# Towards a Knowledge Graph-based Data Mesh for Smart Manufacturing

Irlán Grangel-González[1,*], Marc Rickart[2], Oliver Rudolph[2] and Rui Dias[3]

[1]Corporate Research, Robert Bosch GmbH, Renningen, Germany
[2]Robert Bosch GmbH, Automotive Electronics, Reutlingen, Germany
[3]Bosch Car Multimedia Portugal, S.A., Braga, Portugal

## 1. Motivation

Manufacturing business competition is driven by efficiency in order to offer the best price for products. Of paramount importance to achieve this efficiency is to get the right information at the right time. Manufacturing is a very complex process. To manage such a complex process a lot of data are required. These data are diversely spread out in different IT systems or silos, e.g., Enterprise Resource Planning (ERP), Manufacturing Execution Systems (MES), and Master Data (MD). These silos comprise no explicit semantics. They also contain differences in the way real-world concepts are modeled, i.e., Semantic Interoperability Conflicts (SIC) [1], thus hindering data re-usability. To tackle these problems Knowledge Graph (KG)-based applications have emerged. For instance, the Line Information System (LIS) [2] for manufacturing enables semantic harmonization, i.e., the resolution of SICs of data on production lines. However, despite this and other previous efforts at Bosch using KGs [3, 4, 5, 6, 7] for handling semantic harmonization many more data is being generated and consumed (cf. Figure 1). In addition, there are still no mechanisms to fulfill the FAIR principles [8] in manufacturing scenarios at Bosch. Of key relevance here is to have the FAIR principles in action, i.e., the data consumers should be capable of finding, accessing, and reusing data whenever required. Moreover, these data should be interoperable which remains as a huge challenge. To accelerate the data exchange and to meet the expectations of data consumers, it is required to move from an application mindset to a data centric one, where KG-based data products present concrete solutions to the manufacturing domain. Despite having just one data product in place, i.e. LIS, many more data from other domains than manufacturing are required by consumers.

## 2. KG-based Data Mesh for Manufacturing

To tackle the data reusability problems in manufacturing, we propose a KG-based data mesh [9, 10, 11]. Our approach has the KG-based products at its core resolving SICs and exposing

*SemIIM'23: 2nd International Workshop on Semantic Industrial Information Modelling, 7th November 2023, Athens, Greece, co-located with 22nd International Semantic Web Conference (ISWC 2023)*
irlan.grangelgonzalez@de.bosch.com (I. Grangel-González)

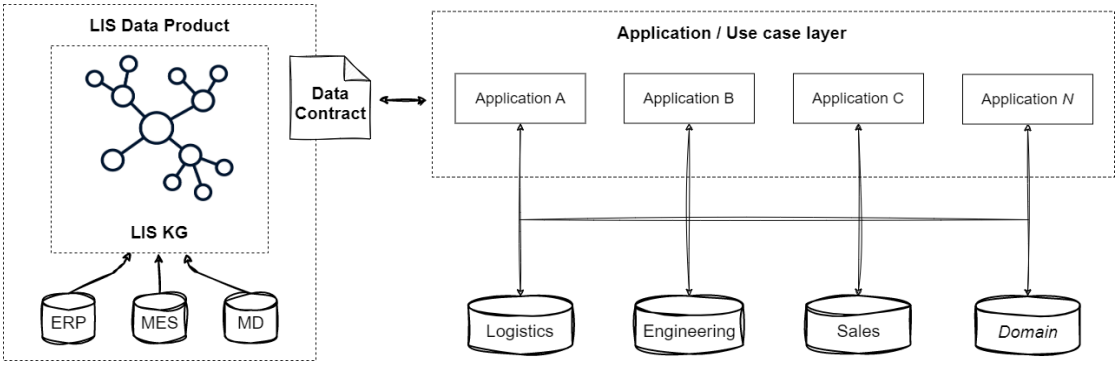


**Figure 1: Motivating Example**. LIS as a KG-based data product resolves the FAIR principles for just a certain part of the data silos in manufacturing. However, in reality, many more silos are needed to meet the demand of the smart manufacturing at Bosch.

semantically clean data to consumers (cf. Figure 2). Furthermore, being a source-oriented solution where users can discover who are the data owners, source origin, sample data sets or quality metrics. As being understandable means leveraging semantics to explain the syntax of the datasets, the format on how the data is presented is also important, e.g., serialization, queries to execute, proper ontologies. Then domain ownership is defined based on the data products. Furthermore, support functionalities need to be implemented, e.g., policies to be defined and governed, training, and consultancy of the organization, and provisioning the data as a self-service. The platform for data as a self-service provides all domains with their data products as host for all other consumers to integrate them in their applications. This as a starting point, the decentralized approach accelerates the deployment of the data mesh and thus faster data sharing and reusability via KGs and described ontologies.

## 2.1. Line Information System LIS as a data product

serves for instance other domains like product engineering or technology development for reuse in their respective data products driven by KGs. Common concepts in manufacturing as used in LIS are defined by the set of ontologies of the Core Information Model for Manufacturing (**CIMM**) [12]. The case of the Internal Defect Costs (IDC) as a KG-based solution has to deal with cost avoidance in a process failure for electronic products in the Surface-Mount Technology (SMT) area at Bosch. The typical approach for the IDC project would have been to reinvent the wheel by trying to semantically integrate manufacturing data that is already covered in the LIS data product. With our approach, several months of man-hours are saved due to the fact that the IDC project was able to reuse data out of the LIS KG. Having LIS as a data product enables IDC to reuse related relevant plant manufacturing data, e.g., materials, processes, machines, lines and even aggregated data by plant. This gives experts across all domains a deeper insight in defect costs along with reducing the time of decision making with more precise and accurate data for defining the product cost. Like this getting the edge on manufacturing business competition improves efficiency and best product price can be offered. Without LIS, there would be a danger of reverting to siloed data with continuous requests for data in file formats with costly human

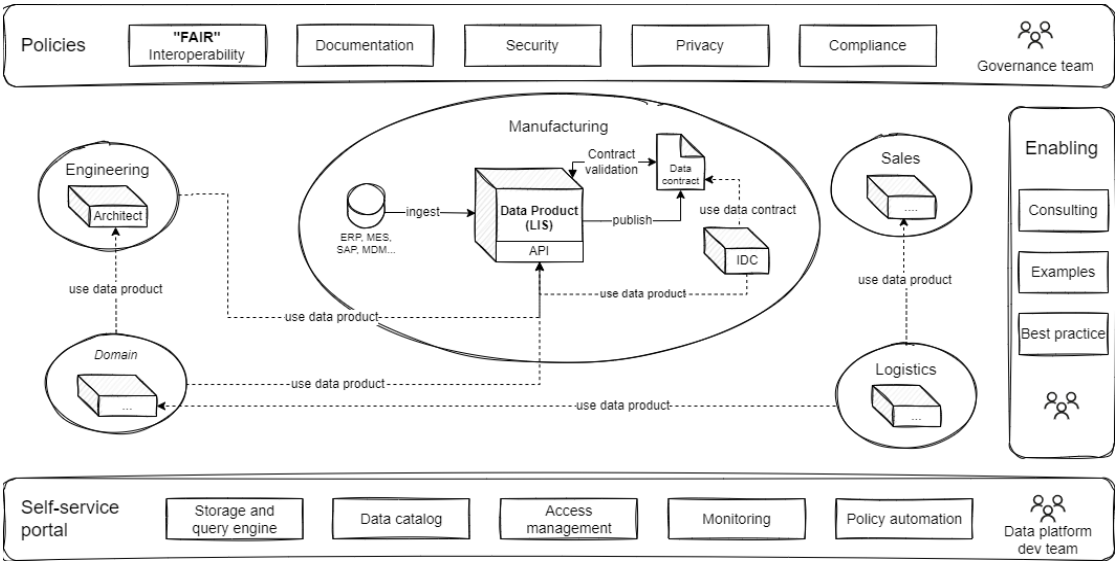


**Figure 2: KG-based Data Mesh for Manufacturing data**. In the center KG-based Data products like LIS offer semantically harmonized and clean data to other data products and data consumers. Different domains associated to manufacturing, e.g., Logistics, Engineering, and Sales, offer their data products which are interlinked with each other. In the side bar the evangelists deal with establishing best practices, standardization with concrete examples. This approach is only possible with a deep governance practice, thus, top and bottom bars refer to supporting functions, e.g., FAIR principles, Data catalog, Access management, Documentation, Data Security, etc. [10]

interaction, increasing data loss and time taken in decision making. Therefore, main driver is the focus on semantic integration of the data that are still not part of the LIS data product for further improvement work. For manufacturing domain this means higher performance with less cost and additionally data available which can be reused by other domains and applications.

## 2.2. Insights and feedback of the organization

In established market enterprises the competition is tough and use of data will provide an advantage. The organization in these enterprises is usually more hardware centric than data oriented and thus data receives a different prioritization as if it was the only source of income. Roles and responsibilities are equally different in that they are focused on the hardware product development and assembly, while information technology remains a support function only. In that setup, a strong lead on the tech stack and its application is missing. This creates the opportunity for the individual domains to establish their own tech stacks, thus resulting in a plethora of data storage technologies. As metadata has to be applicable to any and every source of at least the structured data, a decoupling of the metadata layer from the storage layer is advisable. Furthermore, implementing the FAIR principles nevertheless allows the organization to adapt faster to use of the data and metadata offered. At Bosch the data product LIS is offered to many other domains for reuse by APIs with defined data contracts. The data product is described by KG-based technologies semantically and provides links between fields

not necessarily on same data source system. In order to avoid SICs link prediction methods are embedded and by active use of the metadata system the instances themselves offered. Any data quality concern of mismatch in semantics will be spotted by all of the data consumers and can be fed back, while the other metadata system can simply be ignored and data consumers may start having their own description tables put in place. That is exactly what we observe in our organization. As for future work, we envision to enable further KG-based data products, e.g., for engineering, logistics, and sales to be able to cover a wider range for manufacturing applications.

## References


[1] I. Grangel-González, M. Vidal, Analyzing a Knowledge Graph of Industry 4.0 Standards, in: J. Leskovec, M. Grobelnik, M. Najork, J. Tang, L. Zia (Eds.), Companion of The Web Conference, Virtual Event / Ljubljana, Slovenia, April 19-23, ACM / IW3C2, 2021, pp. 16–25.

[2] I. Grangel-González, M. Rickart, O. Rudolph, F. Shah, LIS: A knowledge graph-based line information system, in: C. Pesquita, E. Jiménez-Ruiz, J. P. McCusker, D. Faria, M. Dragoni, A. Dimou, R. Troncy, S. Hertling (Eds.), The Semantic Web - 20th Int. Conf., ESWC 2023, Hersonissos, Crete, Greece, May 28 - June 1, Proceedings, volume 13870 of *LNCS*, Springer, 2023, pp. 591–608.

[3] I. Grangel-González, F. Shah, Link Prediction with Supervised Learning on an Industry 4.0 related Knowledge Graph, in: 26th IEEE Int. Conf. on Emerging Technologies and Factory Automation, ETFA, Vasteras, Sweden, September 7-10, IEEE, 2021, pp. 1–8.

[4] E. G. Kalayci, I. Grangel-González, F. Lösch, G. Xiao, A. ul Mehdi, E. Kharlamov, D. Calvanese, Semantic integration of Bosch manufacturing data using virtual knowledge graphs, in: J. Z. P. et al. (Ed.), 19th Int. Semantic Web Conf., Athens, Greece, November 2-6, Proceedings, Part II, volume 12507 of *LNCS*, Springer, 2020, pp. 464–481.

[5] M. N. Mami, I. Grangel-González, D. Graux, E. Elezi, F. Lösch, Semantic data integration for the SMT manufacturing process using SANSA stack, in: A. H. et al. (Ed.), The Semantic Web: ESWC 2020 Satellite Events, Heraklion, Crete, Greece, May 31 - June 4, volume 12124 of *LNCS*, Springer, 2020, pp. 307–311.

[6] A. Mehdi, E. Kharlamov, D. Stepanova, F. Loesch, I. Grangel-Gonzalez, Towards Semantic Integration of Bosch Manufacturing Data, in: Proc. of ISWC, 2019, pp. 303–304.

[7] B. Zhou, Z. Zheng, D. Zhou, G. Cheng, E. Jiménez-Ruiz, T. Tran, D. Stepanova, M. H. Gad-Elrab, N. Nikolov, A. Soylu, E. Kharlamov, The data value quest: A holistic semantic approach at Bosch, in: P. G. et al. (Ed.), The Semantic Web: ESWC Satellite Events - Hersonissos, Crete, Greece, May 29 - June 2, Proceedings, volume 13384 of *Lecture Notes in Computer Science*, Springer, 2022, pp. 287–290.

[8] L. Gleim, J. Pennekamp, M. Liebenberg, M. Buchsbaum, P. Niemietz, S. Knape, A. Epple, S. Storms, D. Trauth, T. Bergs, C. Brecher, S. Decker, G. Lakemeyer, K. Wehrle, Factdag: Formalizing data interoperability in an internet of production, IEEE Internet of Things Journal 7 (2020) 3243–3253. doi:10.1109/JIOT.2020.2966402.

[9] V. K. Butte, S. Butte, Enterprise Data Strategy: A Decentralized Data Mesh Approach, in: Int. Conf. on Data Analytics for Business and Industry (ICDABI), 2022, pp. 62–66.
[10] L. V. Jochen Christ, S. Harrer, Data Mesh Architecture, https://www.datamesh-architecture.com, 2022. Accessed: 10-03-2023.
[11] I. A. Machado, C. Costa, M. Y. Santos, Data Mesh: Concepts and Principles of a Paradigm Shift in Data Architectures, in: M. M. Cruz-Cunha, R. Martinho, R. Rijo, D. Domingos, E. Peres (Eds.), CENTERIS 2021 - Int. Conf. on ENTERprise Information Systems Information Systems and Technologies, Braga, Portugal, volume 196 of *Procedia Computer Science*, Elsevier, 2021, pp. 263–271.
[12] I. Grangel-González, F. Lösch, A. ul Mehdi, Knowledge Graphs for efficient integration and access of manufacturing data, in: 25th IEEE Int. Conf. on Emerging Technologies and Factory Automation, ETFA, Vienna, Austria, September 8-11, IEEE, 2020, pp. 93–100.