Development and Evaluation of a Process Management and Analytics Platform for Small and Medium-sized Enterprises

Jonas Kallisch
Emden/Leer University of Applied Sciences, Emden, Germany

Christoph Wunck[†]
Emden/Leer University of Applied Sciences, Emden, Germany
Iowa State University, Ames, USA
OFFIS Institute for Information Technology, Oldenburg, Germany

Abstract

This paper presents a prototype for an information exchange system, which allows information exchange between companies without actually sharing data. First, the need for such an intercompany exchange platform is explained and the value for supply chains resulting from such a platform is described. A literature review presents the existing concepts and techniques contributing to the development of an architecture. Finally, the information exchange concept and the prototype implementation are explained in detail.[*]

Key Words: Small and medium enterprises (SME), analytics, machine learning, data exchange.

1 Introduction

Machine learning, advanced analytics and other methods have become a major element of modern production systems. Reduction of waste, improvement of manufacturing timings and quality are some of the contributions that these methods have brought to the companies own manufacturing infrastructure [17]. Because of this fact, many companies have started to analyze their shop floor data to profit from the benefits described. The increasing availability of customizable analytic tools and the decreasing of their prices enables even small businesses to use them. As a result, companies are more empowered than ever to identify and address the vulnerabilities of their manufacturing infrastructure. Parallel to this development, supply chain cooperation has deepened, which means that value creation depends more than ever on the cooperation of the companies[11].

Despite this reliance on collaboration, many companies, especially small and medium enterprises (SME), do not share data with their customers or suppliers and so do not optimize their joint manufacturing [23]. There are many reasons for this lack of data exchange. An interview series conducted by the authors investigated into these reasons, found that a common issue is that companies have recognized the value of the data captured in manufacturing, and fear disadvantages from sharing it.

The fear is that the companies could lose control over the data shared in the supply chain (SC), for example if the data is stolen [5]. In some cases, the losing of control about their intellectual property could even threaten the existence of the affected company, because competitors could use the information to improve their own production. Therefore, the fear of many companies of sharing data from their manufacturing systems is justified and understandable. It can be concluded that the first requirement of companies for a system for collaborative analysis of data in a supply chain is the guarantee to maintain data sovereignty. Ideally, a solution should be sparing with the sharing of data or completely avoid the sharing of raw data. Requirement #1 is therefore the preservation of the company's data sovereignty.

Another issue for SMEs is their lack of technical knowledge to create their own platforms and systems to exchange data with other SMEs [32]. SMEs often work on a low budget with external technical service providers, which must be able to support the solution of the company to exchange data with a customer or a supplier. This situation, which particularly affects SMEs, results in the requirement that a solution must fit the company's budget and must also be available in a form that can be implemented and maintained by service providers. Requirement #2 can be described as companies expecting a lightweight solution.

The advantages of sharing data rely on the ability to analyze them and understand the results of the analysis. Many SMEs don't have these abilities within their businesses, as our interview partners told us and other studies found out as well [30]. Thus, the third requirement for a solution can be described as the ability of companies to perform cross-company analyses on their own or at least in cooperation with their IT service provider. Requirement #3 can be summarized with the term applicability with the capabilities of the companies.

However, there are numerous reasons for analyzing data across companies or even for sharing data, even if this sharing

contradicts the claim to maintain data sovereignty. By waiving to share shop floor data in the supply chain they give up many opportunities to improve the overall production efficiency along the value creation inside their SC [23].

A first step to data integration is to reduce the costs of data generation inside the SC [1]. Much of the data used by the participants is collected multiple times for different purposes. This results in costs that could be lowered by sharing the data between the members of the SC. In addition, the companies that are involved in the SC can be seen as members of a community of trust, as they should have a shared interest to improve the results of the group like quality or efficiency.

A second step towards integrating SC data is to create new data through a combination of already existing data. By combining data from different stages relations between incidents appearing in late stages of the manufacturing process and data patterns appearing at earlier stages of the SC can be found [24]. This helps to reduce waste and optimize the overall production system.

To summarize, it can be stated that data exchange on a shop-floor level is not widespread and SMEs in particular rarely exchange manufacturing data with other companies. The authors propose that in addition to the lack of trust between companies, the existing solutions for data exchange do not meet all the requirements of companies.

To demonstrate this, the paper will first present the existing architectures and evaluate them with a view to the identified requirements #1, #2 and #3. Therefore, particular attention will be paid to the aspects of data security and usability for SMEs. The paper will show that existing architectures and concepts do not meet the requirements for all companies and that a concept is therefore needed for the secure networking of corporate data assets. Such a concept and a prototype have been developed by the "Zukunftslabor Produktion" [Futurelab Manufacturing] (ZLP) [35]. This interdisciplinary project is developing solutions and models especially for small and medium businesses in the manufacturing sector. Based on a use-case of the production of die-cast aluminium parts, the project examines how technical manufacturers can use data analytics and other industry 4.0 methods to improve their shop floor systems and create more resilient SCs. The concept and the reference implementation will be presented in the third part of this paper. The paper will close with a discussion of the findings.

2 Existing Concepts and Solutions

To identify existing concepts and implementations of intercompany communication in SCs, more than 100 papers have been examined and evaluated in a structured literature review. This section presents the concepts identified by the literature review and shows which SME requirements they do not meet.

2.1 Data Spaces

Data spaces have been defined by Franklin, Havely and Maier as a next step in the evolution of data integration architectures [8]. The evolutionary step is that data spaces combine storing of data with services, to merge data from different sources to extract information. Key of this approach is the integration of data from different domains and the mapping of their different data elements. Therefore, data spaces can represent a SC as a number of related data sources, which can be connected. This data connection allows the whole system to improve its value. Similar to a SC the data spaces' value depends on the level of compatibility – mapping and matching – between the different suppliers [27].

One implementation of these data Spaces is the International Data Space (IDS) [9]. The IDS is a system of data providers interacting on a platform. Each participant can be provider or user of data and has the right to negotiate about the rights on the data. The platform allows to connect the data on a contract base. The full model of the IDS is displayed in Figure 1.

The model consists of data providers and consumers that hold their data in their own physical storage locations. However, these physical storages do not interact directly with each other. The interaction is controlled by the broker, which offers two main services to the participants. Firstly, the broker lists and categorizes the data delivered by the data providers. This service enables the searching companies to find the data they need. The companies can search based on labeled data or based on the types of data they already have – e.g., a certain type of
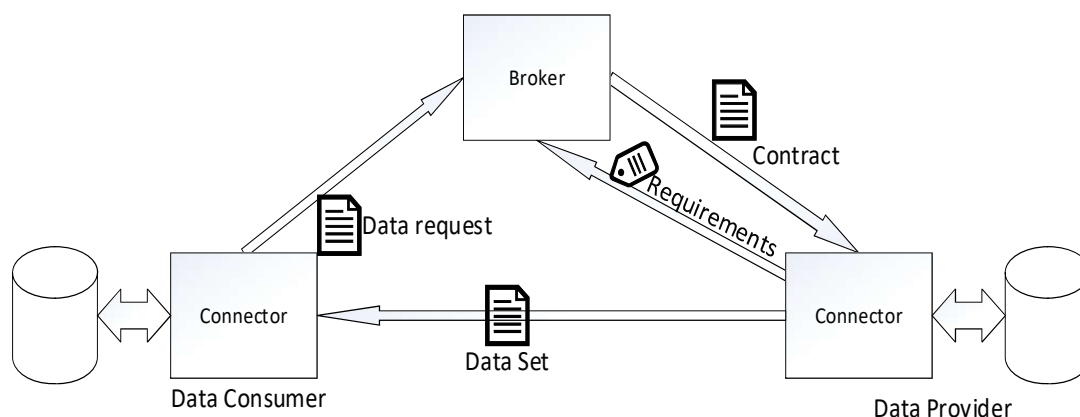

Figure 1: Model of the IDS [13]

application or asset. If the system includes a data provider with the needed data, the broker connects the two – or more – companies. As the data can be traded anonymously, it is even possible to buy data from trusted sources without knowledge about the very company supplying the data.

Eventually, a service provider intermediating between the entities is part of the model. A service provider can be part of a whole ecosystem of service providers. These can for example be infrastructure services – e.g. data storage, computing power or other infrastructure – that allow participants to be part of the data space without having their own hardware. [29] Beyond this Infrastructure as a Service (IaaS), other providers can add services – e.g., data analysis or searching services – or Software as a Service (SaaS) – e.g., analytics tools or data mining software – via a dedicated App Store. [24] Therefore, the IDS delivers a solution for sharing data between companies on a contract base. However, the concept explicitly assumes sharing of data and thus cannot preserve the data sovereignty of the participants. Therefore, the concept cannot be used to network companies without exchanging data and does not meet the requirement #1.

2.2 GAIA-X

The European GAIA-X project can be considered as a solution for connecting different kinds of data stores. These data stores might also be IDS infrastructure elements. It contributes an architecture concept that allows the sharing of data in a public catalog, where everybody can see the available data, but can only access data when the access is granted [10]. Similar to IDS brokers, GAIA-X enables companies to give permission for sharing data on an individual level.

Core component for the concept is the federated catalog. This is a public catalog, where companies can publish the data they are able to share. Every company connected with the catalog is able to see the data that is available, but can only access data when the access is granted by the data owner. Therefore, the companies can choose to give permission for sharing data, on an individual level.

The GAIA-X foundation guarantees this kind of data sovereignty as a service. This means that participants have the capability to fully self-determinate their data exchange and sharing. The secure exchange is realized by a function called Data Contract Transaction. This service initiates a handshake between the data provider and the requesting party. The service validates the contract and, if the content is valid and both parties have confirmed the transaction, the Data Contract Service distributes the Data Contract to both companies. After that, the requesting company can access the requested data and may analyze it. The distribution of data is observed by a function called Data Exchange Logging, which enables companies to restrict the usage of their data to a certain extent or for a specific purpose.

The model of GAIA-X allows sharing of data in a secure and customizable way but still needs to actually exchange the data to analyze them inside of the SC. A very interesting part of the solution is the way the catalogue combines data identification

and services by self-description. The value of this for companies that are interested in sharing data has also been examined by Dumss et al. [6]. They suggest an architecture model called EuProGigant, which allows exchanging data in a scalable way. They also describe how services can enrich the generated data and emphasize the importance of self-description in GAIA-X. However, they did not give a suggestion how to secure the data exchange or how to keep the intellectual property of the companies. This means that the concept, as it is proposed at this point, is not able to protect the data ownership interests of the SC companies.

Summed up the GAIA-X foundation provides a reliable, effective and secured solution for sharing data. For companies that are interested in sharing or selling their data the GAIA-X Federated Catalogue is a fitting solution. In the case of an interconnected SC GAIA-X might also be a very good communication platform, but by itself is not able to create an information exchange without actually exchanging the data. By that, the author's opinion is that GAIA-X does not preserve data sovereignty and therefore misses requirement #1.

2.3 Catena-X

A development based on GAIA-X is the concept of Catena-X. As an architectural concept, Catena-X offers a decentralized data platform system that enables companies to exchange and use data securely. [12] The platform is based on a smart contract system and an open data architecture. In the concept of Cantena-X, data exchange is realized by the so-called Eclipse Dataspace Connector (EDC).

The EDC enables participants to exchange data from their internal systems, such as ERP, CRM or PLM systems. [22] To do this, a smart contract is created in the platform between the participants who want to exchange data. This contract defines the conditions of the exchange and manages the process of data sharing.

Figure 2 shows the structure of the Catena-X platform and how the data exchange is managed. As can be seen, the components of the federated catalog, as well as the Identity Service of GAIA-X are used in the concept. The contract is concluded via these and guarantees that the transaction partners and offered data are correct. Unlike in the GAIA-X concept, the data exchange itself does not take place via the central platform. Instead, each participant uses its individual EDC in one of two ways. Data can either be performed via direct transfer between the EDC systems of the respective companies, which is indicated in Figure 2 as Option A, or via a service provider, shown as Option B. The second option can also include other data services, such as conversion or analysis. With SMEs in mind, it is also possible to operate the EDC as software as a service. These can therefore participate in the common data analysis without having to build up their own know-how.

Overall, the concept of Catena-X enables the secure sharing of data between companies. The concept of locating data assets within the system is well suited for use between partners in a supply chain, as it is based on GAIA-X. The data can be transferred directly between the partners. The direct transfer of
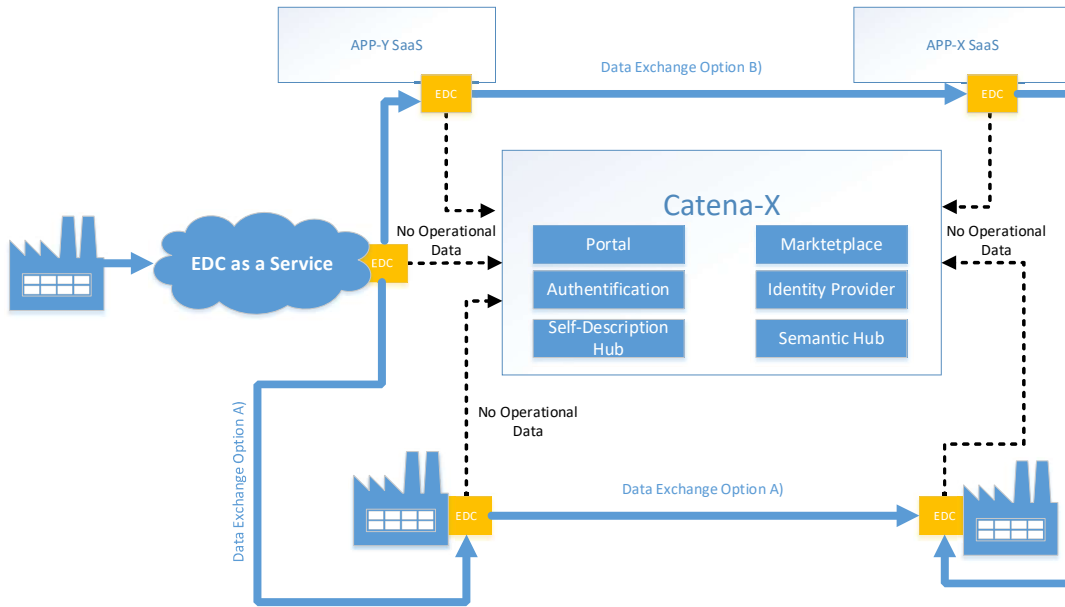
Figure 2: Architecture of Catena-X (based on [15])

data between companies addresses one of the caveats to analyzing data across companies. Because the data is not stored centrally, it is easier to protect data ownership and from unauthorized access to the data. Nevertheless, even in this architecture, sharing data on an individual level becomes necessary. This means that protection against unauthorized access becomes more difficult with each additional location or company where the data is stored as a copy. Besides this problem, the concept of Catena-X does not offer protection against misuse of the data by a partner within the supply chain. The fears of many SMEs of competitive disadvantages due to the loss of control over their know-how can therefore not be solved by Catena-X. Therefore, requirement #1 is not met, even if Cantena-X enables a better control of the data, since these are not stored in a central data storage.

2.4 Federated Learning

Federated Learning (FL) is a concept to analyze datasets, which are distributed over different devices that are connected with a central station [36]. It can be divided in horizontal and vertical FL [34]. The difference between these two types of FL is the selection of elements they share. As Figure 3 shows horizontal FL shares features, e.g., temperature measuring or other kind of data points, but not the samples – a concrete case of measurement, while vertical FL shares samples, but not features.

The more common case of FL is the horizontal FL as it is used in mobile devices, to improve their ability to analyze their user data, without transferring them. A horizontal FL starts with an initial algorithm, which is created on a data sample. In a second step, the model is decomposed into sub models, matching the

data elements of the different storages. The different data stores, e.g., smartphones, then train the model on their own data. In the fourth and last step the results of the individual models can be transferred to the central application to improve the model. The result is that the different data stores are able to improve their analytical models of their data without sharing with each other. The limitation of the horizontal FL is that it requires similar

Horizontal Federated Learning
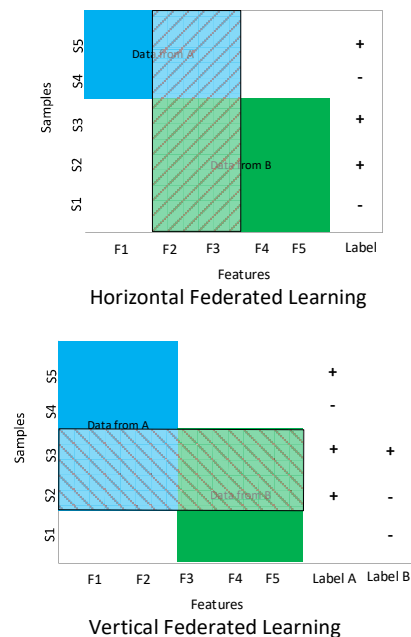

Vertical Federated Learning

Figure 3: Difference between types of federated learning (based on [34])

data structures on the different devices.

In SCs, that do not share common data architectures between different companies, vertical FL can be used. As shown in Figure 3 the data stores in a vertical FL model are not sharing the same features – data structures – but the different data stores share the same samples. For example, two companies in the same city might not collect the same data, but collect the data from the same customers. If these companies share a common interest, they could combine their data and use it to improve the quality of their prediction algorithms.

Therefore, the usage of vertical FL requires to exchange the data or at least the labels of the data between the companies. A way to solve this problem of exchanging data between the company might lay in the work of Yang, Liu, Chen and Tong [34]. They propose a framework for secured vertical FL, which allows participants to exchange analytical models but frees them from having to share their data with each other. According to this, the issue could be considered as solved, but even these variant of vertical FL comes with some limitations.

The first limitation of the model is that it still requires the companies involved to share their labels of all samples to improve the training of the model. This might lead to leakage of information possible as Bagdasarya, et.al. have shown [4]. The second limitation comes with some assumptions on the data of the participants. One of these is, that all participants are sharing the same labels and have a combined goal, e.g. reduction of waste or overall cost reduction [16]. This would be an issue, if a company is involved in different SCs, with different goals or might profit from a certain kind of waste. The third and last-mentioned limitation relates to the value of information contributed by each individual party involved in the vertical FL. The problem is called 'unbalanced clients'. This means that some participants of the SC are able to contribute more to the whole system than others. The issue of this is that a federated learning architecture isn't able to balance these different features, without exchanging of datasets [36]. It should be mentioned that Zhang, et.al. have found that by selecting an adaptive number of local training rounds for each party can lead to better models, but this also increases the danger of data leakage. Obviously, FL does not deliver a lightweight solution within the capabilities of SMEs. Though FL bears great potential as a concept preserving data sovereignty, requirements #2 and #3 are not fully met.

2.5 Commercial Solutions

Beside the implementation of any of the concepts evaluated in the preceding sections the question remains if there is any existing commercial solution to securely connect SME SCs without giving up their individual data ownership interests. The authors have investigated various available products to see if commercial solutions are available that solve the problem away from scientific issues. The products investigated are described below and, in the authors' view, represent a good range of the systems available on the market.

- SAP Business One

SAP Business One is focused on small businesses and is able to deliver a ERP system that does not share much with the main SAP Products S3 and S4 [28]. Companies can chose to run the software on their own server or use a cloud server hosted by SAP [26]. The product allows built-in analytics and SC automation of business transactions [25]. Business One is a solution that can help small companies to get to the earlier described state of the art of SC communication. It also allows the companies to get analysis of their business decisions, but is not able to solve the issue of intercompany communication.

- SAP Business ByDesign

SAP Business ByDesign aims at medium businesses [25]. The product is cloud based and provides a customizable ERP system, which can be integrated in a SC [14]. The solution is able to create build in analytics to create real time dashboards of the company´s situation [28]. In addition to the features of BusinessOne, it delivers the features of SC Management, like functions to support sourcing and purchasing [28]. As BusinessOne, Business ByDesign did not involve the SC partners manufacturing infrastructure.

- Microsoft Dynamics

Microsoft Dynamics delivers functions for integration of warehouses, material flow planning and collaboration with other companies [19]. The SC management component delivers similar functions to the given by the SAP product [7]. The review also found that all investigated ERP systems enable the integration of suppliers on the level of business communication. Another study also shows the state of ERP MES integration in Microsoft Dynamics and the lag of a integration of intercompany shop floor data exchange in the solution [21]. A current development of Microsoft Dynamics is the product Dynamics 365. The product allows the usage of so called "supply chain data sharing templates". These templates provide options for cloud-based sharing of transactional information, like payment and logistics data, sharing of order pool, warehouse management for example for commission purposes or delivery management.

- Microsoft Azure

Microsoft Azure enables collaboration between companies in a SC with a cloud application that integrates inventory, factory status and logistical data in a twin of the SC elements [18]. The core focus is on using machine learning, optimization algorithms and artificial intelligence on distributed data to improve the decision making inside of the SC. An element called control tower creates a dashboard, which shows this state. The concept of these control towers is that a SC member delivers all relevant data to the gateway and connects them to

the data of the other members. The results can be combined with public data, for example weather information [31]. An example for an SC using Azure can be found in the agriculture sector. These use cases show how the connection of different SC data storages can improve the whole SC's success, but also shows that the members involved must be willing to share their data, as the architecture provides the risk of losing control about the data. It should also be mentioned that the project focus was not on connecting SC members but on connecting singular IoT solutions and product used by different farmers involved. Therefore, the translatability to industrial SC´s may not be given.

2.6 Conclusion

This section has found that existing approaches to creating cross-enterprise data sharing do not meet all of the requirements identified in this paper. In particular, the guarantee of data sovereignty is insufficiently fulfilled by the existing solutions. Nevertheless, the existing concepts provide a basis for building a solution that meets all requirements. Table 1, on the next side, summarizes which requirements are met by which solutions.

3 Towards a Data Securing Information Exchanging Architecture

As the previous chapter has shown, none of the existing concepts or products meet all requirements to solve the problem of information exchange between companies without direct data exchange. Based on the requirements and the existing approaches, an architecture was developed that on the one hand offers the option to interact with data ecosystems such as IDS or GAIA-X in the future and on the other hand protects the control over the intellectual property of the companies. Our concept provides for the sharing of information via a linking platform, which neither requests data for centralized processing, nor passes it on to other participants in the platform. To create these links, existing connections in the data traffic can be used. In an SC, these usually consist of the flow of goods between the respective companies.

In many cases, these goods are identifiable, for example via a batch or serial number. These identifiers can be used to find and link the products within the cross-company workflow and the data collected during processing.

In theory, the data values recorded by one company A therefore, have a direct influence on the measurement results of a second company B. If company B finds that different results are determined for comparable process variables in its own data stock, the cause can therefore lie in a deviating process variable in company A or in a pre-production stage of A. By using a process support platform, these correlations can be analyzed without direct data exchange. In addition, the leakage of information is less likely, since only in defined cases, analysis will be started and the results do not even have to be shared between the companies.

This description of a possible network results in a concept whose core is a value-oriented relationship between at least two partners. As displayed in Figure 4 the partners are connected via a process support platform that distributes information based on the stored exchange relationships. The participants would be company A with data $A1$ and the derived information $A2$, company B with data $B1$ and the derived information $B2$. $A2$ and $B2$ are exchanged in the scenario. Thus, participant A holds $A1, A2$ and $B2$ and participant B holds $A2, B1$ and $B2$, whereby each participant can determine the concrete information provided. Based on this core concept, the participants of the process support platform can improve their own analyses through the exchange and agree on individual agreement services. Such a service can be, for example, a cross-company fault cause recognition, which is passed on to supplier A in the event of a fault at B that cannot be explained by its own processes. A would then be obliged to determine whether this error can be traced back to processes in its production.

The concept for such a data saving and cross-company information sharing architecture consists out of three layers:

- Communication layer
- Service layer
- Analytics layer

The first of these layers, the communication layer, is intended to establish data exchange between the individual enterprise is to be transferred to link the data sets of the various supply

Table 1: Fulfillment of requirements

|	Protection of data sovereignty	Lightweight solution	Capabilities of companies
Data Spaces	requirement not met	requirement not met	requirement partially met
GAIA-X	requirement not met	requirement partially met	requirement fully met
Catena-X	requirement partially met	requirement partially met	requirement partially met
Federated Learning	requirement fully met	requirement not met	requirement not met
Commercial Solutions	requirement not met	requirement fully met	requirement fully met
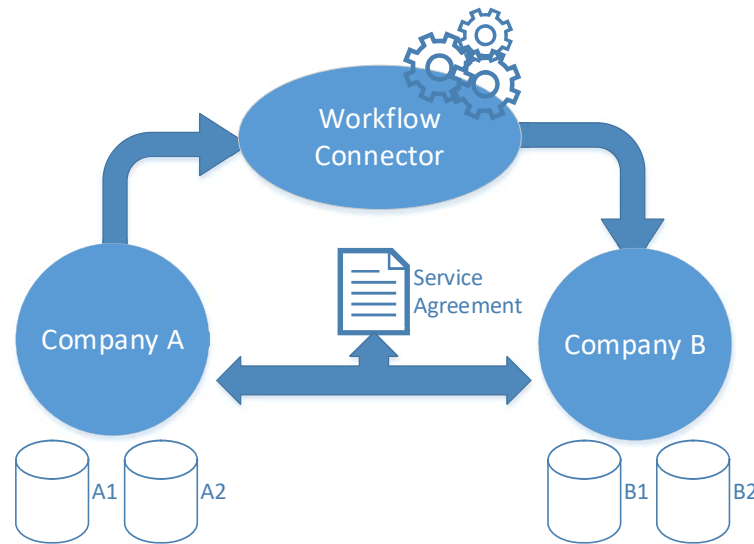
Figure 4: Concept of the project

chain participants. The correct selection of these data points and information is thus not monitored or determined by an internal or external actor. From the point of view of the concept developers, this is not necessary, since all participants in a supply chain should be interested in optimizing joint value creation. Nevertheless, it is conceivable to have the selection and set-up performed by a common, preferably independent, service provider. This would also have the advantage that knowledge of the entire data structure of all participants can be incorporated into the establishment of the link.

The second layer is used to implement services between the participating companies. These services allow companies to create individual applications for their supply chain. The layer in the interconnection platform describes these services and stores the rights of the respective participants to use the services. Requests to use services are forwarded via the communication layer and lead to the creation of an agreement to use the service. It would also be conceivable to extend the service system to an open system in which providers of data or analytics services could also interact in the supply chain. Essentially, however, this would require a partner to provide hosting of the platform and enable it for companies not in the supply chain. A connection with existing architecture models, such as IDS or GAIA-X would also be an option for the integration of these external services.

The final layer of the concept is for cross-enterprise data analytics. The layer is intended to provide a set of basic analytics that can be used by companies within their data platforms. Results of the cross-company analyses are then exchanged via the link in the communication layer. Here, participants should be able to restrict the sharing of data to prevent the extraction of data through targeted queries. This layer is implemented with the use of federated learning models in mind, as these provide a suitable basis for implementing cross-enterprise analyses. In order to compensate for the

identified weaknesses of federated learning, companies should be able to configure the analysis models themselves and thus influence the respective parameters in their access area. The goal is thus to compensate for the problem of unbalanced clients.

This model resulted in a basis for the implementation of the architecture with one of ZLP's industrial partners. This forms the basis for answering further research questions and optimizing the model.

4 Prototype of a Supply-Chain Workflow Connector

The concept described in the previous section was developed in the form of a prototype in cooperation with an industry partner of the ZLP. This prototype is to be tested on industrial data to determine the potential of the ZLP's described data-saving concept. For this purpose, the analysis results of the prototype will be compared with a classical, central data analysis. By looking at the differences between the two approaches, any potential for optimizing the system will be identified.

In preparation for implementation, the data sets intended for exchange were transferred to a data platform. This serves to separate the operational infrastructure from the infrastructure connecting the supply chain. In view of the large number of different formats within the data sets provided by the practice partners, these individual database systems were realized in the form of customizable data platforms. For the selection of the data platform system, different systems were compared with each other. The architectures compared included the aforementioned commercial solution Microsoft Azure, which is used in a combination with Apache Kafka [2] as a data stream message broker, MongoDB [20], Hadoop [3] and Apache Superset [33]. Due to the industry partner's specifications, only systems with OnPremise options were evaluated.

The various systems were built as examples and evaluated in terms of their utility values and the effort required to build them. Since one of the ZLP's goals is to provide a modular solution that is easy for companies to implement, the software's deployability for companies with little IT knowledge of their own was also considered a key criterion. All of the platform systems considered are basically capable of meeting the requirements of the ZLP. In terms of the complexity of the structure, the various systems examined did not differ significantly from one another. With regard to the performance of the platforms, differences can be seen in specific situations, which were also described by other studies.

As a decision for an architecture it was decided to first realize two platforms that represent two actors within the supply chain of the ZLP and to show the transferability of the concept. These form between the tool mold fabricator and the aluminum die caster. Hadoop and MongoDB were implemented as the systems. As a system, these offer many advantages in terms of extensions and existing analysis systems.

The systems are linked via the communication layer. This was realized with the help of the multiagent system Jade. Jade is a framework that enables the creation of agents that can also exchange messages and data packets across systems.

The respective agents are used to communicate with the central Workflow Connector and can make requests to other participants in the platform. Which rights and which relationships exist between the companies regulates how the requests are handled and forwarded in the platform. The companies are able to configure these relationships themselves and also select the respective data elements.

Based on this link with Jade, services or agents were first defined that enable the exchange of quality parameters via the platform. On the basis of these services, quality deviations of the created parts are to be compared with the history of a tool without merging the data. The analysis of this data will take place in a next step, once different options for implementing the analysis layer have been evaluated and the most suitable solution has been identified.

5 Conclusion and Further Work

This paper described the need for a secure information exchange system for companies in SCs. The current lack of suitable concepts and solutions was shown. The authors devised an architecture that protects the data ownership of the involved parties and enables information sharing. The advantages and disadvantages of the architecture have been briefly discussed and will be further evaluated within the project prototype and be addressed in future work. The concept achieves networking through an exchange platform that allows companies to exchange information based on individual contracts without having to disclose data. The concept adopts some mechanisms and ideas of existing concepts. On the one hand, this improves the quality of the architecture by using tried and tested systems and, on the other hand, it enables companies to integrate into platforms such as IDS or GAIA-X. In the future, the authors will build further implementation variants, such as a

microservice architecture, and investigate them with regard to their potential. In addition, the ZLP will evaluate the prototype with support of our industrial partners. For this purpose, experimental setups with industrial partners and on-site laboratory infrastructure will be used.

References

[1] V. Alcácer and V. Cruz-Machado, "Scanning the Industry 4.0: A Literature Review on Technologies for Manufacturing Systems," *Engineering Science and Technology, an International Journal*, 22(3):899-919, 2019, doi: 10.1016/j.jestch.2019.01.006.

[2] Apache Kafka, *Apache Kafka,* [Online], Available: https://kafka.apache.org/ (accessed: Feb. 16 2023).

[3] Apache Software Foundation, *Apache Hadoop,* [Online], Available: https://hadoop.apache.org/ (accessed: Jun. 28 2022).

[4] Eugene Bagdasaryan, Andreas Veit, Yiqing Hua, Deborah Estrin, and Vitaly Shmatikov, "How to Backdoor Federated Learning," *International Conference on Artificial Intelligence and Statistics*, pp. 29382948, 2020, [Online], Available: https://proceedings.mlr. press/v108/bagdasaryan20a.html

[5] C. Colicchia, A. Creazza, C. Noè, and F. Strozzi, "Information Sharing in Supply Chains: A Review of Risks and Opportunities Using the Systematic Literature Network Analysis (SLNA)," *SCM*, 24(1):5-21, 2019, doi: 10.1108/SCM-01-2018-0003.

[6] Stefan Dumss Markus Weber, Clemens Schwaiger, Clemens Sulz, Patrick Rosenberger, Friedrich Bleicher, Manfred Grafinger, Matthias Weigold*.,* "EuProGigant – A Concept Towards an Industrial System Architecture for Data-Driven Production Systems," *Procedia CIRP*, 104:324-329, 2021, doi: 10.1016/j.procir. 2021.11.055.

[7] F. M. Elbahri, O. I. Al-Sanjary, M. Ali, Z. A. Naif, O. A. Ibrahim, and M. N. Mohammed, *Difference Comparison of SAP, Oracle, and Microsoft Solutions Based on Cloud ERP Systems: A Review*, 2019, [Online], Available: https://www.researchgate.net/profile/omar-al-sanjary/ publication/333457778_difference_comparison_of_sap_ oracle_and_microsoft_solutions_based_on_cloud_erp_sy stems_a_review/links/5ffbad6292851c13fe02ecd5/differe nce-com-parison-of-sap-oracle-and-microsoft-solutions-based-on-cloud-erp-systems-a-review.pdf

[8] M. Franklin, A. Halevy, and D. Maier, "From Databases to Dataspaces," *SIGMOD Rec.*, 34(4):27-33, 2005, doi: 10.1145/1107499. 1107502.

[9] Fraunhofer-Gesellschaft, *International Data Spaces,* [Online], Available: https://www.fraunhofer.de/en/ research/lighthouse-projects-fraunhofer-initiatives/ international-data-spaces.html (accessed: Aug. 27 2020).

[10] *GAIA-X: Technical Architecture,* [Online], Available: https://www.data-infrastructure.eu/GAIAX/Redaktion/ EN/Publications/gaia-x-technical-architecture.html (accessed: Jul. 13 2021).

[11] M. Holweg, S. Disney, J. Holmström, and J. Småros,

"Supply Chain Collaboration," *European Management Journal*, 23(2):170-181, 2005, doi: 10.1016/j.emj. 2005.02.008.

[12] *Home | Catena-X,* [Online], Available: https://catena-x.net/en/ (accessed: Feb. 16 2023).

[13] International Data Space Association, *International Data Space Reference Architecture Modell,* [Online], Available: https://www.fraunhofer.de/content/dam/zv/en/fields-of-research/industrial-data-space/IDS-Reference-Architecture-Model.pdf, 2019.

[14] K. Kurbel and D. Nowak, "Customization of On-Demand ERP Software Using SAP Business ByDesign as an Example," *Innovation and Future of Enterprise Information Systems*: Springer, Berlin, Heidelberg, pp. 289-297, 2013, [Online], Available: https://link.springer.com/chapter/10.1007/978-3-642-37021-2_21

[15] W. Lehmann, "Präsentation Konzept Cantena-X," [Online], Available: https://catena-x.net/fileadmin/user_upload/Vereinsdokumente/Catena-X_UEbersicht.pdf, 2022.

[16] L. Lyu, H. Yu, J. Zhao, and Q. Yang, "Threats to Federated Learning," *Federated Learning*: Springer, Cham, pp. 3-16, 2020, [Online], Available: https://link.springer.com/chapter/10.1007/978-3-030-63076-8_1.

[17] J. A. Marmolejo-Saucedo, M. Hurtado-Hernandez, and R. Suarez-Valdes, "Digital Twins in Supply Chain Management: A Brief Literature Review," *2020 – Intelligent Computing and Optimization*, pp. 653-661, 2020.

[18] Microsoft Corporation, *Advancing Reliability through a Resilient Cloud Supply Chain | Azure Blog and Updates | Microsoft Azure,* [Online], Available: https://azure.microsoft.com/en-us/blog/advancingreliability-through-a-resilient-cloud-supply-chain/ (accessed: May 30 2022).

[19] Microsoft Corporation, *Supply Chain Management | Microsoft Dynamics 365,* [Online], Available: https://dynamics. microsoft.com/de-de/supply-chain-management /overview/ (accessed: May 30 2022).

[20] MongoDB, *MongoDB Data Lake,* [Online], Available: https://www.mongodb.com/de-de/atlas/data-lake (accessed: Feb. 16 2023).

[21] S. Oman, R. Leskovar, B. Rosi, and A. Baggia, "Integration of MES and ERP in Supply Chains: Effect Assessment in the Case of the Automotive Industry," *Teh. vjesn.*, 24(6):1889-1895, 2017, doi: 10.17559/tv-20160426094449.

[22] projects.eclipse.org, *Eclipse Dataspace Components,* [Online], Available: https://projects.eclipse.org/projects/technology.edc (accessed: Feb. 16 2023).

[23] P. Radanliev, D. C. de Roure, J. R. C. Nurse, R. M. Montalvo, and P. Burnap, "Supply Chain Design for the Industrial Internet of Things and the Industry 4.0," 2019.

[24] K. Rong, G. Hu, Y. Lin, Y. Shi, and L. Guo, "Understanding Business Ecosystem Using a 6C Framework in Internet-of-Things-Based Sectors," *International Journal of Production Economics*, 159:41-55, 2015, doi: 10.1016/j.ijpe.2014.09.003.

[25] SAP, *ERP-Software für kleine Unternehmen | SAP Business One,* [Online], Available: https://www.sap.com/germany/products/business-one.html (accessed: May 25 2022).

[26] SAP, *Übersicht Produkte | Unternehmenssoftware für KMU | SAP Mittelstand,* [Online], Available: https://www.sap.com/germany/products/sme-business-software/products.html (accessed: May 30 2022).

[27] A. D. Sarma, X. Dong, and A. Y. Halevy, "Data Modelling in Dataspace Support Platforms," *Lecture Notes in Computer Science*, *Conceptual Modeling: Foundations and Applications: Essays in Honor of John Mylopoulos*, A. T. Borgida, V. K. Chaudhri, P. Giorgini, and E. S. Yu, Eds., Berlin: Springer, 5600:122-138, 2009.

[28] K. Schneider, "SAP Business By Design – Integrated Cloud Suite," (in De;de), *HMD*, 55(1):44-61, 2018, doi: 10.1365/s40702-017-0380-8.

[29] S. Shah, M. Bolton, and S. Menon, "A Study of Internet of Things (IoT) and its Impacts on Global Supply Chains," *2020 International Conference on Computation, Automation and Knowledge Management (ICCAKM)*, Dubai, United Arab Emirates, pp. 245-250, Jan. 2020.

[30] J. Stentoft, K. W. Jensen, K. Philipsen, and A. Haug, "Drivers and Barriers for Industry 4.0 Readiness and Practice: A SME Perspective with Empirical Evidence" *Proceedings of the 52nd Annual Hawaii International Conference on System Sciences: January 8-11, 2019, Maui, Hawaii*. Honolulu, HI: University of Hawaii at Manoa Hamilton Library Scholar Space, 2019, [Online], Available: https://scholarspace.manoa.hawaii.edu/handle/10125/59952.

[31] J. Sunny, N. Undralla, and V. Madhusudanan Pillai, "Supply Chain Transparency Through Blockchain-Based Traceability: An Overview with Demonstration," *Computers & Industrial Engineering*, 150:106895, 2020, doi: 10.1016/j.cie.2020.106895.

[32] S. Wang and H. Wang, "Big Data for Small and Medium-Sized Enterprises (SME): A Knowledge Management Model," *JKM*, 24(4):881-897, 2020, doi: 10.1108/JKM-02-2020-0081.

[33] *Welcome | Superset,* [Online], Available: https://superset.apache.org/ (accessed: Feb. 16 2023).

[34] Q. Yang, Y. Liu, T. Chen, and Y. Tong, "Federated Machine Learning," *ACM Trans. Intell. Syst. Technol.*, 10(2):1-19, 2019, doi: 10.1145/3298981.

[35] ZDIN, *Zukunftslabor PRODUKTION,* [Online], Available: https://www.zdin.de/zukunftslabore/produktion (accessed: Jun. 27 2022).

[36] Jie Zhang Song Guo, Zhihao Qu, Deze Zeng, Haozhao Wang, Qifeng Liu, Albert Y. Zomaya, "Adaptive Vertical Federated Learning on Unbalanced Features," *IEEE Trans. Parallel Distrib. Syst.*, p. 1, 2022, doi: 10.1109/TPDS.2022.3178443.

Jonas Kallisch is a research assistant at the Emden/Leer University of Applied Sciences (Germany). He is researching the networking of manufacturing companies and their manufacturing infrastructures as part of the Future Lab Production. He holds a master's degree in industrial engineering. His research interests are data analytics, digital business ecosystems and federated data architectures.

Christoph Wunck is a full professor of Business Information System at Emden/Leer University of Applied Sciences (Germany). He serves as a member of the executive board "manufacturing" at OFFIS Institute of Information Technology in Oldenburg (Germany). He was appointed Affiliate Professor with Iowa State University, Department of Computer Science. He holds a doctoral degree from the Faculty of Mechanical Engineering and a master's degree (Diplom-Ingenieur) in Electrical Engineering, both from RWTH Aachen University. His research interests cover technical architectures for digital business ecosystems and analytics in manufacturing execution systems.