prostep ivip

Recommendation

prostep ivip Recommendation PSI 31

Cloud-based Production Collaboration

Challenges and Lessons Learned for Successful Implementations

Version 1.0

Abstract

This recommendation summarizes the results of the Cloud-based Production Collaboration (CBPC) project group. The objective of the CBPC project group is to evaluate possibilities and solutions for heterogeneous cloud-based collaboration scenarios within and between companies in the manufacturing industry with focus placed on production.

Disclaimer

prostep ivip Recommendations (PSI Recommendations) are recommendations that are available for anyone to use. Anyone using these recommendations is responsible for ensuring that they are used correctly.

This PSI Recommendation gives due consideration to the prevailing state-of-the-art at the time of publication. Anyone using PSI documentation must assume responsibility their actions and acts at their own risk. The prostep ivip Association and the parties involved in drawing up the PSI Recommendation assume no liability whatsoever.

We request that anyone encountering an error or the possibility of an incorrect interpretation when using the PSI Recommendation should contact the prostep ivip Association (psi-issues@prostep.org) so that any errors can be rectified.

Table of Contents

Figures

Tables

Abbreviations, Definitions, References

Abbreviation	Meaning
API	Application Programming Interface
ASE	Advanced Systems Engineering
bdd	Block Defininition diagram
CAD	Computer-Aided Design
CDLC	Cross-Discipline Lifecycle Collaboration
CPO	Code of Openness (open initiative of prostep ivip)
CWM	Common Warehouse Metamodel
DDP	Digital Data Package (working group in prostep ivip)
EMOF	Essential Meta-Object Facility (OMG Standard)
FMEA	Failure Mode and Effects Analysis
FMI	Functional Mock-Up Interface (Open-source standard)
FMU	Functional Mock-Up Unit
FTA	Failure Tree Analysis
GfSE	Gesellschaft für Systems Engineering (German chapter of INCOSE)
ibd	Internal block diagram
IF	Implementor Forum (groups in prostep iViP)
(e)HSUV	(Extended) Hybrid Sports Utility Vehicle
INCOSE	International Council on Systems Engineering
IoT	Internet of Things
IREB	International Requirements Engineering Board
IP	Intellectual Property
JT	Jupiter Tesselation
MBSE	Model-Based Systems Engineering
MDK	Model Development Kit
MMS	Model Management System
MOF	Meta-Object Facility (OMG Standard)
OEM	Original Equipment Manufacturer
OMG	Object Management Group

Abbreviation	Meaning
OpenMBEE	Open Model-Based Engineering Environment
OSLC	Open Services for Lifecycle Collaboration
PDM	Product Data Management
PLM	Product Lifecycle Management
ReqIF	Requirements Interchange Format (OMG Standard)
RFI	Request for Information
RFP	Request for Proposal
RQM	Requirements Management
SE	Systems Engineering
SmartSE	Smart Systems Engineering (working group in prostep iViP)
SOI	System of Interest
SoS	Systems of Systems
SSP	System Structure and Parameterization (Modelica Standard)
SpecIF	Specification Integration Facility (Open-source standard of GfSE)
SVG	Scaleable Vector Graphics (W3C Standard)
SysML	System Modeling Language (OMG Standard)
UML	Unified Modeling Language (OMG Standard)
UUID	Universally Unique Identifier
VDA	German Association of the Automotive Industry
VDI	German Association of Engineers
V&V	Verification & Validation
WF	Workflow Forum (working group in prostep ivip)
WLTP	Worldwide Harmonized Light-Duty Vehicle Test Procedure
WPx	Work Package x
XMI	XML Metadata Interchange (OMG Standard)
XMI-DI	XMI Diagram Interchange
XML	Extensible Markup Language (W3C Standard)
XSL	Extensible Stylesheet Language (W3C Standard)

1 Background and objectives of the group

1.1 Initial situation

Cloud technologies offer considerable potential for company-wide and value chain-wide collaboration in the manufacturing industry. Heterogeneous architecture concepts and forms of collaboration, however, present companies with major challenges. On the one hand, this leads to specific initiatives for developing digital twins and (internal) Industrial Internet of Things (IIoT) platforms and, on the other hand, to external platforms for exchanging data and information that are often customer driven. This poses major challenges, especially for companies with a large number of collaboration partners, and there is a need for harmonized IT services and interfaces as well as good practices for architectures and forms of collaboration.

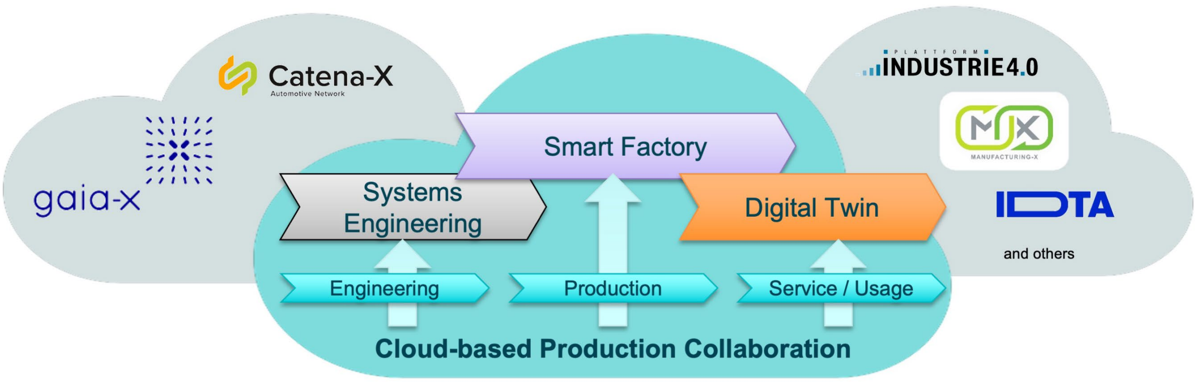

Figure 1: Motivation and scope of the CBPC group

1.2 Objectives

This need was also identified among the members of prostep ivip Association, and the Cloud-based Production Collaboration (CBPC) project group was established. The objective of the group is to evaluate possible forms of collaboration and solution patterns in production and encourage an exchange of information between the members of the project group via application scenarios and terminologies. The aim is to discuss the impact of cloud-based collaboration on different production areas and use cases in production:

- Requirements for new collaboration scenarios
- Internal versus external exchange of information
- Costs and benefits as well as lessons learned from existing implementations

Basic terminology is therefore presented first of all below. This followed by a discussion of possible collaboration scenarios and, in particular, their cloud-based architecture concepts, as well as experience gained during implementation and the challenges that implementation poses.

2 Terminology

The terminology used in the context of the cloud and the forms of collaboration in production based on it are subject to heterogeneous definitions and perspectives. Therefore, key concepts and terminology are first defined. The following consensus was reached within the group:

Cloud-based production collaboration is characterized by:

- horizontal and/or vertical collaboration between multiple stakeholders
- on the basis of shared data
- with the aim of creating added value in one or more processes in the value network

The envisaged added value depends on the use case and the stakeholders involved and is specified in the recommendation using practical examples. Specific characteristics and challenges of production are in particular:

- Brownfield systems with hardware lifecycles that are significantly longer than software lifecycles
- Heterogeneous IT and process landscapes that have evolved over time
- Data that needs to be interpreted and thus requires semantic enrichment
- Specific (process) models, databases, PLM/ERP systems

The collaboration scenarios of prime interest to the project group are presented in greater detail in section 3, while this section provides a more detailed description of the terminology used in the context of the cloud and edge computing.

2.1 Edge and cloud computing

A general distinction can be made between tasks and arithmetic operations that take place near the system (edge) or in the cloud. **Cloud computing** includes technologies and business models that make IT resources available dynamically and charge for their use based on flexible payment models. Unlike static and often limited IT resources (e.g. individual computers or servers), the availability of cloud structures is flexible and they offer scalable computing power. Cloud computing is often provided at competitive conditions by hyperscalers (Microsoft Azure, Amazon Web Services, Google Cloud, etc.) that offer specialization and economies of scale, but it can also be provided in-house in data centers and computing clusters. **Edge computing** describes processing units at the "edge" of a network, e.g. close to machines and systems. The edge devices are both logically and physically in close proximity to the data sources and subsequent applications and also provide the interface to network structures and thus also to cloud environments.

2.2 Private, public and hybrid clouds

A **public cloud** is an offering provided by a freely accessible provider that makes its services available publicly to anyone via the Internet. These services can be free or subject to a charge.

On the other hand, there are also **private clouds**. In this case, the cloud infrastructure is provided solely for the use of an organization (e.g. a company). The cloud can be owned, managed or merely operated by the organization. The reason for this type of provision is usually data protection or IT security.

Hybrid clouds are a combination of the first two cloud environments. Certain services are provided by public providers via the Internet, while applications that are data protection critical are operated within the company. The challenge lies in the systematic classification and separation of data and business processes into data protection critical and data protection non-critical processes.

2.3 Service level

A distinction can also be made with regard to the service level of a cloud-based collaboration scenario (Figure 2):

On premises: Infrastructure, runtime environments, operating systems, data and applications are all managed locally; there is no cloud-based collaboration.

Infrastructure as a Service (IaaS) describes the lowest service level of a cloud. Applications, data and the operating system are managed by the user. The provider merely provides the infrastructure, i.e. virtualized computer hardware resources such as computers, network structures and storage.

Platform as a Service (PaaS) clouds also offer access to programming and runtime environments on which users can develop or run their own software applications.

Software as Service (SaaS) clouds also provide these software applications and therefore represent the highest service level.

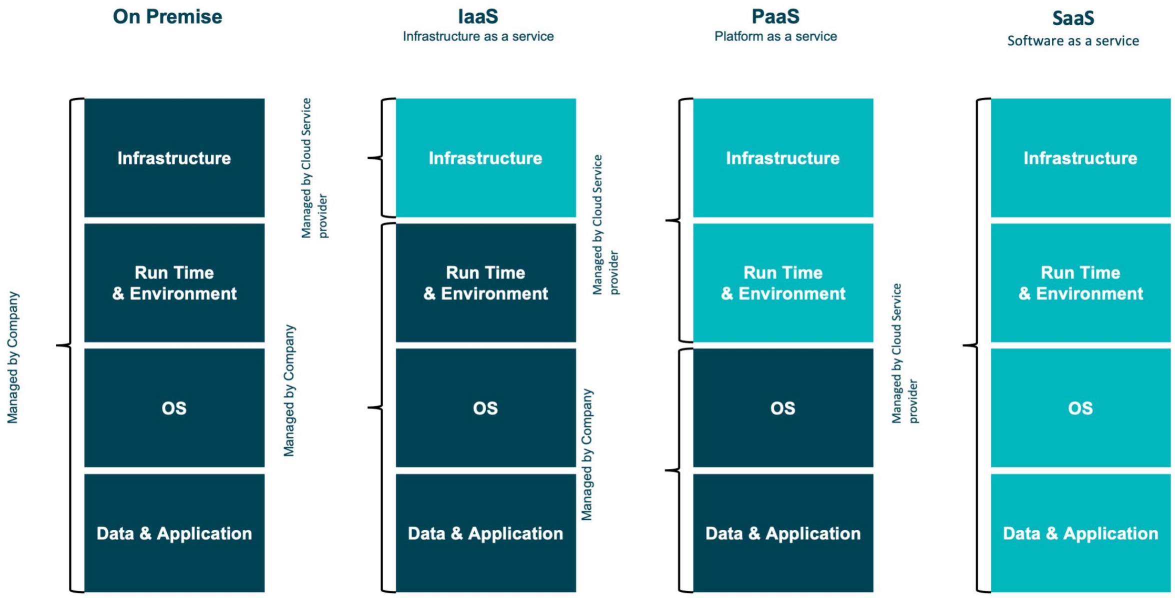

Figure 2: Service levels of cloud systems

3 Forms of collaboration

3.1 Horizontal and vertical forms of collaboration

Cloud-based systems make it possible to implement different collaboration scenarios. These can be categorized according to the degree of information exchange. On the one hand, information can be exchanged internally (within a company), for example to improve and analyze processes or implement services such as predictive maintenance. On the other hand, information can also be exchanged across companies, for example to create transparency regarding inventories throughout an entire supply chain and thus improve collaboration. Figure 3 provides an overview of different horizontal and vertical application areas for cloud-based production collaboration scenarios.

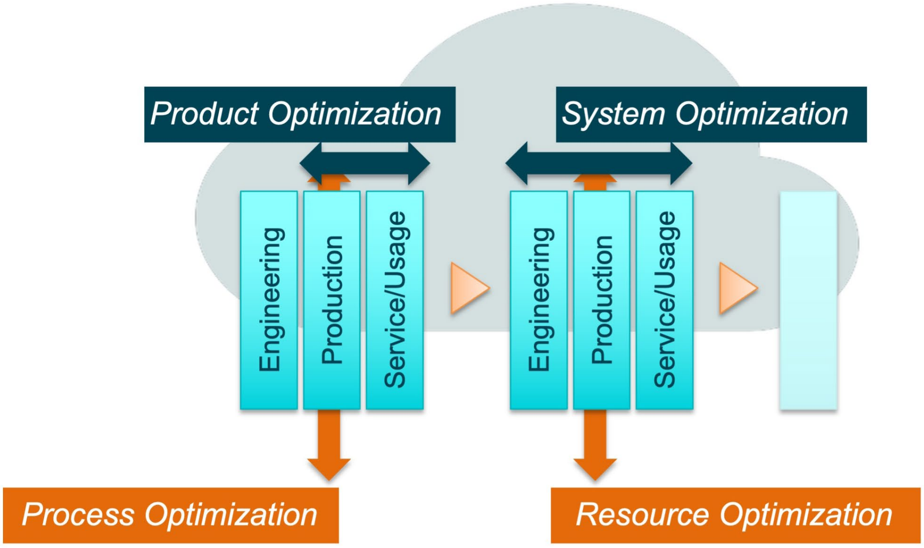

Figure 3: Overview of horizontal and vertical application areas for collaboration as per (Rauen et. al. 2018)

External cloud-based collaboration can, for example, not only take place with (parts) suppliers and customers but also with machine and system manufacturers, service providers and platform, infrastructure and software providers. Internal collaboration is possible across specialist departments and individual stakeholders such as management, IT, purchasing, sales and optimization teams all the way through to the cloud-based integration of different locations and production areas.

Throughout the product engineering process (PEP), there are also numerous forms of collaboration in product planning, product development, process planning, production system design, production through to the end of the product lifecycle, including aftersales, disposal and/or recycling. In the early phases of the PEP prior the start of production (SOP), manufacturing companies, suppliers, machine and system manufacturers, and integrators work together on heterogeneous data exchange platforms all the way through to product lifecycle management (PLM) systems. Cloud-based PLM platforms address the new developments.

Vertical applications can include:

- Condition monitoring of production on site
- Process optimization in production
- Analysis of quality problems and the cause of errors through to quality forecasts
- Predictive maintenance

- Analysis and forecasting of energy requirements (e.g. peak loads)

Horizontal applications include, for example:
- Requirements-driven inventory management
- Determination of the product carbon footprint (PCF) or digital product passport (DPP)
- Roll-out of scalable services and applications
- Cross-site condition monitoring
- Collaborative product and production system design

3.2 CBPC group's collaboration scenarios

The CBPC project group analyzed and prioritized 11 collaboration scenarios from the group members with the aim of identifying heterogeneous cloud-based collaboration scenarios and discussing specific challenges and lessons learned. The result was a selection of four collaboration scenarios (see Figure 4) – a more detailed description of the selected scenarios can be found in the appendix (see section 7.1).

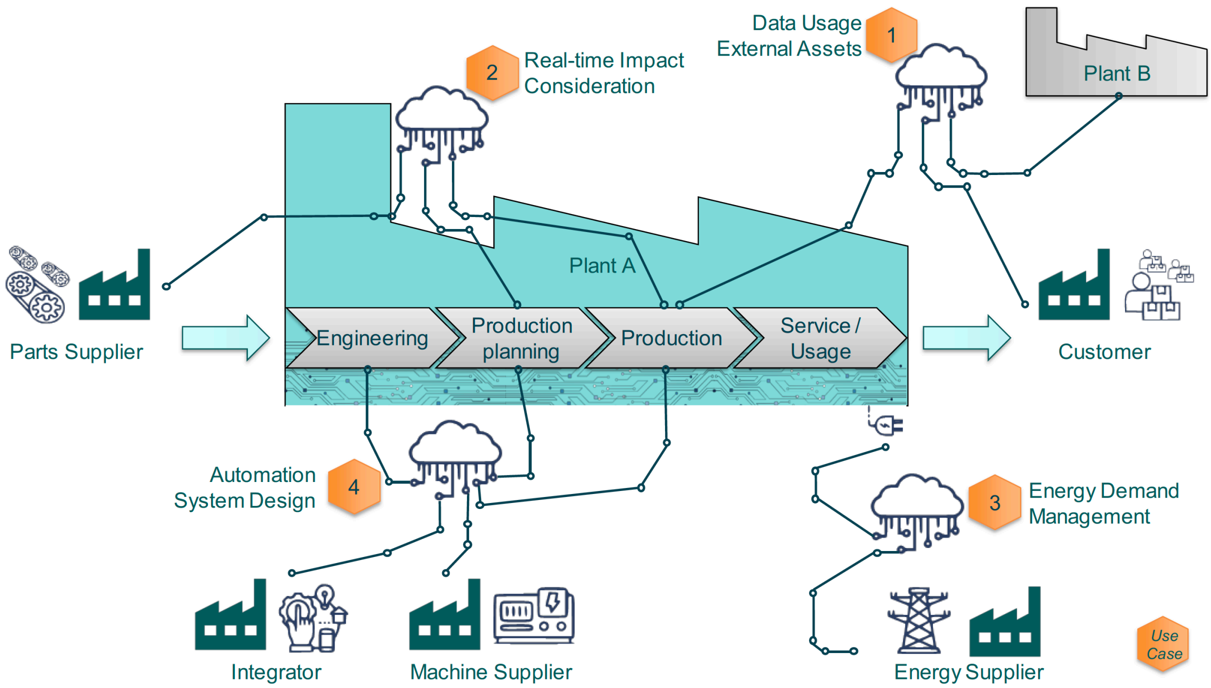

Figure 4: Overview of the CBPC group's collaboration scenarios

The application scenarios were subsequently analyzed in more detail together with the partners. Focus was placed on mapping product, process and resource data (PPR) as well as the interaction between IT systems and stakeholders who collaborate or provide data and information. The result is shown in the following figures. The group's focus was then further narrowed. The collaboration scenario Data Usage External Assets was selected for this purpose and enriched with other industry scenarios. The other collaboration scenarios are the subject of future work. The following section presents the implementations and versions of the use case at Dortmund Technical University's Institute of Production Systems (IPS) and, in the next step, at the industrial partners.

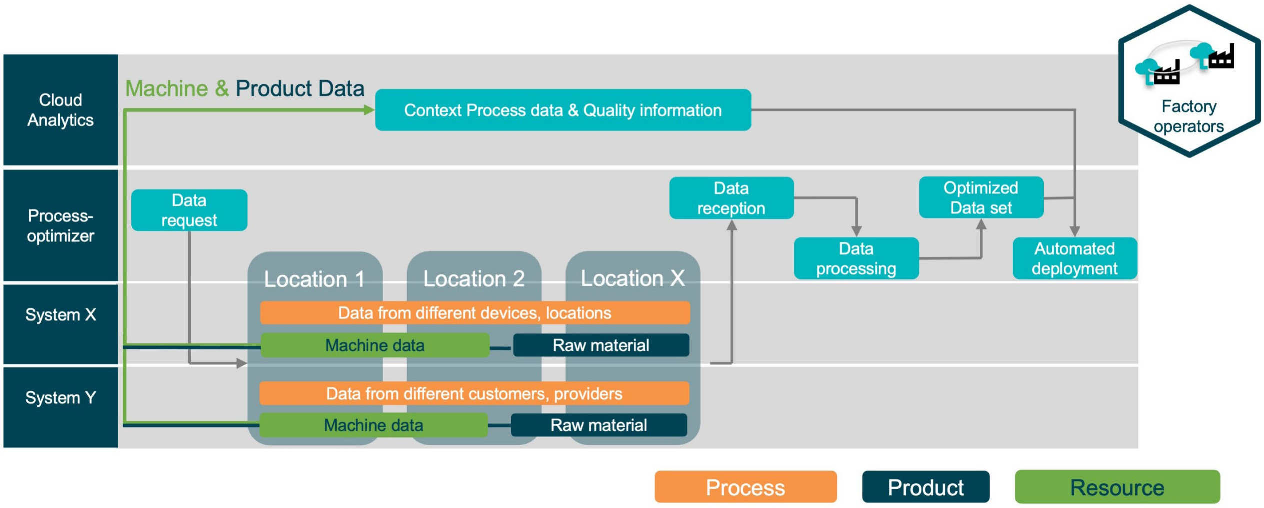

Figure 5: Specification of collaboration scenario: Data Usage External Assets

4 Collaboration scenarios examples

The use of cloud technologies requires new methodological approaches to collaboration and new architectures for this collaboration. In conventional scenarios, the exchange of data and information is often manual and proprietary, which leads to huge amount of time and effort, redundant data storage and synchronization problems. The collaboration scenario selected for initial specification in greater detail involves the use of external data to optimize the company's own products or processes. Different collaboration scenarios among the group members were considered in detail for this purpose. It is intended that not only the architecture but also the experience gained during implementation and the challenges posed by implementation be highlighted.

4.1 Data-driven recipe optimization in the cyber-physical brewing lab

4.1.1 Collaboration scenario

The use case focuses on the production of beer in a cross-location experimental brewery in Dortmund and in Sydney. The stakeholders involved as end users are the brewmaster or product planner as well as process engineers, IT departments, data scientists and customers, who provide subjective feedback on the recipe. The aim is to guarantee a high level of product quality, which is reflected by both objective criteria, such as the color of the beer or the original gravity, and subjective criteria, such as opinions regarding taste. It is also intended that the resources be put to optimum use to ensure, for example, that the right amount of malt is used to achieve the desired taste. The production processes and production parameters are mapped as digital twins with the aim of enabling system monitoring with early detection of deviations. The focus of the application is planning new recipes based on product and production data as well as customer analyses, thus promoting product innovations. The data is collected with the help of edge devices and merged in the cloud, and the application is used across all locations. The added value of collaboration therefore lies not only in the larger volume of data that can be used to train machine learning (ML) models but also in the scaled use of applications at multiple locations and the integration of customer and supplier data.

Each cyber-physical brewery (see Figure 6) comprises a 0.5 hl brewhouse and three fermentation tanks, each with a capacity of 1.2 hl. Each brewery executes the brewing process (mashing, lautering, boiling, whirlpool) and the fermentation and storage processes. The brewhouse comprises two kettles, an operating panel (HMI) for controlling the system, a pump and a piping system. In addition, temperature, flow and volume sensors are used for condition monitoring and automation. The fermentation tanks are also equipped with temperature sensors, a cooling unit and a variety of valves for temperature control. Production is batch-based.

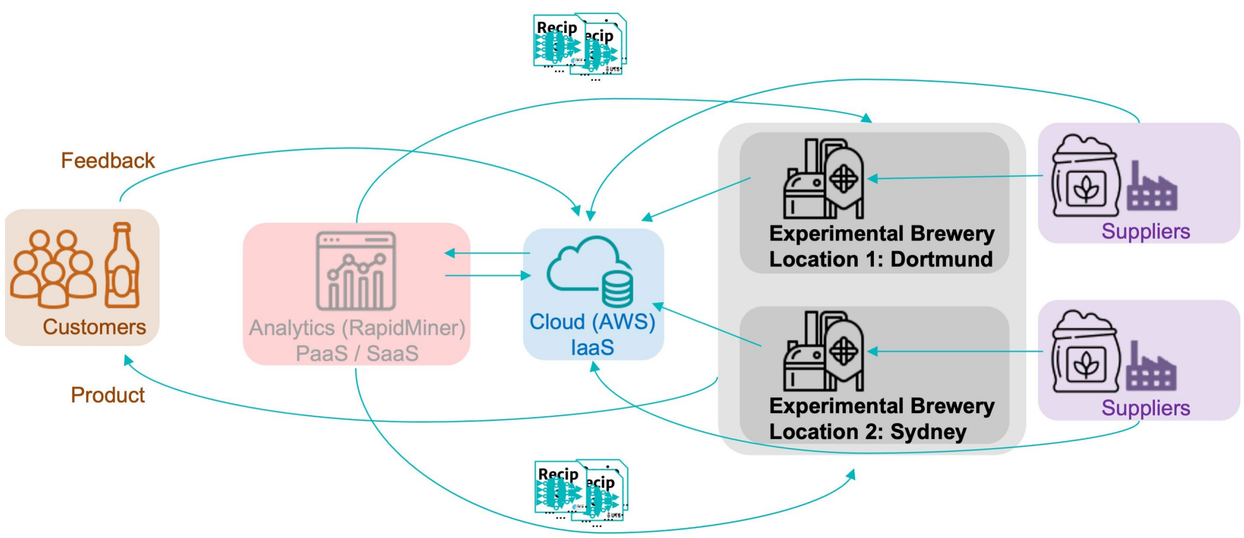

Figure 6: Overview of collaboration scenario: Recipe optimization in cyber-physical brewing lab

4.1.2 Implementation

The implementation was based on the process model established for the cross-industry standard process for data mining (CRISP-DM, cf. Chapman et. al. 2000). The first step involved the definition and specification of the target variables and possible influencing factors. The first major challenge encountered in the next step was integrating the large number of heterogeneous data sources and developing the collaboration environment. The data comes from a variety of on-premises IT systems. Of all the assets, the Zipmatic process control system (PCS) is the key system for defining recipes and controlling the automated processes as well as for status monitoring and process data acquisition. Aggregated process and system data is stored locally at each site in a PostgreSQL database belonging to the PCS. A Simatic ET-200EP PLC, which is controlled by the PCS via OPC-UA and supplies information to the PCS, offers a higher resolution and more direct access to the process data. Because the PLC only communicates bidirectionally with the PCS in its original state but does not have access to its database, an edge device was implemented that stores the raw data from the system. In addition, it is not possible for the system in its as-delivered state to record the fermentation process. Therefore, another external sensor (Tilt Pro Hydrometer) was installed, which continuously measures specific gravity based on its tilt angle. The data is stored in the SaaS solution Google Spreadsheets via the edge device.

Data collection is heterogeneous at the product level. Every supplier of yeast, hops and malt provides different product data sheets and laboratory analyses in PDF format. In addition, a number of different measurement methods are used for intermediate products in the brewing process. These include not only iodine tests but also density measurements (e.g. for different process states) using an Anton Paar EasyDens density meter, which exports values from the corresponding application as CSV files. A laboratory information and management system (LIMS) with a graphical user interface (GUI), which enables both manual product data creation and linking with the brewing process, was created to integrate the heterogeneous product data. The product data is augmented with the final quality assessment via sensory analyses. The data for the tastings is stored systematically on the Microsoft Forms platform.

The solution comprises a variety of different data sources and IT systems that need to be integrated. Central databases were implemented as an IaaS platform on the AWS Elastic Cloud to serve as a uniform database. This was then used as the basis for implementing a RapidMiner AI Hub as a central collaboration platform for recipe optimization, on which the interdisciplinary teams at both locations now work together (Wöstmann et. al. 2022)

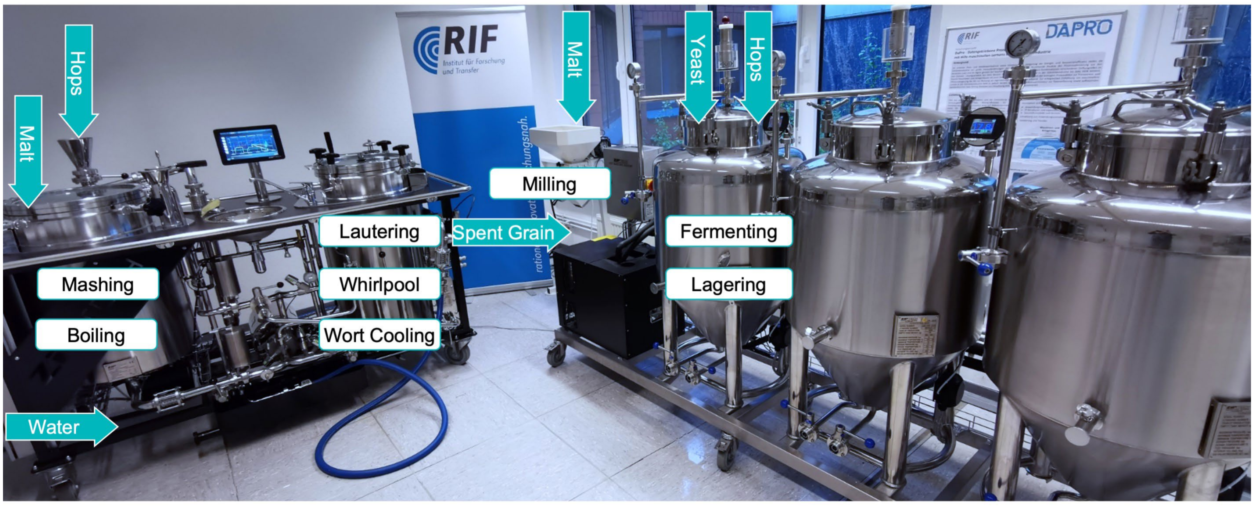

Figure 7: Cyber-physical brewing lab in Dortmund. A similar asset is located in Sydney

4.2 Cloud-based analysis of the Light Flextrack drilling processes

4.2.1 Collaboration scenario

During production, aircraft manufactured by Airbus move through structural, equipment and final assembly lines. Light Flextrack drilling robot systems are used in certain structural assembly lines for the production of A320 and A350 aircraft. This applies in particular to the more recent lines. Monitoring of the drilling process and an assessment of the logged operating steps performed by the Light Flextrack drilling robot systems on various aircraft from the A320 product family showed that, in some cases, there were both qualitative defects in the drilling process and positioning problems, which were revealed by repeated searches for incorrectly placed drill holes. Preparation and evaluation of the logs proved to be a time-consuming manual process. Insights gained from earlier preparation operations involving process and inspection data only slightly reduced the amount of work involved when repeating the tasks.

If the time required to evaluate each individual log and identify generally applicable patterns is to be reduced, it is important to minimize the effort involved in data preparation and modeling.

The application scenario was implemented in a heterogeneous team comprising domain and technical experts, a central orchestrator, data scientists and the IT department (Wöstmann et. al. 2021). At the heart of the collaboration lies better data continuity and cooperation between the heterogeneous stakeholders on a standardized platform. This can be used to derive reports appropriate to the different target groups. It is also possible to apply the analyses to a large number of other drilling processes.

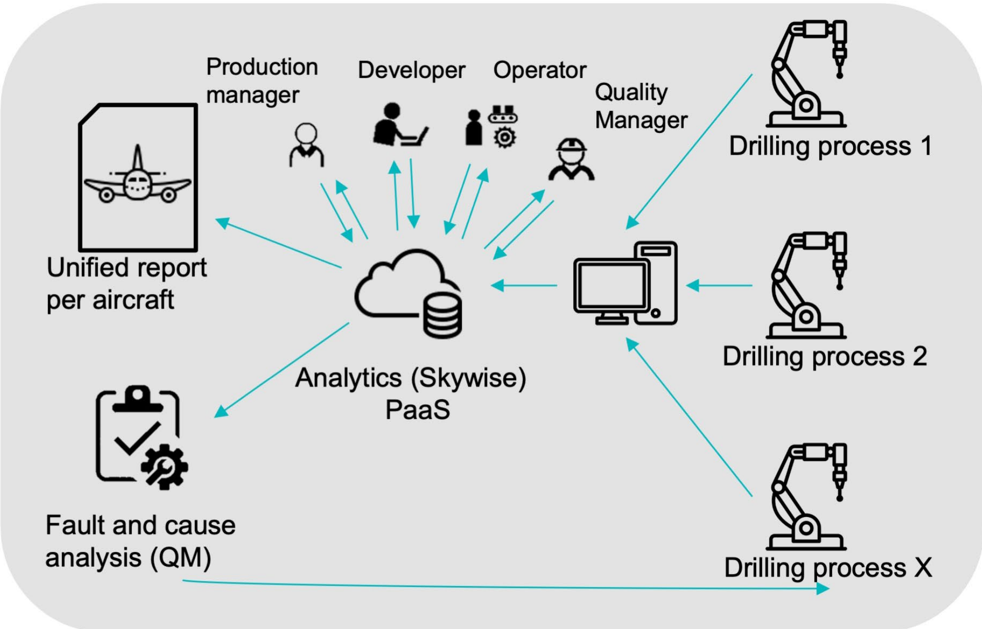

Figure 8: Overview of the collaboration scenario: Analysis of the Light Flextrack drilling processes

4.2.2 Implementation

The main aim of the project was to standardize and automate a data-analytical approach to evaluating the decentralized log files recorded by the Light Flextrack robots during operation. A standardized procedure and high level of automation are necessary because each aircraft is unique in its specific details due to customer-specific modifications, material properties and outside influences.

The newly developed analysis process ensures the automated upload of the Light Flextrack's logs to Airbus's Skywise data analysis platform, which is based on the Palantir Foundry platform. It is hosted on-premises as a local cloud and also serves as a PaaS for the specialist departments. Each log file is automatically formatted, cleansed and checked for plausibility. Automation of the data flow and the production of cleansed datasets were achieved by means of SQL scripts, data schemas and an identical log file header structure. The workload has been dramatically reduced thanks to minimization of the manual effort involved in data preparation and modeling. The standardized process, based on the cleansed datasets, has resulted in the automated generation of a report for each aircraft, which contains general information about the drilling process and recorded deviations as well as referencing difficulties and the possible reasons for these difficulties. The uniform datasets made available in this way also provide the basis for further data-based analyses for the optimization of the Light Flextrack algorithm.

Figure 9: Light Flextrack drilling robot system (MTM Robotics)

4.3 Quality forecasting in the production of sheet metal parts for OEMs

4.3.1 Collaboration scenario

The manufacturing process for sheet metal formed parts requires large and capital-intensive machines that can handle the forces involved. In addition, technology cycles are long and are primarily characterized by the use of new materials. This is why machines are operated for a long time and, if necessary, overhauled by means of a retrofit. Changes result first and foremost as the result of new product families that are defined within the company. The sector is characterized by long-term relationships with suppliers and local supply chains. The automotive industry, most notably, faces intense competitive pressure and exhibits reduced vertical integration in the value chain, which means that assembly activities in particular play a primary role. Core processes, e.g. in bodywork construction, are however carried out independently in order to safeguard know-how, for example. The large number of plants that use standardized processes means that data and models are easy to structure in a uniform way. It therefore makes sense for the OEM in this use case to operate its own platform and in the future optimize it for stakeholders involved in the collaboration, especially in light of intense competition.

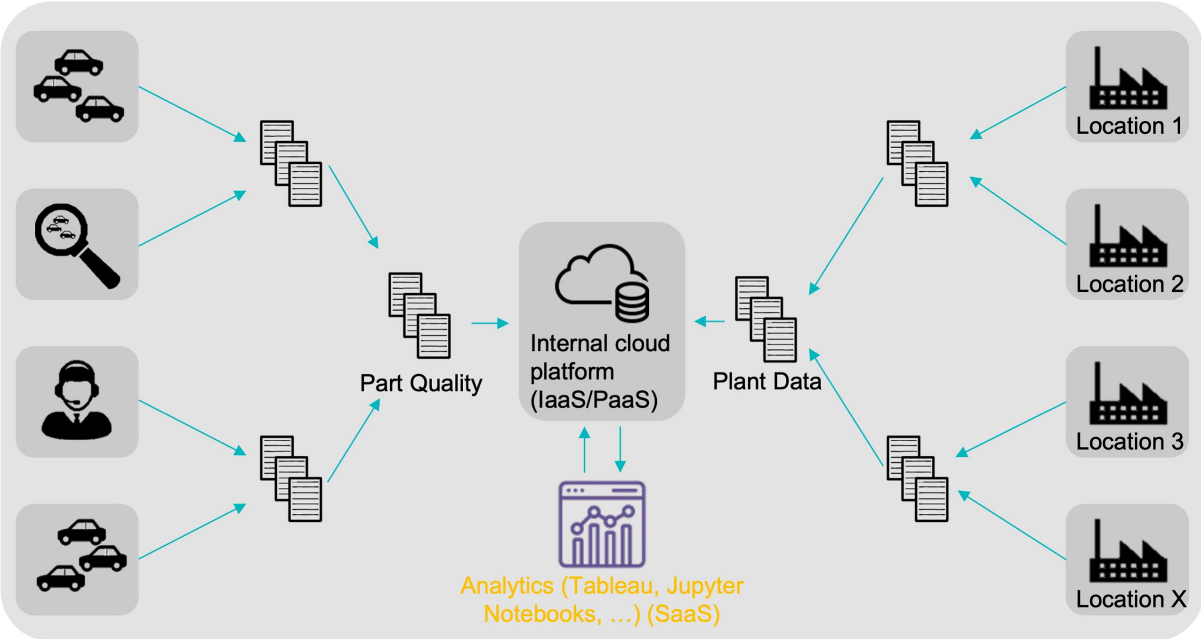

Figure 10: Overview of collaboration scenario: Quality forecasting in the production of sheet metal parts

4.3.2 Implementation

The OEM in the collaboration scenario provides comprehensive in-house services for planning MES systems and for data acquisition and analysis in production. Although the degree of digitalization in production is high, the also high proportion of information systems that have evolved over time means that the company is faced with the challenge of linking highly process-specific information from distributed data sinks (e.g. databases) so that it can be used effectively. In concrete terms, this means for example that data from machines rarely includes semantics, which leads to problems interpreting the data (e.g. conversion from metric to the Anglo-American system of measurement units). Discrepancies between production requirements and the technologies now available, particularly in the context of preparing data, pose additional challenges. In the future, it is intended that an abstraction layer that makes subsequent data processing easier will be established as a service (e.g. as an API-layer that ensures data quality). Standardized MES systems with interfaces, e.g. OPC UA, are used. Tableau is used for the visualization of prepared data from static information flows. Prototypical applications are implemented using Jupyter Notebooks and Pytorch. The company is subject to both standardized legal and internal guidelines in order to meet data protection requirements. In addition to complying with legal requirements, data also has to be classified internally. The implementations are therefore predominantly based on internally hosted private cloud structures, which are made available to the specialist departments as IaaS or PaaS. The aim is to make the applications available to users as SaaS.

4.4 Collaborative data use in transmission assembly

4.4.1 Collaboration scenario

The assembly of transmissions for the automotive industry is a variant-rich process that component manufacturers distribute over multiple assembly lines in different locations. Customer requirements, which vary greatly, result in a large number of variants that are covered by adaptive assembly systems. Assembly lines are not always designed in the same way, which means that the individual functions within the assembly system are also designed differently. The process chain itself, however, basically remains the same and is independent of the variant. The collection of data across all the final assembly lines offers an opportunity for finding optimum solutions and implementing them throughout the company, ultimately increasing quality and reducing costs. The aim is to access and collect data from different equipment – for automated, semi-automated and manual activities – and make it available to all the stakeholders in the company's production process. This can be used to derive two different collaboration scenarios from the perspective of the platforms:

The characteristics of the domain mean that the customers have great market power, which means that suppliers often only work for a small number of customers. It is therefore primarily the customers that specify the technology stack, which in turn means that customer-specific platforms dominate.

An alternative feasible scenario is driven by the iterative deployment of cloud-based technologies. In this scenario, parts of a production system are integrated step by step into higher-level digital systems, creating a large number of platforms. The next logical step is to interlink these heterogeneous platforms with the aim of generating added value across all the platforms.

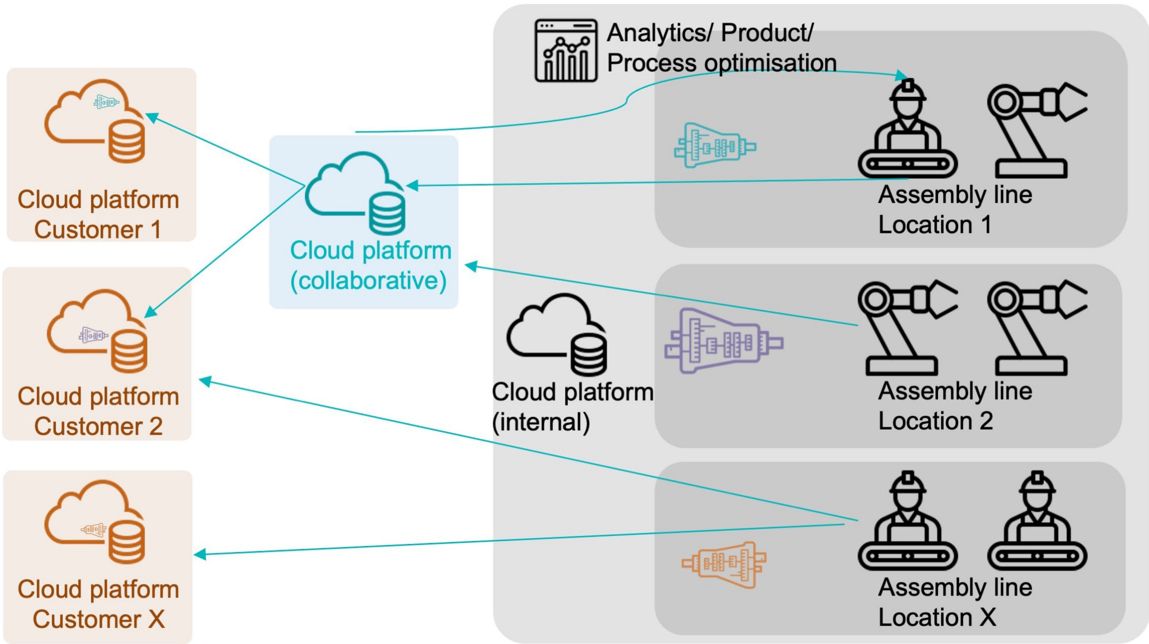

Figure 11: Overview of collaboration scenario: Data usage in transmission assembly

4.4.2 Implementation

Typical protocols are used at shop floor level to transfer data, depending on the degree of automation implemented and the equipment selected. PLC systems, for example, transfer their data to higher-level systems via OPC UA. However, the fact that the semantics used with the same protocols are not identical poses a problem when the data is consolidated and harmonized. The digitalization of every piece of equipment means that the connection to MES and SCADA systems is ensured and an initial database is available. The types of data involved are also different, which means that binary information such as OK/not OK, time series, nominal information (measured values, torques) can be grouped together and assigned to specific workstations with the help of IDs. Here again, the aim is to enable data-based analyses, which range from simple static evaluations for assessing individual processes through to AI applications for implementing predictive quality approaches.

5 Lessons learned from the implementations

In the second year of the project, the CBPC project group conducted an interview-based study with the partners involved in the collaboration scenarios with the aim of summarizing the experience gained and addressing the questions regarding costs and benefits and the lessons learned from existing implementations posed at the beginning. The structured evaluation is based on the CRISP-DM process model, which divides the implementation of the analytics projects on which the collaboration scenarios are based into six phases: business understanding, data understanding, data preparation, modeling, evaluation and deployment. In the following, the synthesized lessons learned in the individual phases are summarized and recommendations for action are derived in the context of the recommendation.

5.1 Business understanding

Organizational challenges

- Companies often lack a comprehensive strategic and organizational objective. This is characterized by a lack of overarching financial and human resources (e.g. project-independent/location-independent) for maintaining, integrating and making a basic stock of data available.
- The definition of interfaces and data structures at information level are needed if data from different production sites or machine manufacturers/variants is to be used in a uniform manner (harmonization of interfaces, protocols and semantics).
- The creation of a proof of concept (PoC) is helpful when dealing with new issues (especially in the context of analytical objectives). Here, for example, student employees, who have a high cost/benefit ratio when it comes to new challenges, can be integrated. It is recommended that the fact that a PoC may by all means deviate from the company standards regarding productive operation or minimum viable product (MVP) be communicated clearly. A successful PoC is often a prerequisite for the release of funds.
- Describing problems to financial backers may be difficult due to different perspectives or an investment backlog. A lack of transparency between developers/solution providers and those responsible for cost centers means that the investment needs for higher-threshold implementations, which cannot be carried out using standard resources, cannot be met. One possible solution would be to provide and manage a pool of venture capital or apply for partial funding using public (research) funds.

Human resources

- The commitment of the project team plays a key role when implementing a CBPC application. Heterogeneous teams and new technologies mean that various participants often have to deal with new topics. Internal motivation and a common objective shared by the entire team therefore have a significant impact on successful implementation.
- Production technology-specific process knowledge must be carried into the IT department. ML and IT experts can provide valuable input when it comes to generating ideas on how to improve products or processes. Nevertheless, there must be a transfer of know-how as to which production-related problems can be solved. A combination of domain-specific expertise and IT skills is helpful. It is recommended that the project team includes an employee with a close connection to the plant. A role model was developed, for example, in the Data Preparation for Data Analytics (DPDA) project group (prostep 2021).
- If collaborative work on the project is to be performed in different locations, a binding agreement regarding capacities must be reached for all locations to ensure that the project is not too strongly influenced by one location. One lesson learned in this respect is that this requires employees who perform different roles (connection to plant, IT, etc.) at each location.
- It is recommended that personnel capacities be built from the bottom up, clear-cut individual interviews be conducted and resource planning then be carried out again and not be dictated from the top down.
- It often turns out that the headcount is difficult to estimate. Resources cannot always be clearly allocated in advance as a very accurate information base regarding the qualification profiles of the resources/persons is needed. This, however, leads to difficulties in the survey.

Asset management

- major hurdle is often the lack of asset management across multiple plants. This is a prerequisite for CBPC applications and is reflected in the resource analysis. An iterative approach in the business understanding phase is often required here: first create an overview of the assets and then redefine the objectives and resources together with all the stakeholders.

Specific objectives

- Stakeholders should be open to redefining objectives and taking an agile approach. There is a difference between the conventional approach used in the PEP and in software development (e.g. waterfall vs. Scrum). An MVP should be the first order of the day, especially in the context of analytical issues for which a solution is not immediately obvious.
- The impact of CBPC applications on key performance indicators depends on a large number of factors, which means that the target figures actually achieved later may differ.
- The concepts of continuous integration/continuous delivery (CI/CD) cannot be implemented in monolithic applications whose frameworks do not support this. In general, silo thinking in the early phases hampers the successful implementation of cloud-based solutions.

5.2 Data understanding

Identification and analysis of data

- Identifying the data sources required and integrating existing data sources and sinks is a key challenge in the project.
- In the data understanding phase, the project may be terminated if it is established that the previously defined analytical objectives cannot be achieved on the basis of the underlying data. The alternative is to return to the business understanding phase.
- Plant-specific properties and configurations must also be taken into account in the data understanding phase, especially when connecting field devices.

Standards, harmonization, catalogs

- The way in which data (sources) are cataloged in companies, if indeed they are cataloged, often does not meet expectations. The extent to which parameters like the recency of the data or the format in which it is displayed are maintained is not always adequate. More important than conventional data catalogs, however, are the user's perspective and the way in which data is currently used, for example via data flows (as close as possible to the value added).
- As is the case with business understanding, the harmonization of interfaces and data formats boosts efficiency dramatically.
- An integrated data model with centrally merged semantics as part of an overarching data strategy makes it possible for companies to decide which components are mapped locally or in a cloud solution. It is also expedient to start with certain aspects and not to demand universal semantics. What is important is consistency.
- New devices must be measured on the basis of predefined standards. Systematic compliance on the part of external and internal users is also essential.

Scaling in the context of the cloud

- For technical or financial reasons, a PoC does not necessarily have to satisfy all the requirements placed on production software used in the company. The challenge posed by scaling should, however, be taken into consideration in all the phases and should be communicated early on.
- The decision regarding the design of the cloud architecture needs to be addressed as soon as data needs to be accessed for the first time. It is important to specify whether both the application and the data are hosted in their entirety in the cloud or, for example, (also) on-premises. This often has an impact on which data is available. Different data sources are also often assigned to different cost centers, which makes it difficult to manage them in a shared architecture.

5.3 Data preparation

Resource planning for this phase

- The time and effort involved in the data preparation phase depends to a great extent on the type of application involved, e.g. an AI application or an evaluation based on simple calculations. A high level of collaboration between the team members is required for beneficial solutions, as not everyone has the same background knowledge. Appropriate resources must be provided for this purpose.
- Requiring expert knowledge for the annotation of data depends on the participation of the respective colleagues and therefore requires the provision of appropriate resources to avoid inferior data quality.
- Data labeling (e.g. selecting parts for feeding back quality-related data) can, to a certain extent, be very subjective. Use case-specific standards for analysis-oriented documentation are inadequate or can often only be provided by experts. Personnel must be provided for this purpose. Sufficient time and effort must also be planned for checking and verifying the data.
- Learning processes are based on data. Great importance should therefore be attached to the data and its pre-processing (e.g. feature engineering, cleansing, etc.).

From machine to vectors and tensors

- A lack of documentation during the development of the PoC or MVP can lead to problems, first and foremost when cleansing data and during transformation and integration. Developers must be able to track changes. If the model is to deliver the desired result, the exact same process must also be reconstructed for the production environment. A possible solution here is the exclusive use of defined APIs; alternatively, an API may first have to be implemented in the context of the MVP.
- Even if data formats, semantics and other technologies are standardized, this does not necessarily apply to the use of data. Standards do not necessarily have also to be created for the use of data; sometimes best practices may also be sufficient.
- In general, it often turns out that there is a big difference between the learning environment and the production environment, particularly when it comes to selecting, cleansing and transforming data. The aim should be to keep a PoC as close as possible to the system architecture of the production systems, or to take this into account in the MVP at the latest. Data access and intermediate steps should be taken into account when versioning the PoC to ensure that the flow of data from the machine to vectors and tensors can also be traced later.

5.4 Modeling

Versioning and organization

- During model creation, many ML models are iteratively trained and their performance evaluated. The versioning and documentation of the steps and results are particularly important in this context, especially when working together in a team.
- Collaborative data science platforms offer solutions for managing project statuses, data, processes and results.
- An agile approach replaces the rigid division into project phases in practice.

On-site expertise

- If collaboration takes place across multiple locations, data science skills are required in a number of different places.

5.5 Evaluation

Scope for testing and evaluation

- Concurrent testing and evaluation with the staff on the shop floor is difficult due to their integration in day-to-day operations. This means that dedicated personnel must be assigned for this purpose or automated testing be used, if possible.
- If testing, like, for example, labeling, places an additional burden on staff on the shop floor, acceptance of the application is jeopardized.
- Stress tests must also take account of plant-specific dependency.

Real-time data quality requirements

- The discrepancy between the learning environment and the productive environment can sometimes only be determined here – especially with regard to different plants. This is often due to data quality. If inference in the production environment is poorer, further investments are often required to improve data quality. As these do not contribute any short-term quantifiable added value to the originally planned target, they are costs that are difficult to calculate and also difficult to communicate to responsible persons.

5.6 Deployment

Acceptance on the shop floor

- In general, it should be possible to explain metrics to staff on the shop floor and make them easy to understand. Even if the underlying models are, for example, deep neural networks, inputs and outputs need to be explained. Otherwise, the level of acceptance can fall sharply, as the system may for example call the expertise of an employee into question.
- The concepts of explicit and implicit knowledge in the system and in people have to be communicated to the employees. For example, if an employee "senses" that a part is a reject but the system determines the opposite based measured parameters, this can lead to expertise-related conflicts. Information and integration are key to acceptance in this context. It must be made clear that AI systems are systems that provide support and are not intended as replacements.
- Personal data also plays a major role. Different legal opinions or ways of working with data in different regions can still have an impact on the project in the final phase.

Organization and responsibilities

- The operation and maintenance of cloud-based applications results in a multitude of responsibilities. These responsibilities must be checked and clarified beforehand.
- Deployment planning not only means giving due consideration to time zones and the different data centers but also includes local staff. It is important that local staffing be taken into account.
- The maintenance of smaller applications can be carried out efficiently by a DevOps team. Larger applications/platforms, however, need separate build and run teams. The organizational structure in the IT teams must be able to cover both concepts.
- IT teams often build custom applications for internal customers. Sharing the cost of cloud-based products, however, often constitutes an obstacle due to the fact that not only fixed costs but also utilization-related costs are incurred. This means that a payment/cost model to which the individual customers agree needs be defined beforehand. This could, for example, involve an allocation of centralized costs and pro rata levels of use.
- In conclusion, it can be said that the organizational challenges are usually bigger than the technical challenges in development, testing and deployment.

IT infrastructure

- Concepts such as identity, permissions and roles must be taken into account when developing the application. This is required during deployment in the company network at the latest. If data models allow individual permissions, this makes granting access permissions easier. Objects that only allow one set of roles often include too much data.
- If concepts for standardized interfaces and the use of applications already exist in companies, this will also need to be introduced for data in the future. The standards for this are still incomplete.
- When it comes to access, the IT infrastructure is becoming increasingly complex and costly. Zero trust will become the de facto standard in companies and preparations are often currently being made. This must be taken into account when implementing and operating solutions.
- Careful consideration must be given to the various security methods when ensuring the security of cloud platforms. Whitelisting in particular makes data-based collaboration difficult. This creates a conflict between comparatively secure whitelisting and blacklisting, which tends to be less secure.

- The use of distributed systems and distributed data sources/sinks generally reduces the resilience of the overall system. Increasing resilience is however costly due, for example, to the purchase of additional required technology components in the form of buffers or mirroring. This must be taken into consideration when using distributed systems.
- IT infrastructures in the production environment, especially in the context of the shop floor, are often characterized by outdated hardware but first and foremost by a network infrastructure that is too small. This can lead to intermittent load-specific failures. Retrofitting components in outdated IT systems is time-consuming, costly and sometimes not even possible.
- This can also mean that IT infrastructure and its bandwidths do not grow in step with new computing resources on the shop floor. Retrofitting the IT infrastructure is much more complex and is often not included in the budget for a project designed to introduce value-added services. The lack of bandwidth in network infrastructures or the lack of performance of local databases is particularly evident when connecting cloud computing resources.

6 Summary and outlook

6.1 Summary of lessons learned

The work performed by the CBPC project group placed particular focus on forms of collaboration for the shared use of production data with the aim of making it possible to compare systems, products and locations through to and including data-driven product and process optimization. To this end, examples of applications from the manufacturing industry were outlined in the first year of the project. In the second year of the project, the collaboration scenarios were analyzed in greater detail and an interview-based study was conducted in order to abstract success factors and lessons learned from these and other implemented applications.

The wide variety of protocols in heterogeneous IT systems that have evolved over time poses a major challenge. This also applies to the interfaces between new and existing systems. Success factors include the uniform use of IDs and batch reference numbers across processes and systems that enable comprehensive data integration between collaborating stakeholders. This poses a challenge as material and serial numbers are often anonymized for the purpose of IP protection and IT security. In addition, different subnetworks require numerous additional transmission nodes on the IT organization side.

In addition to technical collaboration issues, the organization plays a particularly important role. People with the skills required for setting up data connections, implementing and managing platforms, as well as for data analysis and deployment need to be found and trained. Complementary and heterogeneous teams are an important success factor. The integration of domain-specific expertise is also critical to success. Users and engineers need to be involved on a continuous basis. It is important to raise awareness and provide free space in everyday working life so that collaboration scenarios can be implemented in line with requirements and be put into practice.

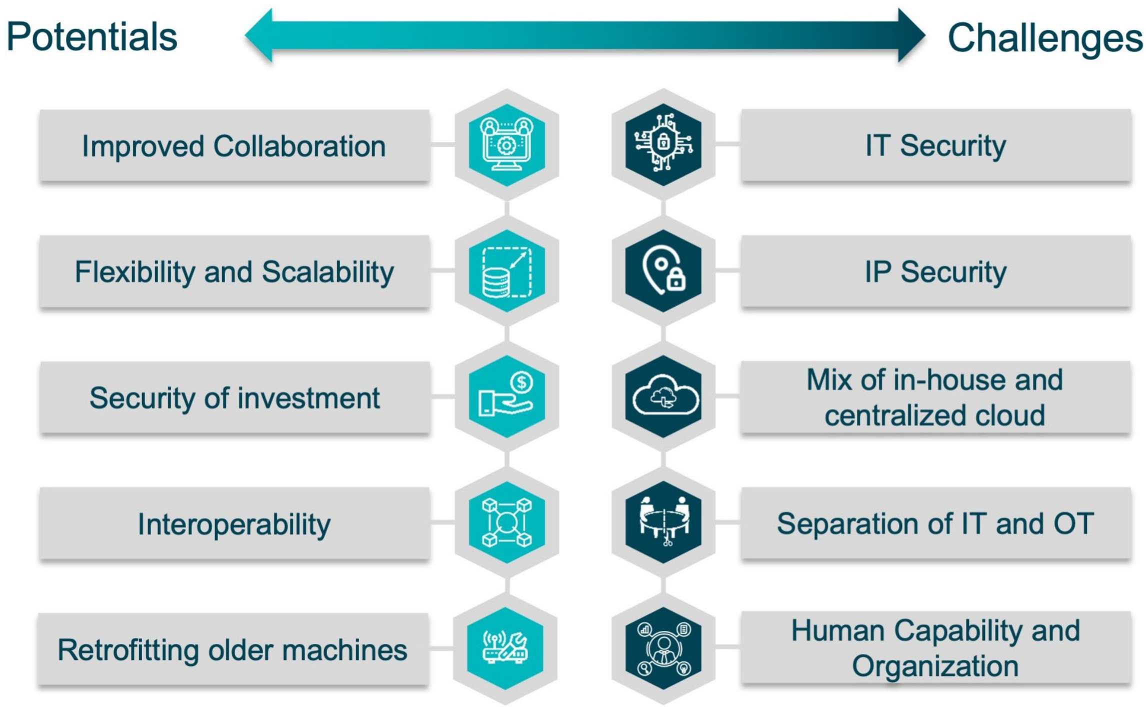

Figure 12: CBPC: potential and challenges

6.2 Outlook

In general, there has been a sharp increase in cloud-based forms of collaboration, and this momentum will not diminish as the level of digitalization continues to grow. Future tasks include topics such as data security and sovereignty in the context of using cloud-based platforms and services. National and European initiatives such as Catena-X for the automotive industry and Manufacturing-X in the mechanical and plant engineering industry promise to develop new solution patterns for the sovereign and interoperable exchange of data and thus promote, for example, traceability and the cross-company consolidation of product and process data through to active collaboration between participants in value networks. However, it remains to be seen how these mechanisms will fare in the free market as the majority of global hyperscalers and the relevant growing economic areas are not located in Europe. This highlights the need for international standards and approved practices for secure and sovereign cloud-based collaboration. At the same time, organizations are also facing a major change as tasks are becoming both more diverse and more specialized and require heterogeneous teams – and a shortage of skilled workers is looming. The latter can lead to a boost in higher service levels thanks to the outsourcing of tasks, although the increasing transfer of prototypes to productive systems means that the trustworthiness of data-based applications in particular is becoming increasingly important, which means that the need to map the corresponding competencies internally is growing.

7 Appendix

7.1 Overview of the collaboration scenarios

UC Type	External collaboration		IT/OT-System/Functionalities	Internal collaboration => supply chain collaboration
UC #	3	3a	5	7
UC name	**Data usage external assets**	**Energy demand/management**	**Automation system design**	**Real-time impact consideration**
Descrip-tion	Learning from the big amount of users, the learning cannot be connected to the individual user	Energy demand (finalized product) shall be available and transparent. Therefore, data must be collected for each part and sub-assembly up to the entire product, from extraction of the raw materials up to the final assembly.	Interpretation of the data (e.g. signals) * system state data is needed * Data shall be given continuously or on demand	Depending on the available data the best option shall be chosen. Possibly using AI
Situation	Process expert is interested in -bottleneck forecast (Engpass Prognose), capacity forecast (Kapa Prognose) - predictive X - receiving and evaluating machine data => getting findings for XY	Sustainability is gaining in importance. Today the energy demand is rather estimated, data about the entire supply chain is not available.	Panner must constantly have available all process-relevant data of all available production resources.	Production relies on the supply chain and internal processes. There are always short term impacts => prediction, not planning. Real time information may not be available (truck is late/had an accident).
Goal	To be able to obtain data in order to optimize the production processes	Be able to report the specific energy demand per finalized product.	Coordination of data exchange between planning and automation system design	To be able to obtain external impact parameters, clarify the impact, understand options and react
As a	Process Expert	Sustainability manager	Planner	Production controller (manufacturing and assembling)
I want to	use knowledge and data from external assets (findings from their experience transferred to my environment)	have all digital information available	know if it is possible to receive process data and how.	see the impact on my production by suppliers and customers in real time
in order to	to optimise my own production process and asset usage	to be able to report the specific energy demand for production and delivery	be able to plan the future production	adapt my program to ensure max OEE
Produc-tion relation	production process and assets	production is the focus of consideration	Automation system designer (OT)	Production process
Cloud relation	high => XaaS CatenaX => Traceability & Bedarfs- und Kapa Analyse	Cloud is not necessarily the desired technology, but an option. Option: Including IoT platform	Pre-processing of data may be possible in the cloud	high, higher for external collaboration
Se-quence	see swimlane presentation	see swimlane presentation	see swimlane presentation	see swimlane presentation
Infor-mation	Betriebsdaten, Machinen-daten, Energiedaten (Effizienz), Footprints/Sustainability (CO2, ...), Energie/1 Product ...	Energy demand per component: production data, product data, internal data and data from supply chain	Device/machine status, current process runtime, set-up status of the device (Rüstzustand), Data as xml Information	
Desired benefits via cloud techno-logy	What are the options	Not cloud but shared data, possibly private cloud GaiaX like (federated services, not public cloud)	Automated alignment options desired through cloud, otherwise more time-consuming	Cloud may be an option, but is not neccessarily the solution
Comment	2 demonstrators available * what do they do with my data? Project partner is optimizing his process with my data		working group existent zum semantischen Modell	cloud not internally only, but incl. ext. access, cloud from external provider, outside of my corporate network. cloud => no defined location, anywhere

Table 1: Details of the CBPC Group's application scenarios

7.2 Specification of additional collaboration scenarios

Use Case 3a energy demand

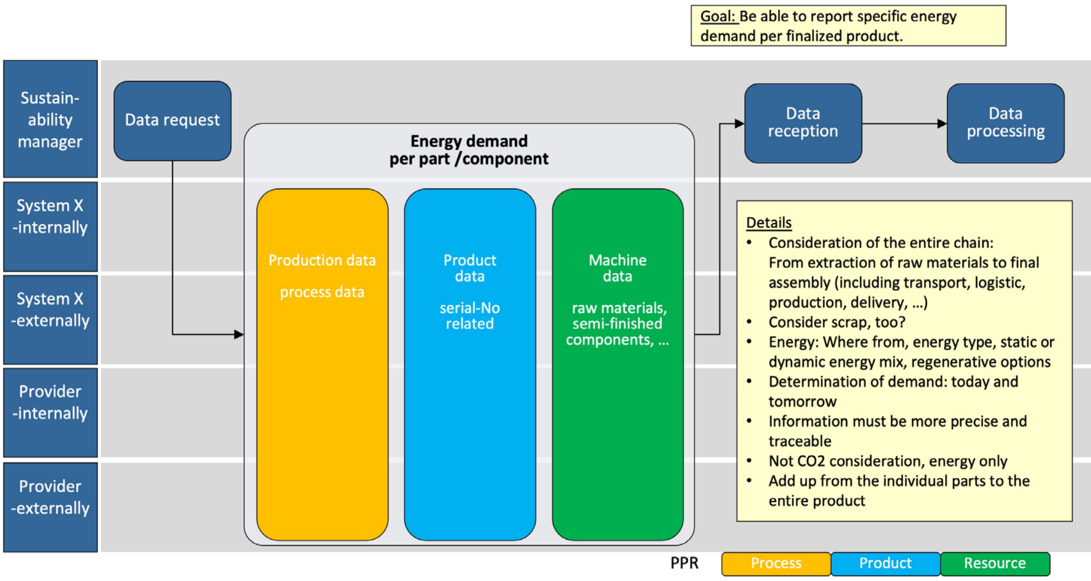

Figure 13: Specification Use Case 3a: Energy demand

Use Case 5 automation system design

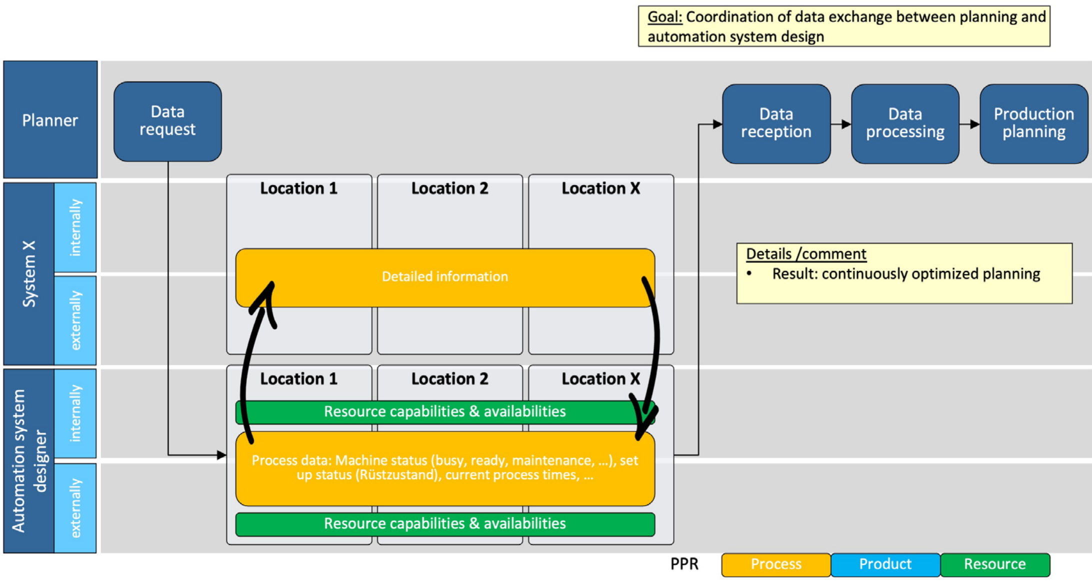

Figure 14: Specification Use Case 5: Automation system design

Use Case 7 real-time impact consideration

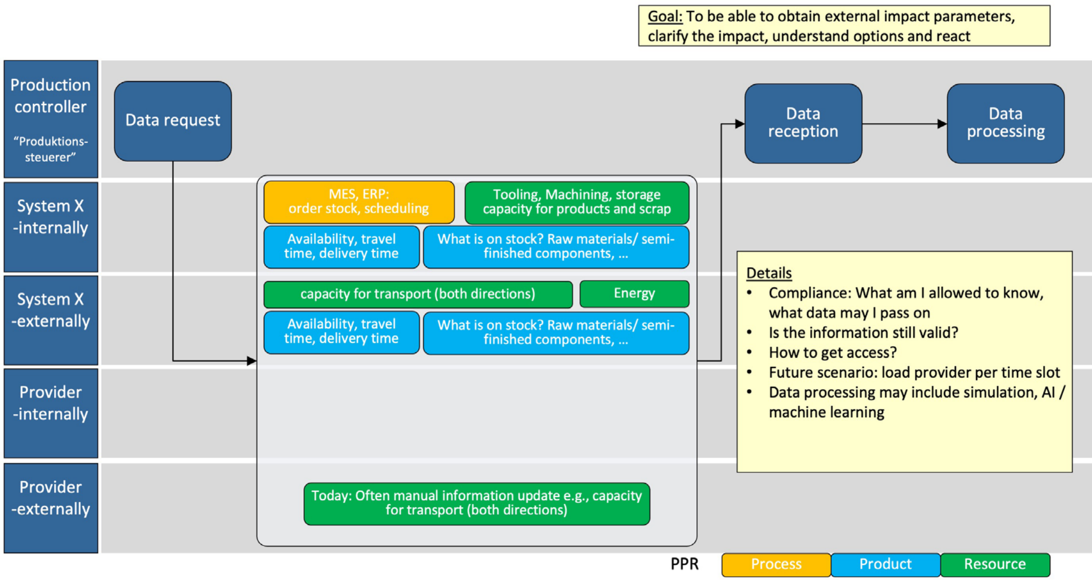

Figure 15: Specification Use Case 7: Real-time impact consideration

8 References

Chapman, Pete; Clinton, Julian; Kerber, Randy; Khabaza, Thomas; Reinartz, Thomas; Shearer, Colin; Wirth, Rüdiger (2000): CRISP-DM 1.0. Step-by-step data mining guide: CRISP-DM Consortium.

prostep ivip (2021): Management Summary "Data Preparation for Data Analytics" – Data preparation for industrial value creation. prostep ivip Association. Darmstadt

Rauen, Hartmut; Glatz, Rainer; Schnittler, Volker; Peters, Kai; Schorak, Markus H.; Zollenkop, Michael; Lüers, Martin; Becker, Lorenz (2018): Plattformökonomie im Maschinenbau. Herausforderungen – Chancen – Handlungsoptionen. Roland Berger GmbH (ed.), VDMA e.V., Deutsche Messe AG. München.

Wöstmann, René; Brünnhäußer, Jörg; Büttner, Gerd; Beckmann, Felix; Deuse, Jochen; Stark, Rainer (2021): Data Preparation for Data Analytics and Artificial Intelligence. Data preparation for industrial value added. In: ProductData Journal 28 (1), pp. 16–20.

Wöstmann, René; Borggräfe, Thorbjörn; Janßen, Sascha; Kimberger, Josef; Ould, Solomon; Bennett, Nick et al. (2022): Data-driven recipe optimisation based on unified digital twins and shared prediction models. In: Michael Affenzeller, Agostiono G. Bruzzone, Emilio Jimenez, Francesco Longo und Antonella Petrillo (Hg.): Proceedings of the 34th European Modeling & Simulation Symposium (EMSS 2022). Rome, Italy, 19-21.09.2022, pp. 1–11.

prostep ivip Association

Dolivostraße 11
64293 Darmstadt
Germany

Phone +49-6151-9287336
Fax +49-6151-9287326

psev@prostep.com
www.prostep.org

ISBN 978-3-948988-29-6
PSI 31
2023-12/Version 1.0