INTERNATIONAL DESIGN CONFERENCE – DESIGN 2022
https://doi.org/10.1017/pds.2022.71

Data Networking for Industrial Data Analysis Based on a Data Backbone System

A. Eiden [1,✉], T. Eickhoff [1], J. C. Göbel [1], C. Apostolov [2], P. Savarino [2] and T. Dickopf [2]

[1] Technische Universität Kaiserslautern, Germany, [2] CONTACT Software GmbH, Germany

✉ eiden@mv.uni-kl.de

Abstract

Industrial Data Analytics needs access to huge amounts of data, which is scattered across different IT systems. As part of an integrated reference kit for Industrial Data Analytics, there is a need for a data backend system that provides access to data. This system needs to have solutions for the extraction of data, the management of data and an analysis pipeline for those data. This paper presents an approach for this data backend system.

Keywords: data mining, big data analysis, product lifecycle management (PLM), data aggregation, data integration

1. Introduction

The spread of modern information and communication technologies (ICT) and the technological ability to systematically and comprehensively collect data in decentralized industrial networks make it possible to build dynamic information repositories of previously unknown size and quality (Eickelmann et al., 2015b). Using ICT-supported solutions makes it possible to open up previously isolated data areas and make the often highly heterogeneous data available in a standardized manner (Appelrath et al., 2014). The interpretation and efficient use of the knowledge implicit in these information repositories to support decision-making in processes of dynamic value networks is increasingly becoming the focus of companies (Geisberger and Broy, 2012; Deuse et al., 2014).

Against this background, modern, industrial data analysis approaches offer solution strategies as a key technology for value-creating, cross-company collaboration (Köksal et al., 2011; Harding et al., 2006; Niggemann et al., 2017; Eickelmann et al., 2015a). Thus, the integrated application of data-analytic methods enables efficient knowledge generation about processes, procedures, methods, and technologies to make them specifically usable, expand them, and make them available to all company divisions across the board (Abele and Reinhart, 2011). Thus, in addition to the possibility of a holistic increase in efficiency in the value creation network, this approach offers a high potential for opening up new forms of cooperation as a basis for developing innovative, data-driven services and business models (Gadatsch and Landrock 2017).

At the same time, however, there is a growing realization that companies are unable to use their resources to make sensible use of the necessary building blocks of data analysis technology and develop them further in a targeted manner, from standardized procedures and use cases to cross-network business model developments (Stark et al., 2014). As a result, there is a lack of the necessary competencies and implementation strategies in the companies themselves, as well as a lack of strategically oriented, possible service and technology offerings to enable SMEs in particular to use the massive potential of modern methods of data networking and analysis in a sustainable, low-cost and application-oriented manner.

2. Research Project AKKORD and Contribution

The overall goal of the research project AKKORD is the integrated usage of industrial data analysis for competency-driven collaboration in dynamic value-creation networks. The project addresses the sub-goals of value-creating collaboration, integrated and interconnected data analysis, development of personnel competencies and decision recommendation, and creating an integrated data basis considering the triad of human, technology, and organization.
The goals are to be reached by employing a modular reference kit that supports organizations in setting up and integrating dynamic collaboration for industrial data analysis. The modular kit covers the domains of Collaboration and Business Models, Competences and Action Recommendations, and Analysis Modules and Configuration. Function-wise the reference kit covers the fast and effortless setup and the use-case-specific customization of systems for industrial data analysis in dynamic value-creation networks. Therefore, an integral part of the project is developing a robust data backend system for the management and semantic linkage of heterogeneous data as part of the modular reference kit.

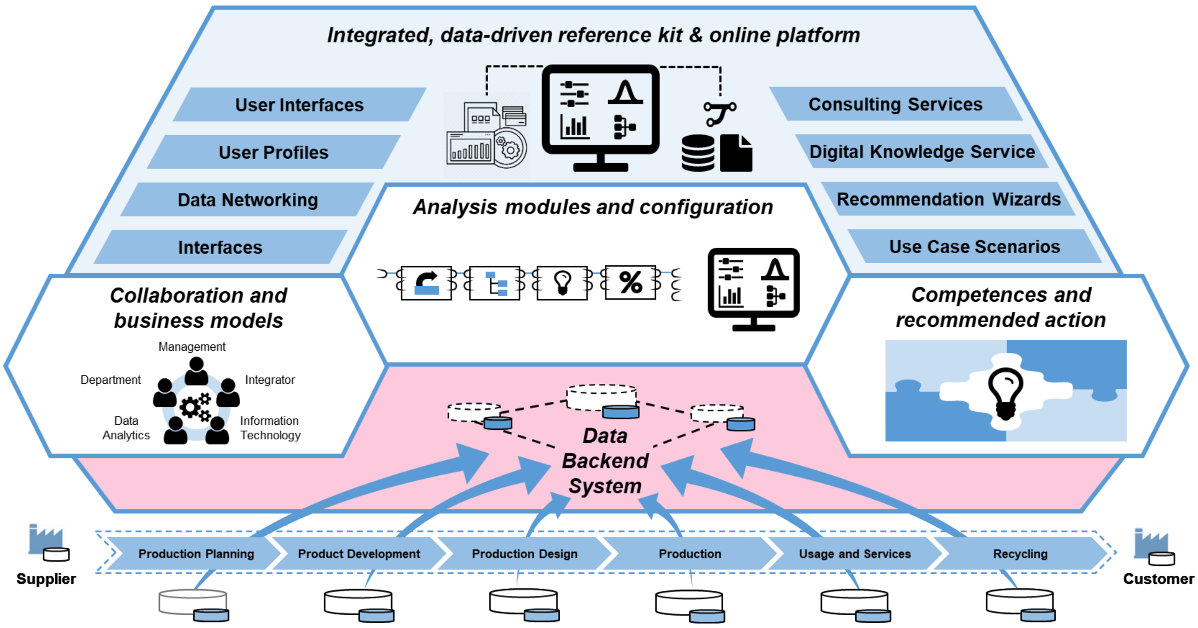

Figure 1. General setup of the AKKORD project

This contribution presents the work of creating a solution for the AKKORD data backend system as part of the reference kit described in this chapter. The results were developed according to the CASE research cycle (clarification – analysis – synthesis – evaluation) after Müller (Müller, 2013). Here, the initial position and the task of this contribution are clarified. Section three provides insight into the analysis phase results by giving a short overview of state of art in the corresponding fields. The current state of the approach's synthesis for creating a comprehensive and networked database for industrial data analysis is described in section four. Finally, an initial evaluation of the approach through a case study is implied in section five and appertains to our future intends and will be briefly addressed in the outlook section.

3. Current approaches in data integration and analysis

This chapter presents main results of our analysis according to the CASE research methodology. It was conducted mainly by studying literature and was supplemented by empirical interviews with project and research partners.

https://doi.org/10.1017/pds.2022.71 Published online by Cambridge University Press

When it comes to Industrial Data Science, there are different solutions for many aspects, but there is no holistic solution that manages data, processes, and analytics in a unified approach. Therefore, in the following, possible approaches for some aspects of Data Integration, Industrial Data Science, and platforms for those processes are presented.

Alongside the product lifecycle, dozens of tools generate or store data about products, processes, or surrounding information. For example, product Lifecycle Management (PLM) or Enterprise Resource Planning (ERP) software aims to store as much company-relevant information as possible inside their database. However, they are rarely used as a monolithic approach due to historical reasons. Instead, companies tend to use many parallel software tools for data management, even using more than one tool for the same task (Quirmbach, 2015) and up to a hundred tools to store data alongside the product lifecycle. Therefore, Data Integration has been a topic in science and industry for many years. One challenge is data availability, but the bigger one affects the mapping of different data sources and the data itself so that human users and machines can understand the meaning of data. There are three different solutions regarding this problem: Data mapping using ontologies and semantic web technologies (Woll et al., 2015, Abramovici et al., 2016), data mapping using standards (Eckert et al. 2005), and data mapping using RESTful web services (Franke et al., 2014, Pencius et al., 2014). Eickhoff et al. give a deeper insight into those three approaches (Eickhoff et al., 2020).

Another problem is the large variety of data structures. A heterogeneous IT architecture with various systems fulfilling different tasks has different data structures. For example, a well-structured bill of materials in mechanical engineering, but network-alike structures in software development (Eigner et al., 2016), or unstructured data from IoT sources. Inside the research project AKKORD, there are data sources from various phases across the product lifecycle. Engineering data is stored in a various specialized databases, quality data in an SQL database and business data inside an ERP system to name some examples.

Industrial Data Science combines computer science, statistics, and engineering science (Bauer et al., 2018). It aims to optimize products and processes on an industrial scale. Arnarsson et al. 2016, 2018) show potential use-cases in product development, Strauss et al. (2018) and Wöstmann (2019) show predictive maintenance applications of Industrial Data Science in brownfield situations. A considerable part of Industrial Data Science and process models like CRISP-DM (Cross-Industrial Standard Process for Data Mining) or the newer ASUM-DM (Analytical Solutions Unified Method) (IBM, 2016) is the data acquisition and pre-processing. As described above, data is scattered alongside the product lifecycle, and data integration tools need to be connected to Industrial Data Science tools. Current research projects try to develop platform concepts for Industrial Data Science that integrate as many process steps as possible. For example, (Wöstmann et al. 2020) developed a reference architecture for machine learning in the process industry. Such a reference architecture needs a data connection management, where pipelines of data streams can be established from databases to the training environment or the deployment environment of machine learning applications. Available commercial solutions like AWS, ADAMOS, AXOOM, MindSphere, Thingworx often support some data sources from one kind, like IoT. However, they cannot unify data from software alongside the product lifecycle like PLM, ERP and other data sources while providing a common interface for users to build and have insights into a metadata model.

4. Conceptual Approach

After the initial analysis, this chapter describes the synthesis of the research process according to the CASE methodology.

As part of the integrated reference kit for Industrial Data Analytics, the AKKORD data backend system must be suitable for a wide range of industrial use-cases and reflect the heterogeneity of company IT architectures. Therefore, the AKKORD data backend system does not provide just one tool but different approaches and respective tools to choose from. In the following, two approaches are focused. First, an integrated System Lifecycle Management (SysLM) Backbone (Eigner et al., 2014), which centralizes the data along the product lifecycle in one system and a semantic metadata repository providing links to the different data sources, connecting them over system borders. The SysLM Backbone is a conceptual solution based on PLM's technology and IT architecture. Therefore, it is best used in engineering-centric

https://doi.org/10.1017/pds.2022.71 Published online by Cambridge University Press

companies that want to expand the scope of their PLM implementation. On the contrary, the semantic metadata repository is best suited in companies, where the IT is very heterogeneous and can fit for any possible system, that makes its data accessible.
Both approaches provide suitable functions for data networking and data analysis, specific the ability to build up a data model of heterogeneous data and the ability to feed this data into an analysis pipeline. To do so, there are two main processes: First, the preparation of the backend itself, linking data and creating a semantic net and second, creating a data pipeline from the backend into data analytic modules. Both processes can be iterative, making the data backend a "living" tool that is extended on the fly.
The preparation of the data backend can be triggered at any time, but especially when preparing a new data analytics use-case. Figure 2 illustrates the process: The relevant data sources and their respective data models have to be selected (1). Then, using the semantic metadata repository, the data model can be generated inside the Data Model Canvas and be used as a basis for data connections between the backend itself and the data sources (2). If necessary data is spread across different source systems, but has a similar meaning or should be linked in any way, those connections are made visible inside the data backend (3). In the last step, additional information can be added to provide more value and insight into existing data (4).
Before the first process run, the data backend is an empty space, but as the process is iterative, it will build up further every time the process is run. As an alternative, the data understanding process can be run without a specific data analytics use-case in mind, just to connect and describe potentially all data sources one company possesses. If a company decides to follow the iterative or the complete alternative is a strategic decision, the AKKORD approach to build a data backend supports both.
In general, the first two steps of the process are conceptual and can be modelled on paper or with the help of an IT solution, the third and forth step have to be run inside a technical solution, either the SysLM backbone or the semantic metadata repository.
To help facilitate building the domain and IT data-models, we created a Data Model Canvas as an easy-to-use tool. Users can create data types and attributes by drawing graph nodes on a canvas and connecting them afterwars.

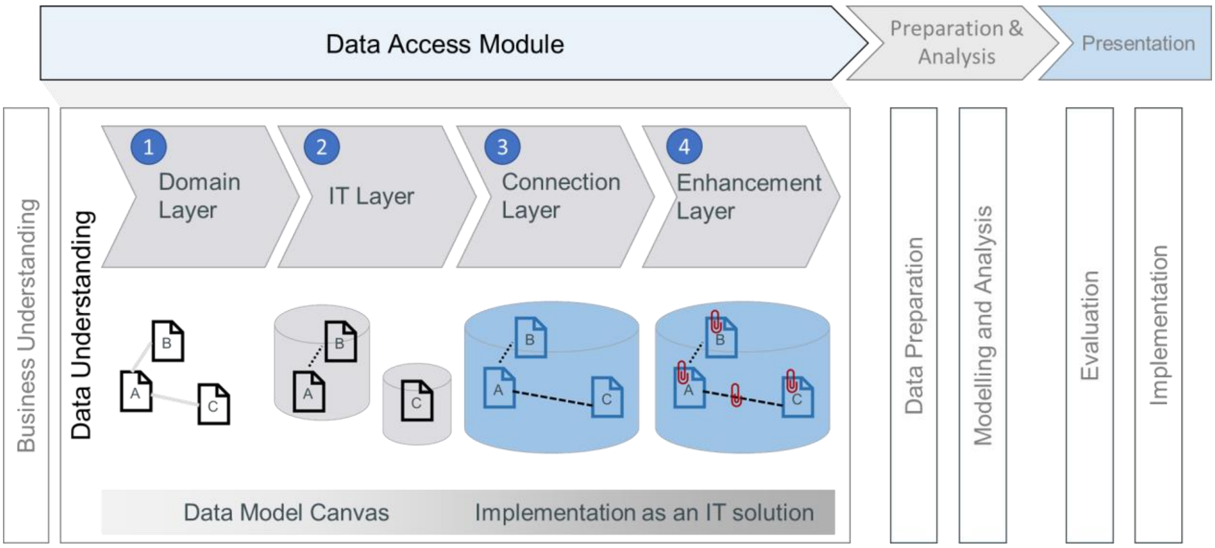

Figure 2. Data understanding process as part of the CRISP-DM

After building up the data backend and its data model, data analytics use-cases need pipelines for direct access to data. Therefore, the second process is now triggered. Data analysts can specify their data needs inside a file or start with a blank process. Either way, they can explore the available data models within the Data Model Canvas. It is possible to look directly in the source systems for further details and example data. Alternatively, they can rapidly build a data pipeline by choosing data types and attributes,

https://doi.org/10.1017/pds.2022.71 Published online by Cambridge University Press

creating a URL to access the data in .csv-format and looking further inside their data analytics solution. After checking the data quality, the data pipeline can be used permanently by storing the URL into a data needs file.

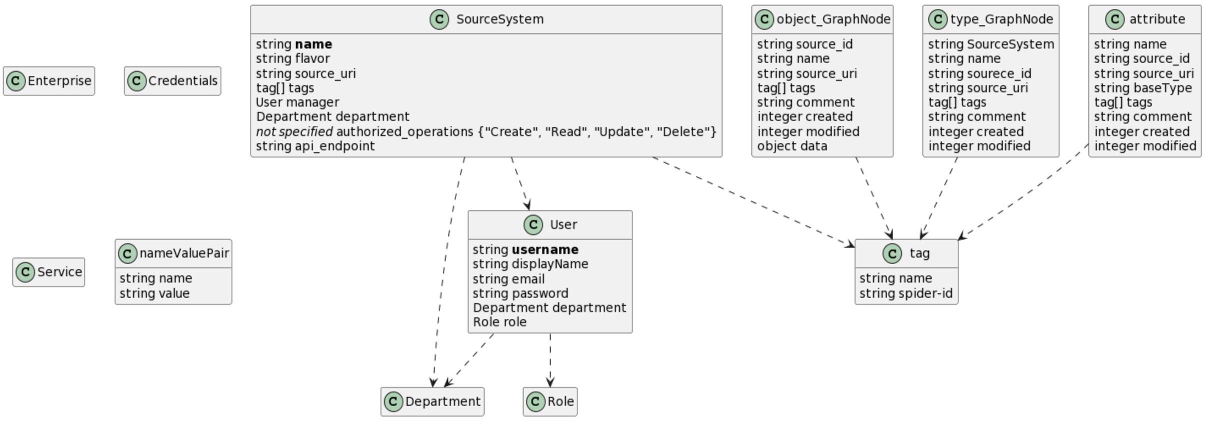

Figure 3. Metadata model

Aside the processes, the approach consists of a metadata model (cf. Figure 3), which allows to identify each data object within connected systems. The stored data is sufficient for building a data model for analysis and with the help of the system-specific connector sent data into an analysis pipeline, even if the underlying system is not capable of providing data by default. Another benefit of the metadata model is the possibility of tagging and commenting every object, making clear connections and providing additional context to data.

The data backend approach is a facilitation for data analysts, as it can directly access and provide data from all connected systems. The data access is live; therefore, the data analysis always has the most recent data at hand, which is a huge improvement to a workflow with downloaded files. Another benefit is data integration. In the traditional approach, data from different systems needs to be downloaded and put into the analytics process. The data backend can do this and therefore integrate data from different source systems and provide it to the data analytics process via one single interface. Consistency regarding the provision and management of data objects along the whole product lifecycle can be challenging, as the lifecycle stages have various requirements concerning data models and objects (cf. chapter 3). For example, organization and resource management follow a more generic management approach, while asset management is highly instance-centric. With the metadata model of the data backend system, these approaches and objects can be mapped and connected.

5. Prototypical Implementation and Evaluation

To evaluate the research, a prototypical implementation of the approach was evaluated according to the CASE research methodology. It is based on the tools of the different project partners (cf. Figure 4). As a SysLM Backbone solution, the CONTACT Elements Platform is used. Here the data from the different phases of the product lifecycle is managed and linked in a consolidated data backend. In addition, the VPE SP²IDER solution includes the possibilities of a semantic metadata repository and the data model canvas, which can also be used for the SysLM Backbone. The data analysis takes place in Rapidminer Studio.

While the CONTACT Elements platform consists of different modules that can address the requirement of managing different data types, it was used as a data backend system for prototypical implementation and evaluation of the concept described. As a result, some of the data backend modules provide and manage data objects of the early lifecycle phases and are covered by the product lifecycle management tool of the platform called CIM Database. In contrast, CONTACT Elements for IoT cover other modules focusing on real-time data objects (Figure 5).

https://doi.org/10.1017/pds.2022.71 Published online by Cambridge University Press

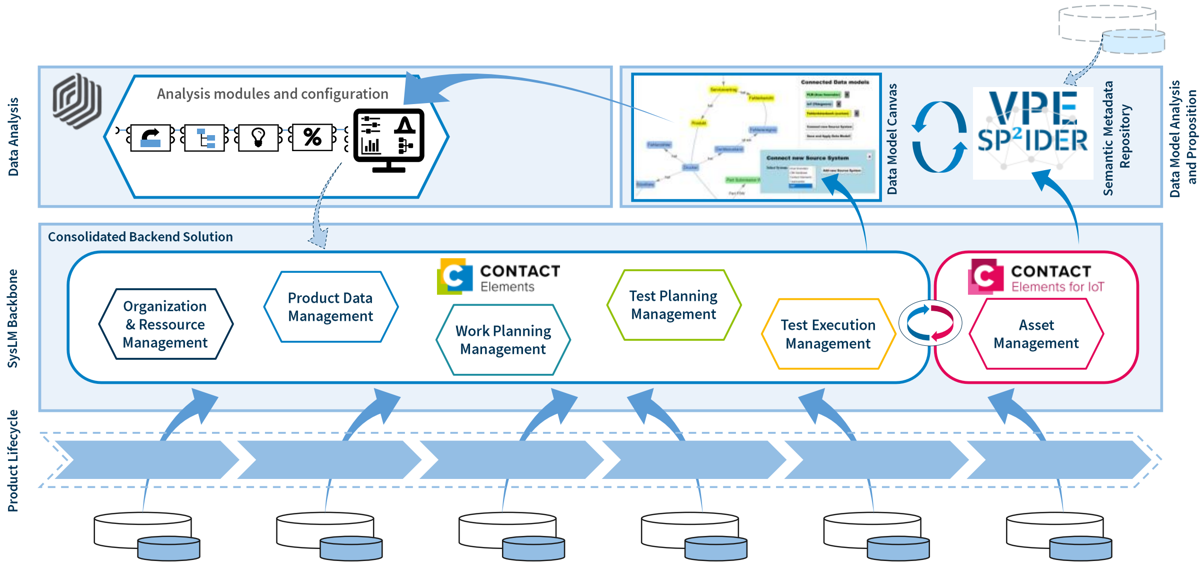

Figure 4. AKKORD Data backend system

Different data objects from modules represent data from different, heterogeneous sources of the product lifecycle in the CONTACT Elements platform: the product data management module consists of objects such as the product itself, which contains different parts on a generic level. The generic product must satisfy specifications that are part of requirements, which are tested in test runs in accordance with acceptance criteria and generate test results. Every tested part is instantiated by assigning a serial, referred to in a test order as part of the test execution management module. Work plans and test plans consist of one or more sequences (respectively test processes) and one or more operations (respectively test steps) and are part of the work planning or test planning module. A workplace represents the organization and resource management module, which is part of an organization or a plant to which persons, e.g., working personal, are allocated. Real-time data are represented by objects such as assets, which describe the digital twin of a product instance. Finally, the production order describes the sequence in which operations are executed. As part of different modules, these objects communicate with the modules for data analysis and the data model analysis and proposition.

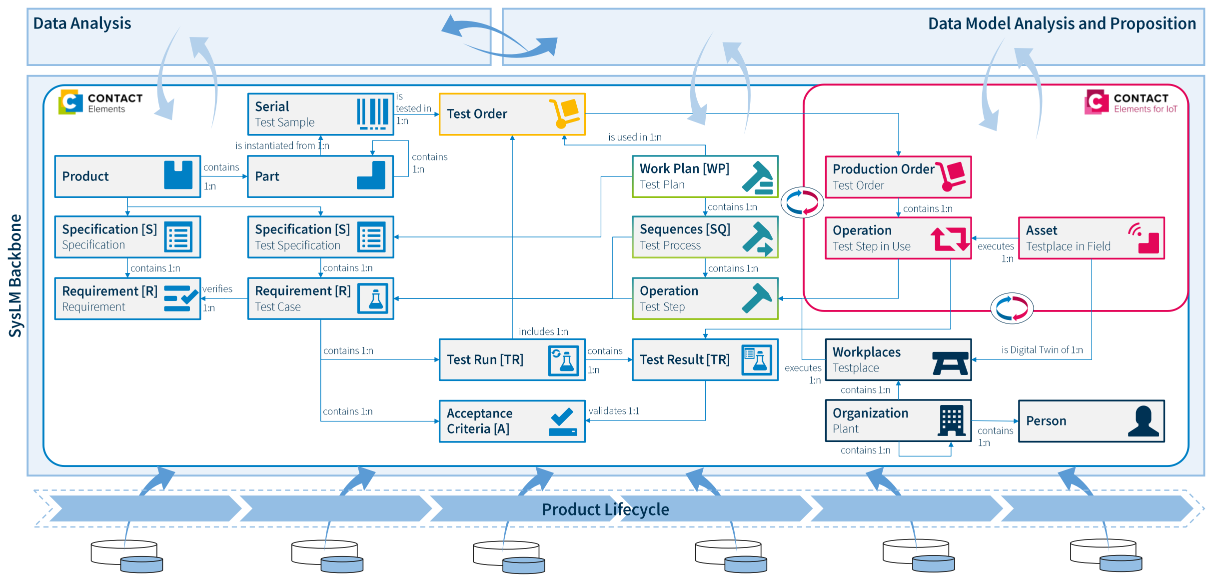

Figure 5. Possible data model of the SysLM Backbone

https://doi.org/10.1017/pds.2022.71 Published online by Cambridge University Press

As described above, the CONTACT Elements platform provides an extensive data model that allows to bring in heterogeneous data from various legacy sources in a state-of-the-art platform and link and semantically enrich these data. It also provides an intuitive, modern user interface for personal data consumption and manipulation, as shown in Figure 6. However, it is not always possible to bring all data needed for extensive industrial analysis into one or several closely integrated systems. In this case, a data integration pipeline based on the SP²IDER technology comes into play as described in the following.

A data linkage and provision pipeline to supply the analysis tool with data from various sources is built based on the SP²IDER and Data Model Canvas (DMC) technology. Figure 6 shows the schematically connected parts of the solution.

The prototype of the SP²IDER solution is a collection of different Flask server applications. Specific connectors provide access to data from the SysLM-backbone of Contact CIM Database and Elements for IoT, Influx DB, Aras and generic SQL databases. The primary user interface is the Data Model Canvas, a Python and Javascript web application where users can interact with the graph-based representation of data models and the data needs module to create data pipelines for analytics use-cases. The URLs can be tested in the browser for their functionality. The data access can then be transferred into Rapidminer for data analysis.

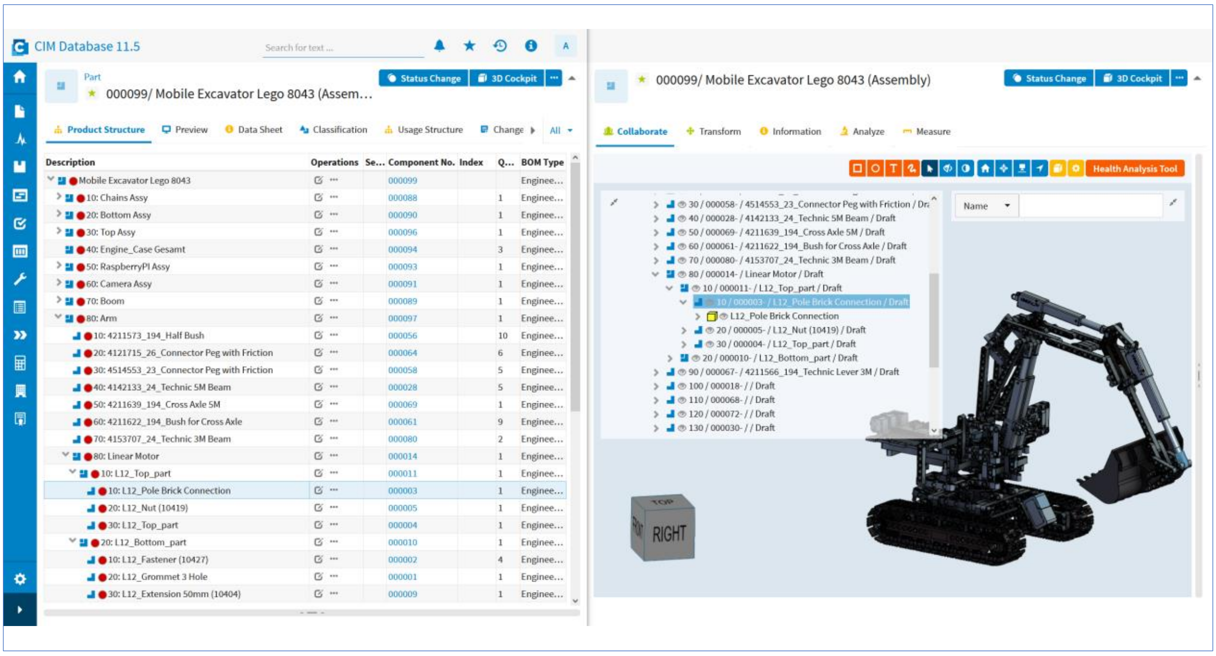

Figure 6. CONTACT CIM Database providing relevant PLM data for consumption to the VPE SP²IDER data pipeline (cf. Figure 5)

Figure 7 shows the AKKORD implementation loop with different data source systems, the SP²IDER data backend approach and gives a glimpse into the user interface of the Data Model Canvas. Behind every graph node is the aforementioned metadata model, which gives enough information to identify the data type and its attributes in the specific source system. This information is then extracted by the connector and the underlying data is transferred and accessed by a REST interface, which can be inserted into the analysis software of Rapidminer and then used for the analysis itself.

https://doi.org/10.1017/pds.2022.71 Published online by Cambridge University Press

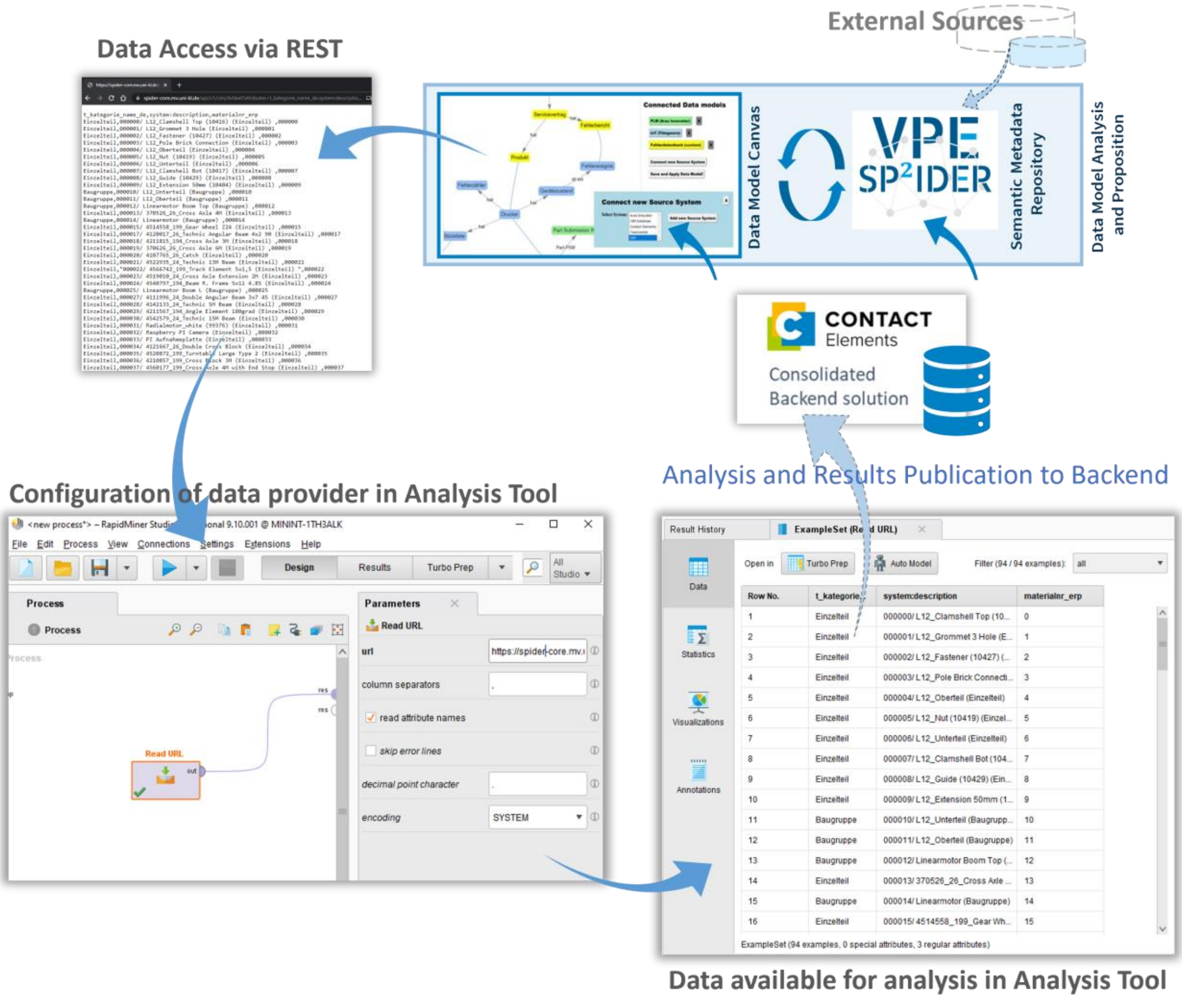

Figure 7. The AKKORD implementation loop with the VPE SP²IDER data pipeline for integrated data proposition to an analysis tool

6. Summary and Outlook

This contribution showed how industrial data analytics could be made tangible in the common cases of relevant data being spread between multiple systems and sources and where heterogeneous data models must be integrated. A state-of-the-art backend system with an extensive data model was developed based on CONTACT Elements technology to host all relevant data from the development to the usage phase of a product when bringing the data to a new system is feasible. Alongside the data availability to analysis, this provides many advantages in terms of data management and visualization capabilities. When this is not a feasible approach, it was demonstrated how an integrated data backend with linking information and an additional data pipeline with data model integration functions and REST endpoint generation could be built based on open technology.
Future work in the AKKORD project includes the connection of a staging area to the data pipeline for performance improvement, as well as the consideration of social and qualification aspects in the usage of the AKKORD data analysis toolkit.

Acknowledgment

This research and development project is funded by the German Federal Ministry of Education and Research (BMBF) within the "Innovations for Tomorrow's Production, Services, and Work" Program (funding number: 02P17D210/02P17D215) and managed by the Project Management Agency Karlsruhe (PTKA). The authors are responsible for the content of this publication. For more details about the project, visit www.akkord-projekt.de/en.

https://doi.org/10.1017/pds.2022.71 Published online by Cambridge University Press

References

Abele, Eberhard; Reinhart, Gunther (2011): Zukunft der Produktion. Herausforderungen, Forschungsfelder, Chancen. München: Hanser.
Abramovici, M., Göbel, J.C., Dang, H.B. (2016), Semantic data management for the development and continuous reconfiguration of smart products and systems. In: CIRP Annals 65, Nr. 1, S. 185–188
Appelrath, H.-J.; Kagermann, H.; Krcmar, H. (2014): Future Business Clouds. Ein Beitrag zum Zukunftsprojekt Internetbasierte Dienste für die Wirtschaft.
Arnarsson, Í.Ö., Gustavsson, E., Malmqvist, J., Jirstrand, M. (2018), Analysis of Engineering Change Requests using Markov Chains. In: Marjanović, D., Štorga, M., Škec, S., Bojčetić, N., Pavković, N. (Eds): Design 2018 Proceedings of the 15th International Design Conference, May 2018, Dubrovnik, Croatia. Zagreb, Fac. of Mechanical Engineering and Naval Architecture Univ, (DS / Design Society, 92), pp. 523–532
Arnarsson, Í.Ö., Malmqvist, J., Gustavsson, E., Jirstrand, M. (2016), Towards big-data analysis of deviation and error reports in product development projects. In: Boks, C. (Eds): Proceedings of NordDesign 2016, August 10-12, 2016, Trondheim, Norway. Bristol, United Kingdom: The Design Society, pp. 83–92
Bauer, N., Stankiewicz, L., Jastrow, M., Horn, D., Teubner, J., Kersting, K., Deuse, J., Weihs, C. (2018), Industrial Data Science: Developing a Qualification Concept for Machine Learning in Industrial Production. In: Archives of Data Science 5, P27 online
Deuse, J.; Erohin, O.; Lieber, D. (2014): Wissensentdeckung in vernetzten, industriellen Datenbeständen. In: H. Lödding (Hg.): Wie intelligente Vernetzung und kognitive Systeme unsere Arbeit verändern. Berlin: Gito (Schriftenreihe der Hochschulgruppe für Arbeits- und Betriebsorganisation e.V), pp. 373–395.
Eckert, R., Mansel, W., Specht, G. (2005), STEP AP233 + standard PDM = systems engineering PDM? In: 2005 IEEE International Technology Management Conference (ICE): IEEE, pp. 1-8.
Eickelmann, M.; Schallow, J.; Sousanabady, R. Jalali; Deuse, J. (2015a): Lebenszyklusübergreifende Qualitätsservices. Cloudbasierte Service-Plattform zur intelligenten Prognose qualitätsbestimmender Daten. In: ZWF 110 (4), pp. 167–171.
Eickelmann, M.; Wiegand, M.; Konrad, B.; Deuse, J. (2015b): Die Bedeutung von Data-Mining im Kontext von Industrie 4.0. In: ZWF 110 (11), pp. 738–743.
Eickhoff, T., Eiden, A., Göbel, J.C., Eigner, M. (2020), A Metadata Repository for Semantic Product Lifecycle Management. In: Procedia CIRP 91 (2020), pp. 249–254
Eigner, M., August, U., Schmich, M. (2016), Smarte Produkte erfordern ein Umdenken bei Produktstrukturen und Prozessen : Digitalisireung, Integration, Interdisziplinarität und Föderation.
Eigner, M., Dickopf, T., Apostolov, H., Schaefer, P., Faißt, K.-G., Keßler, A. (2014), System Lifecycle Management: Initial Approach for a Sustainable Product Development Process Based on Methods of Model Based Systems Engineering, Bd. 442. In: Fukuda, S., Bernard, A., Gurumoorthy, B., Bouras, A. (Eds): Product Lifecycle Management for a Global Market. Berlin, Heidelberg, Springer Berlin Heidelberg, 2014 (IFIP Advances in Information and Communication Technology), pp. 287–300
Franke, M., Klein, K., Hribernik, K., Lappe, D., Veigt, M., Thoben, K.-D. (2014), Semantic Web Service Wrappers as a Foundation for Interoperability in Closed-loop Product Lifecycle Management. In: Procedia CIRP 22 (2014), pp. 225–230
Gadatsch, A.; Landrock, H. (2017): Big Data für Entscheider: Entwicklung und Umsetzung datengetriebener Geschäftsmodelle: Springer Fachmedien Wiesbaden.
Geisberger, Eva; Broy, Manfred (2012): agendaCPS. Integrierte Forschungsagenda Cyber-Physical Systems. Berlin, Heidelberg: Springer Berlin Heidelberg.
Harding, J. A.; Shahbaz, M.; Srinivas; Kusiak, A. (2006): Data Mining in Manufacturing: A Review. In: Journal of Manufacturing Science and Engineering 128 (4), pp. 969–976.
IBM (2016), Analytics Solution Unified Method: Implementations with Agile principles.
Köksal, G.; Batmaz, İ.; Testik, M. Caner (2011): A review of data mining applications for quality improvement in manufacturing industry. In: Expert Systems with Applications 38 (10), pp. 13448–13467
Müller, P. (2013), Integrated Engineering of Products and Services – Layer-based Development Methodology for Product-Service Systems, Fraunhofer Verlag, Berlin.
Niggemann, O.; Biswas, G.; Kinnebrew, J.; Khorasgani, H.; Volgmann, S.; Bunte, A. (2017): Datenanalyse in der intelligenten Fabrik. In: B. Vogel-Heuser, T. Bauernhansl und M. ten Hompel (Hg.): Handbuch Industrie 4.0 Bd.2. Automatisierung. 2. Aufl. Berlin: Springer Vieweg.
Quirmbach, O.(2015), Umfrage Änderungsmanagement 2015.
Penciuc, D., Durupt, A., Belkadi, F., Eynard, B., Rowson, H. (2014), Towards a PLM Interoperability for a Collaborative Design Support System. In: Procedia CIRP 25, pp. 369–376
Stark, R.; Neumeyer, S.; Kim, M.; Deuse, J.; Schallow, J. (2014): Status Quo und Handlungsempfehlungen für die digitale Produktentstehung. In: ProduktDatenJournal 21 (2), pp. 12–17.

https://doi.org/10.1017/pds.2022.71 Published online by Cambridge University Press

Strauss, P., Schmitz, M., Wostmann, R., Deuse, J. (2018), Enabling of Predictive Maintenance in the Brownfield through Low-Cost Sensors, an IIoT-Architecture and Machine Learning. In: Abe, N. (Ed.): 2018 IEEE International Conference on Big Data, Dec 10-Dec 13, 2018, Seattle, WA, USA, proceedings, Piscataway, NJ, USA IEEE, pp. 1474–1483

Woll, R., Hayka, H., Stark, R. (2015), Ontologiebasierte Datenintegration für das Modellbasierte Systems Engineering. In: Maurer, M., Schulze, S.-O., Abulawi, J. (Eds): Tag des Systems Engineering: Bremen, November 12-14, 2014. München, Hanser, pp. 33–42

Wöstmann, R., Barthelmey, A., West, N., Deuse, J. (2019), A Retrofit Approach for Predictive Maintenance. In: Schüppstuhl, T., Tracht, K., Roßmann, J. (Eds.): Tagungsband des 4. Kongresses Montage Handhabung Industrieroboter. 1. Aufl. 2019. Berlin, Heidelberg, Springer Berlin Heidelberg, 2019, pp. 94–106

Wöstmann, R., Schlunder, P., Temme, F., Klinkenberg, R., Kimberger, J., Spichtinger, A., Goldhacker, M., Deuse, J. (2020), Conception of a Reference Architecture for Machine Learning in the Process Industry. In: Wu, X., Jermaine, C., Xiong, L. (Eds.), 2020 IEEE International Conference on Big Data: Dec 10 - Dec 13, 2020, Atlanta, Ga., USA, proceedings. Piscataway, NJ, USA, IEEE, 2020, pp. 1726–1735

Zhan, Y., Tan, K. H., Li, Y., Tse, Y. K. (2018), Unlocking the power of big data in new product development. In: Annals of Operations Research 270, 1-2, pp. 577–595