage(s) for the CPP data streams, which needs to be flexible enough to incorporate data coming from various industrial sectors.
- CPP Big Data Marketplace providing to service providers a single CPP data access point with just one interface (one-stop-shop), as well as support functionalities for easy data mining/analytics. By these means, data customers (Service Providers) just need to set-up and maintain one interface to gather diverse CPP data from different CPP providers. The Marketplace makes the Cloud Storage for CPP data streams seamless to any data consumer taking security, data ownership and data rights into account.
- Controlled access to diverse CPP data streams and optimal management of data ownership and data rights (covering data flow from CPP owners up to Service Providers), applicable to various cross CPP data streams.
- Win-Win value chain for all ecosystem partners, due to the fact that the costs for the ecosystem in place can be shared by a great many data customers, which will make a single service much more economical.

ACKNOWLEDGMENTS

This paper presents work developed in the scope of the project Cross-CPP. This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement no. 780167. The content of this paper does not reflect the official opinion of the European Union. Responsibility for the information and views expressed in this paper lies entirely with the authors.

REFERENCES

- [1] B. Schmarzo, *Big Data: Understanding How Data Powers Big Business*, John Wiley & Sons, 2013.
- [2] W. H. Inmon und D. Linstedt, *Data Architecture: A Primer for the Data Scientist: Big Data, Data Warehouse and Data Vault*, Morgan Kaufmann, 2014.
- [3] J. Pillmann, C. Wietfeld, A. Zarcuła, T. Raugust und D. C. Alonso, „Novel Common Vehicle Information Model (CVIM) for Future Automotive Vehicle Big Data Marketplaces,“ in *IEEE Intelligent Vehicles Symposium (IV)*, Redondo Beach, CA, USA, 2017.
- [4] T. W. Jim Davis, „Towards Composable Manufacturing, Smart Apps and Services Marketplace,“ in *NIST/OAGI Workshop on Smart Manufacturing (SM) and Cyber-Physical Production Systems (CPPS)*, Washington, 2016.
- [5] „Smart Manufacturing Leadership Coalition,“ SMLC, 2017. [Online]. Available: www.smartmanufacturingcoalition.org. [Zugriff am 20.04.2017].
- [6] Automat Consortium Partners, „Deliverable 2.2: Overall Innovation & Technology Transfer Concept,“ 2015.
- [7] ATOS, Automat consortium partners, „D4.3 Full Prototype of Vehicle Big Data Marketplace,“ 2017.
- [8] BMW, „BMW and MINI CarData, Tailored third-party services for bmw and mini drivers,“ [Online]. Available: <https://www.bmwgroup.com/en/innovation/technologies-and-mobility/cardata.html>. [Zugriff am 10.10.2019].
- [9] E. M. a. P. A. C. S. João Bártolo Gomes, „Learning recurring concepts from data streams with a context-aware ensemble,“ in *Proceedings of the 2011 ACM Symposium on Applied Computing (SAC '11)*, New York, NY, USA, 2011.
- [10] R. S. a. P. P. R. João Gama, „Issues in evaluation of stream learning algorithms,“ in *Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining (KDD '09)*, New York, NY, USA, 2009.
- [11] A. Z. a. S. K. Mohamed Medhat Gaber, „Mining data streams: a review,“ *SIGMOD Rec.*, Bd. 34, Nr. 2, pp. 18-26, June 2005.
- [12] Juniper Consortium Partners, „Juniper Project website,“ 2019. [Online]. Available: <http://www.juniper-project.org/>. [Zugriff am 28. August 2019].
- [13] T. Strang und C. Linnhoff-Popien, „A Context Modeling Survey, in Workshop on Advanced Context Modelling, Reasoning and Management as part of the Conference on Ubiquitous Computing,“ in *The Sixth International Conference on Ubiquitous Computing*, Nottingham, 2004.
- [14] D. Stokic, S. Scholze und O. Kotte, „Generic Self-Learning Context Sensitive Solution for Adaptive Manufacturing and Decision Making Systems,“ in *ICONS 2014*, Nice, 2014.
- [15] S. Scholze, J. Barata und D. Stokic, „Holistic Context-Sensitivity for Run-Time Optimization of Flexible Manufacturing Systems,“ *Journal Sensors*, Nr. 17, p. 455, 2017.
- [16] Cross-CPP Consortium, „D1.3 Public Innovation Concept,“ 2019.
- [17] International Committee for Information Technology Standards, Cyber Security technical committee 1, „Information technology – Next Generation Access Control – Functional Architecture,“ INCITS 499, 2013.
- [18] Huuskonen und e. al., „The Operational Weather Radar Network in Europe,“ *Bull. Am. Met. Soc.*, 2014.