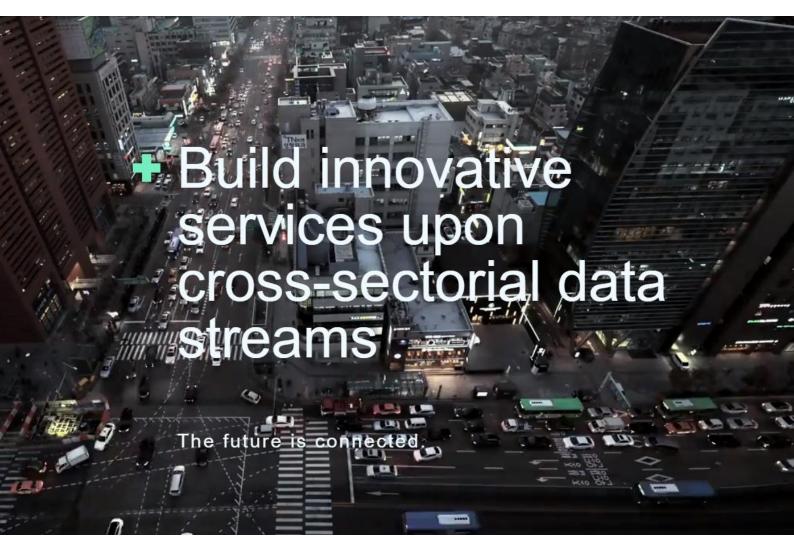


Cross-CPP Ecosystem

White Paper



November 2020





Foreword

Welcome to our Cross-CPP Ecosystem 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 manufacturers (e.g. car makers) to open up their products, by designing a convincing trustworthy Ecosystem.

Recently we have finalized the implementation and integration of the Cross-CPP Ecosystem and have started its validation and assessment by data providers and data customers. Several public presentations of our Data Marketplace solution and Analytic Toolbox have been presented a short time ago in the BDVe Webinar 'How to monetize your data in an open data Marketplace', which is still accessible via the BDVe website.

In this White Paper you will find some more details about the Cross-CPP ecosystem concept as a whole, leading from today's constraints in cross-industrial data access to how Cross-CPP will face those challenges with its innovative ecosystem concept.

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.

In this White Paper, we present to you the key innovations of the Cross-CPP ecosystem concept to face those challenges. These key innovative features concern the *Standard Data Model* fundamental for the brand independent data exchange, the *one-stop shop* concept represented by the Cross-CPP Marketplace acting as the central mediator for the data exchange between data provider and data consumer and the basic data *privacy and security* concept granting full control to the data owner over his data.

The following key points will refer to those main innovative features and show up the high future potential of the Cross-CPP ecosystem concept:

	4
lution to Meet the Challenges	6
oss-CPP Ecosystem	9
all kind of data	9
xploit and control their most valuable data assets	10
ectorial data	11
	12
	lution to Meet the Challenges oss-CPP Ecosystem all kind of data exploit and control their most valuable data assets

Once brought to life, *all* involved stakeholders - service providers as well as data providers and data owners - are in a win-win situation and will profit from the whole Cross-CPP ecosystem.



1 Challenges on the way to build innovative services upon cross-sectorial data streams

Key *motivation of Cross-CPP* project is to give cross-sectorial industries access to the great spectrum of sensor data coming from high volume products from various industrial sectors (vehicles, smart home devices, etc.). With the increasing number of connected sensors and actuators within such mass products, this number will rise in short-term. This enormous amount of data continuously generated by mass products will represent:

- 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
- 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 applications.

Existing CPP Data Access Constraints

However, these business potentials are currently still locked since manufacturing industry producing mass products has not established clear models and tools for such cross sectorial collaborations. Current solutions and offerings in the CPP data domain are driven by OEM specific business approaches that are almost solely focused on their own products and are realized by proprietary solutions. The resulting brand-specific solutions mostly do not provide CPP data to the outside world, hindering long-term value creation by service providers due to fragmented environments and the lack of brand-independent representation of CPP data. Due to this high level of market fragmentation, the situation today is characterized by far too complex and individual value chains resulting in economic inefficiency.

This situation is mainly characterised by the following major difficulties for both, the service providers (data customers) and for the CPP manufactures / owners (data providers):

- 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 ecosystems with
 different data models, causing high efforts and costs for data collection and processing.
- Limited possibilities to use cross-sectorial CPP big data streams, due to missing
 functionalities for an easy access and detection of needed data, as well as of easy to use
 big data analytics functionalities for Service Providers with low big data expertise and
 knowledge.
- Missing preconditions to establish such cross-sectorial data market, as for e.g. the
 absence of an agreed common data model for CPP data coming from various industrial
 sectors, or mechanisms for an optimal management of commercial confidentiality, privacy,
 IPR and ethical aspects applicable to various cross CPP data streams.
- Uneconomical brand-specific service platform solutions, causing high costs for the
 implementation of proprietary data platforms, which have to be financed and justified by
 just a few CPP manufacturer specific services (no single service can bear the cost for
 such system solution -> ROI is not achievable). Thus, today services will be often offered
 for high price not acceptable for the majority of customers.
- Wasted Innovation Potentials by thousands of external experts due to closed brandspecific data platforms. Even very big CPP Manufactures will by far not have the expertise and innovation potentials of the world-wide network of service providers. By that reams of new innovative services remain closed.



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

On the contrary to today's sporadic proprietary CPP ecosystems, which are in most cases restricted to CPP manufacturer specific services and which are not open for third parties interested in these CPP data, the Cross-CPP Ecosystem targets to bring sensor data from CPPs of various industrial sectors to the outside world. Therefore, as key challenges, Cross-CPP has to overcome several obstacles by establishing a CPP Big Data Ecosystem, which has to face the following key challenges:

- Brand independent concept, open for integration of diverse CPP data providers coming
 from different industrial areas, also providing a brand independent standardized cross industrial CPP data model which needs to be flexible enough to incorporate data coming
 from various industrial sectors.
- A Mediator Platform (called the 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.
- Controlled access to diverse CPP data streams and optimal management of data ownership and data rights, 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 large amount of data customers, which will make
 a single service much more economical.



2 The Cross-CPP Ecosystem Solution to Meet the Challenges

How a system solution has to looks like, which can satisfy the stated challenges and which provides the stated features to the stakeholder of the Cross-CPP Ecosystem?

Starting Point of the Cross-CPP System Development

The Cross-CPP Ecosystem development had not to reinvent the wheel and could build upon the results of the past and current projects ProSEco, AutoMat, IASIS, Resilience 2050, and Juniper.

Specifically, results from the project AutoMat, which has established a novel and open Ecosystem in the form of a cross-border Vehicle Big Data Marketplace that leverages currently unused information gathered from connected vehicles, was used as a basis for further development. The validity of the developed brand independent AutoMat concept and standardized common vehicle data model has been proven in the scope of project for the automotive industry by three leading car manufactures (Volkswagen, Renault and FIAT).

In contrast to AutoMat project, which was focused exclusively upon automotive industry, a <u>key mission of Cross-CPP</u> is to analyse and extend the AutoMat concept with regard to the integration of data streams coming from other CPPs than vehicles.

To achieve this transition from AutoMat to a "Cross-Sectorial Ecosystem" the following challenges have to be faced:

- Enlarge the big data pool by providing data streams from various industrial sectors such as automotive, home automation systems and other industries.
- Extend Cross-CPP Data Model, for data streams coming from multiple industrial sources, applicable for various industrial sectors.
- Empower data owners to exploit their most valuable CPP data (vehicle data, smart home device data etc.), simultaneously ensuring that the owner can fully control which data should be harvested and provided to which Service Provider, always under consideration of their commercial confidentiality, privacy, IPR and ethical aspects.
- To offer advanced one-stop-shopping support for Service Providers, especially for SMEs
 or start-ups, by providing instruments enabling an easy access, detection and selection of
 required data, as well as flexible cross-stream analysis tools for large data volumes.

The Three Pillars of the Cross-CPP Ecosystem Structure

The Cross-CPP Ecosystem solution is characterised by an open system approach, enabling that any Data Owner can bring his sensor data to the outside world, to be used by any Service Provider. Thereby, the Cross-CPP System Solution can be grouped into three main pillars, as presented in Figure 1:

- The Left Pillar represents Data Providers (CPP Manufacturers/Data Owner) side, comprising the data harvesting and making CPP data available to the Data Customers in a standardised format.
- The Right Pillar represents Data Customer/Service Provider side from Cross-sectorial industries or manufacturers of CPP using the provided CPP data from various products to create new value out of that data by improving services or establishment of diverse new cross-sectorial services.
- The Middle Pillar represents the mediator between the Data Provider and the Data Customer, responsible for the data handling between both parties. This comprises the storage



of the data in a Data Owner's storage vault. Furthermore, the so called Big Data Marketplace handles the controlled access to the stored CPP data from the various Data Owners and forwards the data to the respective Service Providers.

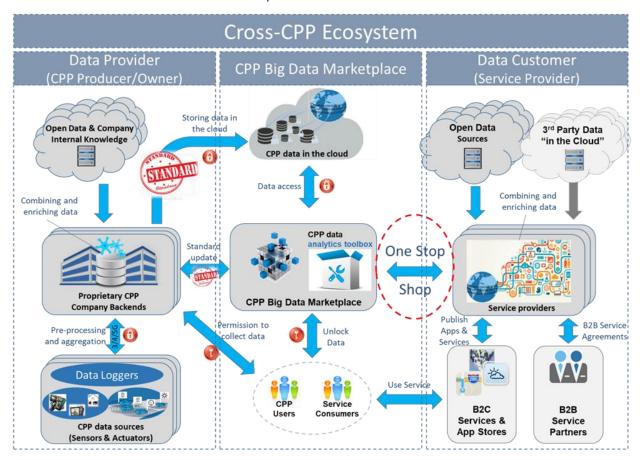


Figure 1: The three Pillars of the Cross-CPP System Structure

For a better understanding of the Cross-CPP Ecosystem Structure in reference to Figure 1 the three pillars are explained in more detailed:

Left Pillar - Data Harvesting and CPP Company Backend

Data Harvesting:

The main role of the Proprietary CPP Data Harvesting is to acquire the data from a proprietary CPP data source and send them to CPP Company Backend. To ensure that the data harvesting process is in consent with end user needs, a data acquisition configuration is downloaded from the CPP Company Backend and deployed in the CPP data logger. CPP data will be measured and stored as CPP data packages. In order to reduce the amount of information to be transmitted the CPP signals can be pre-processed and aggregated. These CPP data will be transmitted to the CPP Company Backend module according to the transmission strategy defined by the CPP manufacturer.

Company Backend:

It represents the central data access point to the CPPs for thousands to millions of CPPs of a brand. It will be based on a proprietary brand specific solution, interpreting and transforming proprietary manufacture-specific CPP data into physical information in reference to agreed owner permissions. Furthermore, the information will be validated and can be masked to enforce privacy. Finally, the information is converted into the required guasi standard data format, the Unified



Cross-Industrial Data Model (CIDM) and is published to the owner's CPP Cloud Storage. The CPP Company Backend has to handle also the consent of the data owner for data harvesting.

Middle Pillar - Cloud Storage and Big Data Marketplace

Cloud Storage:

Cross-CPP targets a cloud based concept for the storage of CPP Big Data. Any Data Owner holds a data vaults explicitly for his data. The storage infrastructure enables the data reception from the CPP Manufacture Backend, as well as data access by the CPP Big Data Marketplace in the standard CIDM format. Furthermore, storage management has to prevent data from unauthorized manipulation, ensure data completeness and the protection of the privacy of data owners, etc. The storage management is also able to handle a CPP Data Model updates.

Big Data Marketplace:

The Big Data Marketplace embodies the central element of the Cross-CPP Ecosystem approach representing the mediator for all the actors involved in the data flow of the Cross-CPP Ecosystem.

For the data customer (service provider) the CPP Big Data Marketplace represents a "One-Stop-Shop", the single point of access to data streams from multiple mass products. Therefore, the marketplace offers instruments enabling an easy access and detection of needed data, as well as a data analytics toolbox and context features, which will provide easy to use big data analytic functionalities for Service Providers with low big data expertise and knowledge. Furthermore, in reference to the agreements between the Data Owners and the relevant stakeholders during runtime the Marketplace will control the data exchange between Data Owner and Service Provider.

Right Pillar - Data customers & Service providers

Right pillar covers the Data Customer/Service Provider representing cross-sectorial industries or manufacturers of CPP using CPP data to improve services or to establish diverse new cross-sectorial services. The Marketplace supports the Service Provider by enabling access to a data catalogue and tools for the selection of relevant data, by setting up data order, as well as by the access right management with the data Owner.

Once all contractual agreements for the data access by service providers have been arranged, for the run-time phase the Service Providers have to enable to bring the cross-sectorial data streams in the CIDM format to their worlds. Based on specialized algorithms for the various services, also addressing features such as data cleaning, filtering, etc. and transforming the CPP information into service relevant input information, the final service products are generated.



3 The Key Innovations of the Cross-CPP Ecosystem

In the following the basic essentials of the Cross-CPP solution concept are presented mandatory to meet the challenges stated above. These key innovative features concern the Standard Data Model fundamental for the brand independent data exchange, the one-stop shop concept represented by the Cross-CPP Marketplace acting as the central mediator for the data exchange between Data Provider and Data Consumer and the basic data privacy and security concept granting full control to the Data Owner over his data.

3.1 One common standard for all kind of data

The key innovation of the Cross-CPP solution to realise a brand independent data exchange between the Data Providers and Data Consumers embodies the Common Industrial Data Model (CIDM). Except the CPP Manufacturer specific data formats along the data harvesting, the data are transformed into the standard CIDM format before they are transferred to the storage area. In the CIDM format the data will be retrieved from the Cloud Storage.by the Marketplace and in this format forwarded to the Service Providers.

The CIDM represents an open and highly scalable big data format, designed to harmonize proprietary data harvested by the CPP Manufacturer into generic datasets. The CIDM is flexible enough to incorporate data coming from multiple industrial sources, applicable for various industrial sectors.

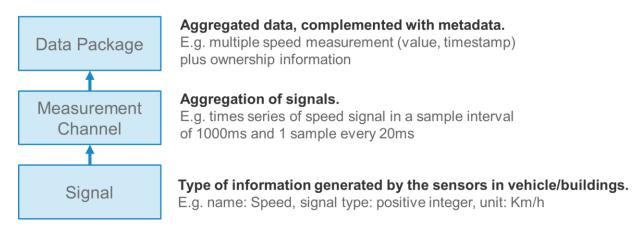


Figure 2: Basic Level Structure of the CIDM

The structure of the CIDM consists of three layers (see Figure 2):

- Starting from the bottom part, Signals describe the type of physical phenomena and chemical quantities of vehicles and buildings etc. 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, the data from the CPP need to be pre-processed and aggregated. The result constitutes a "measurement channel" that include the signals to aggregate, 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. In addition to the data, data packages also contain metadata with support information like ownership and quality assessment. The data packages are forwarded to the cloud store vault of the Data Owner.



However, the CPP Data Model is not rigid, rather representing a living data structure. In reference to the needs of the Service Provider community the number of signals to be recorded, as well as the type of measurement channels can be modified or extended. Based on the decision of a Standardization Board a required update of the actual version of the CIDM by additional signals/channels might be concluded.

3.2 Empower data owners to exploit and control their most valuable data assets

A key innovation of vital importance for the Cross-CPP Ecosystem solution is to follow the General Data Protection Regulation (GDPR) approved by the European Parliament, the Council of the European Union and the European Commission. The GDPR protects people's personal data throughout the European Union (EU). The decree also affects data exports from the EU. It enforces the right for people to lawfully agree with companies to use their private information. It also enforces the right for people to have their private information no longer accessible by a company.

- Data owners may take advantage of data masking features that reflect their privacy requirements
- The Cross-CPP system provides also technical measures capable of enforcing data controls corresponding to the nuances specified by the contracts



In reference to the GDPR, Cross-CPP is aiming to give citizens full control over their data so that the Data Owner always is aware about which of his data are accessed, where they are stored and who is using which of his data for what purpose. Moreover, the Data Owner is able to deactivate specific data accesses at any time. This control of the Owner's personal data causes the necessity to formulate contractual regulations for the interaction between Data Owner and the various stakeholders of the Cross-CPP Ecosystem guaranteeing that the citizens has full control over their personal data. This is implemented by the following introduced key contractual regulations between Data Owner and respective stakeholders:

- CPP Data Owner and CPP Manufacturer: The CPP Manufacturer gets the CPP owner's permission to gather, pre-process and forward data to his storage vault in the cloud.
- CPP Data Owner and Storage Provider: The CPP Owner rents a cloud storage vault explicitly to store his CPP data.
- CPP Data Owner and Service Provider: The Data Owner permits each Service Provider to use his data required to generate the services.

These contractual agreements represent the reference for the Marketplace to control which data of the respective Data Owner are allowed to be retrieved from his storage vault and which of these data can be forwarded to a specific Service Provider. These permissions can be withdrawn by the Data Owner at any time.

The implemented Security and Policy Enforcement system of the Cross-CPP solution provides the functionality to define CPP privacy rules and contract privacy rules. The system enforces the



privacy rules when the service provider consumes the shared data. In order to integrate the functionality provided, the elements created for the integration are the API Client, the sails service and the sails models and controllers.

A Context Sensitive Security Framework of the Cross-CPP Ecosystem ensures that specific security policies are enforced so that access by client applications to data, objects, web services or other protected resources are mediated by policy enforcement points that determine whether access should be granted according to access control policies maintained in a policy server and subject to interpretation according to monitored context. Policy enforcement points may be distributed around a Cross-CPP deployment and the framework is scalable, capable of supporting multiple security policies simultaneously, and adaptable to changing technology, organisational restructuring, and increasing data volumes. Furthermore, a secure data communication between the various stakeholders of the Cross-CPP Ecosystem is guaranteed by the implementation of respective encryption/decryption commercial tools.

3.3 One-Stop-shop for cross-sectorial data

In the Cross-CPP Ecosystem the **CPP Big Data Marketplace** represents the "One-Stop-Shop", the single point of entry for a brand-independent access for Service Providers to create new B2B and B2C data-based products and services. The Marketplace is the meditator for all the actors involved in the flow of the data in the Cross-CPP Ecosystem. Thereby, the Marketplace knows which stakeholders are participating in the Cross-CPP Ecosystem, which data from which data owner are harvested by which CPP Manufacturer and used by which Data Consumer.

In this respect, the Marketplace, in turn, will maintain the mapping between CPP Owners (e.g. CPP_ID, which signals/channels), CPP Manufacturer (e.g. ID, data forwarded storage vaults), Storage Providers (e.g. ID, CPP Owner ID, data available) and Data Consumer: (e.g. registration, data request, data Owner access right). This central stakeholder interaction management by the Marketplace represents the basic concept of the Cross-CPP Ecosystem in order to carry out the access control of the CPP Owner's data by the various data consumers at run-time, as well as being able to answer accurately future requests by any Service Provider.

CPP Big Data Marketplace offers to cross-sectorial Data Consumers, the possibility to search for more than 200 sensor signals, display advance visualization representations (Histograms, Geo-Histograms, Time Series etc.) and retrieve those datasets in a seamless experience thanks to the open SDK-API created.

CPP Big Data-Marketplace functionalities include:

- Provision of a wide data catalogue and delivery of statistics for Service Providers.
- Discovery service of requested data and identification of data owners.
- Management of permissions (future smart contract provision service)
- Secure and reliable end-to-end communication (from the cloud to the Service Providers).
- Transactions accounting.

In the scope of the Data Discovery the Marketplace offers Service Providers a tool to select which data they want to receive from CPP owners. A wide variety of filters are provided in order to help Service Providers to narrow their desired results. Furthermore, a context service is provided. When the Service Provider has selected one or more channels he can also request suggestions of in this context commonly used channels. This is possible thanks to the implementation of the Context Monitoring and Extraction module (CME) within the Data Discovery process. In case



suggestions are requested, the CME will provide a list of channels related to those already selected¹. Channels of interest can be added directly from the suggestion list.

In respect to a stated service provider request to access specific CPP data sets the CPP Big Data Marketplace identifies the relevant Data Owners and supports to manage their permission for the data access by the Service Provider.

During runtime in respect to the data request of a Service Provider the access authorisation will be checked and the relevant data storages identified. The retrieved CPP data from the storage cloud will be decrypted, verified, anonymised where needed and delivered to the Service Providers. In the scope of these activities also an accounting of the data access will be realised as reference for any billing process.

3.4 Find the right data for You!

It may be complicated to get access to huge amount of data. Yet, is this the only challenge? Surely not. As it is well-known in the machine learning community, having data is not tantamount to having knowledge. The Analytics Toolbox simplifies the extraction of the latter, by providing a set of libraries and modules designed to satisfy most data-related needs, and based on the most recent concepts and algorithms developed by the scientific community. It is buttressed by a modular structure, in which new analytics services can be added to fulfil new requirements; and in which multiple algorithms can be chained together, to give answer to even more complex questions.

But how will this Analytics Toolbox help you?

- By enabling fast prototyping. No data to download, no library to develop and deploy inhouse. The Analytics Toolbox enables performing a first feasibility evaluation of a new business idea at essentially no cost.
- By unleashing the power of advanced algorithms. The Toolbox includes modules not easily available in other all-purpose analytics solutions, and specifically designed with CPP data in mind. These span from the analysis of thousands of trajectories, to the representation of network relationships. Again, no in-house development is required: the Toolbox includes everything is needed for a first evaluation.
- By minimising overheads. Filter your data prior to download, for instance through averaging, clustering, or through event-driven triggers. Only download what you need, and when you need it.

The data analytics toolbox, will provide easy to use big data analytic functionalities for Service Providers, supporting an easy access, search/detection and selection of required data. Furthermore, the Analytics Toolbox simplifies the extraction of data, by providing a set of libraries and modules designed to satisfy most data-related needs, and based on the most recent concepts and algorithms developed by the scientific community. It is buttressed by a modular structure, in which new analytics services can be added to fulfil new requirements; and in which multiple algorithms can be chained together, to give answer to even more complex questions. The Analytics toolbox provides services such as:

- Basic statistics and data aggregation: from statistical metrics of location and dispersion, to analysis of distributions and entropies.
- Time series analyses, including the detection of drifts, or sudden changes, through the application of statistical and data mining models.

Suggestions are provided based on a context model that is the result of analysis of the physical relation between signals (e.g. in case temperature is a selected signal, the CME would suggest the humidity and sun intensity



- Trajectory analyses. From the processing of individual trajectories, including interpolations
 and error detections; to multivariate scenarios, as in the detection of clusters of similar
 trajectories see the image for an example.
- Network analyses, for understanding the structures created by interacting elements.

CPP Marketplace

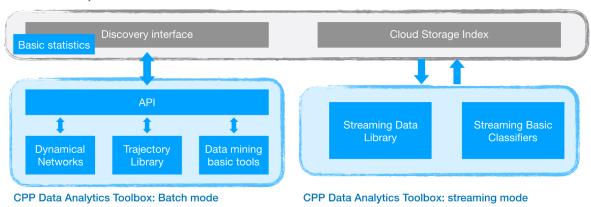


Figure 3: Structure of the Data Analytics Toolbox

During runtime in respect to the data request of a Service Provider the access authorisation will be checked and the relevant data storages identified. The retrieved CPP data from the storage cloud will be decrypted, verified, anonymised where needed and delivered to the Service Provider. In the scope of these activities also an accounting of the data access will be realised as reference for any billing process.





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 (nonindustrial, 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



Every effort has been made to ensure that all statements and information contained herein are accurate, however the Cross-CPP Project Partners accept no liability for any error or omission in the same.

© 2020 Copyright in this document remains vested in the Cross-CPP Project Partners.

