

Invented for life



Moving towards the ideal data strategy in manufacturing

Establish long-term success with digital twins and semantics!

A white paper from Bosch Connected Industry



Table of contents

Management Summary	3
1. Introduction	4
2. A coherent data strategy ensures competitiveness	6
3. The role of semantic data structuring	8
4. Digital twins: the digital data representatives	9
5. Rethink to implement	10
6. User-oriented solution: Bosch Semantic Stack	12
6.1 Blueprint architecture for value-adding solutions	13
6.2 Practical use cases	14
7. Digital twins as future enablers for Agentic AI: efficient data interaction without complexity	17
8. Conclusion: the time is now	20

Management Summary

Products are increasingly shifting into focus as part of the digitalization of manufacturing processes. Thanks to IoT capabilities, the connection to manufacturers no longer ends with the delivery of the products but continues at best throughout the entire lifecycle. Backflowing data streams form the basis for product optimization and digital services – and make it possible to meet new legal requirements such as the digital product passport.

To keep up with these developments as efficiently as possible and to remain competitive, enterprises need a coherent data strategy, i.e. a structured and goal-oriented framework that enables them to use data as a business asset. An important element is a uniform understanding of the meaning (and context) of data. Semantics makes data universally usable regardless of its source and solves one of the biggest current challenges many companies are facing.

This white paper describes the technologies required to implement a coherent and future-proof data strategy that homogenizes data both vertically – i.e. from source to application – and horizontally across the entire lifecycle. The digital twin approach – i.e. the generation of digital representatives of real products, product types, and instances – is at the heart of this approach. Three practical examples show how digital twins and semantics can be used to create significant added value. With the increasing role of artificial intelligence, the added value of digital twins and semantics becomes even more apparent. The more uniform and semantically rich the data is described, the better AI models can interpret it and use it for informed decisions. Semantic description enables effective data interaction because the meaning and context of the data are defined in a machine-readable way.

If they start implementing a coherent data strategy now, companies not only set the course for compliance with future legal requirements, but also lay the indispensable foundation for future success.

1.

Introduction

Digitalization has picked up speed and Industry 4.0 has long become a reality – at least as far as the available database is concerned. The progressive connectivity of production and logistics, as well as the IoT capability of many devices provide the data foundation for innovative solutions that were not possible in the “manual age”: be it tracking a sold product over its entire lifecycle or analyzing large amounts of data from many batches for optimization purposes.

The product itself is now the new linchpin, instead of just the data collected from production machines and quality assurance processes during manufacturing. Every part sold, every machine and technical component, as well as the associated IT systems have become their own individual data source, providing valuable information every day. Whether this data is used to generate added value, however, essentially depends on the maturity of a company’s data handling. To get to the point where valuable insights can be generated from the ever-accumulating mountain of data, a coherent data strategy is needed more than anything else.

A data strategy defines the way a company collects, processes, stores, and uses data. It is future-oriented, designed to expand internal and external connectivity, and to offer and utilize data-driven services as part of a digital, industry-specific ecosystem. The aim of the data strategy is profitable use, in other words a measurable contribution to the success of the respective product on the market. Semantics, i.e. a uniform understanding of the meaning (and context) of the data, is a key aspect. Only standardized semantics make it possible to utilize the data universally, compare it, or use it independently of the data source – a basic prerequisite for the implementation of upcoming legal requirements such as the digital product passport or the calculation of the product carbon footprint (PCF).

Digital twin technology makes this possible. Digital twins are representatives of real products, product types, and product instances that bring together data distributed in different data silos and ensure the semantic structuring of large volumes of data. Both semantics and the closely interlinked digital twin approach are therefore a prerequisite for implementing a coherent data strategy. Without such innovative technologies, the necessary solutions for “data challenges” can only be developed selectively and at great expense, without offering further scaling options or exploiting the full potential of the data.

2.

A coherent data strategy ensures competitiveness

Data is one of the essential components of every manufacturing company and is already used extensively in everyday work. Many processes, particularly in production and in internal and external logistics, are based on decentralized data processing, where data is stored and used “close to the source” or only within the relevant department. The backflow of data from the field, however, is relatively new. It relies on internet-enabled devices with a sporadic or even permanent connection to the producing company. But here too, the traditional approach often still prevails, i.e. the data is kept in its respective silos instead of being made available across departments. The potential that lies dormant in this data is the key to product development and optimization, as well as the provision of data-based services – two important areas that are crucial for a company’s future viability in global competition.

What is a data strategy?

A data strategy is defined within a company as a well-structured and goal-oriented framework that enables data to be used as a business asset. It is the basis for optimizing existing business activities, developing new business opportunities, and optimizing the user experience – all based on data and the information derived from it.

An adequate balance between the added value generated by digitalization and the costs of digital transformation is a decisive factor. Yet one thing is certain: deciding against data-driven digitalization is no longer an option in most industrial sectors. Instead, it is now time to decide how to implement digital transformation. A data strategy geared towards future challenges has a significant impact on the use of costs and resources – and on keeping these as low as possible in the long term. In comparison, a growing accumulation of individual solutions requires significantly more financial and human resources.

Moreover, it is difficult to provide innovative solutions, especially in a market environment where speed is crucial and new global players with a strong IT background are involved. The latter offer products at a high level of innovation, with digital support and services, over-the-air (OTA) update functions, and fast development cycles. To keep pace, the course must be set early for a future-oriented approach – in other words, a coherent data strategy. When choosing the right partner for IT, it may seem tempting to rely on external software specialists who are now also entering the manufacturing industry. However, it is more advisable to focus on solid domain expertise to avoid the classic vendor lock-in and lose control of software-driven functions. In-depth knowledge of production processes and the prospect of long-term cooperation on an equal footing are decisive.

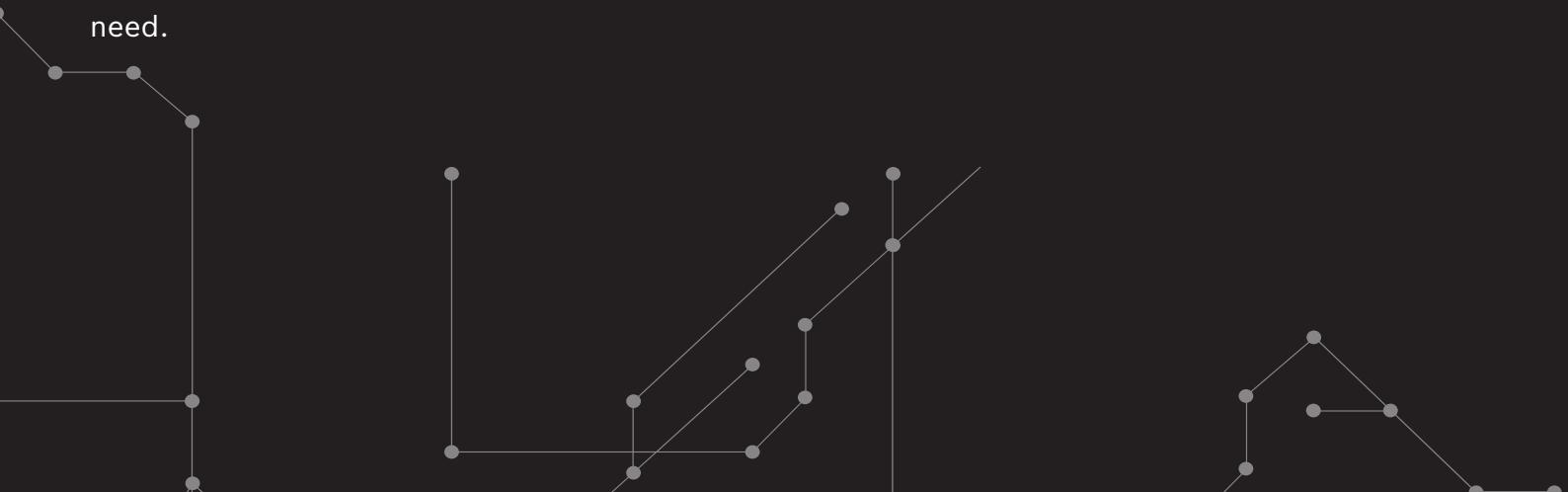


3.

The role of semantic data structuring

Specific digitalization challenges are often addressed with one-off solutions that are highly innovative. However, rolling them out across the entire company, i.e. scaling them up to other areas and products, can be problematic if there is no generally accepted way of handling data. The costs generated by such a solution, particularly in development and data pre-processing, cannot be allocated. If another division wants to offer a similar solution, development starts all over again – with high costs for data retrieval and preparation. Even a direct connection to the successfully established solution is not possible without further effort, as both the data format and the interface are specific to the machine, product, or application.

Semantic data homogenization gives all company data a universal language. No matter where a data package is generated in the company, it can be found, retrieved, understood, and processed at any time. This data is enriched with all contextual information, regardless of its physical source. For example, a data series with the digits 123 135 167 199 is only valuable if it is also known that these are temperature specifications in degrees Celsius. Semantic data structuring can also be used to create so-called aspects that provide very specific information geared towards user groups, such as (dynamically generated) errors and status data, as well as master data and historical information from the development and production period. Software can access these aspects in a targeted manner and provide employees with exactly the information they need.



4.

Digital twins: the digital data representatives

Semantic data structuring is the core of the digital twin approach, i.e. generating a digital data twin of real products – whether from a complete machine, a tool, or a single sensor. All aspects are collected in a twin so that they can be accessed via a central point, which also simplifies rights and access management. To support a data-driven lifecycle, digital twins are ideally created at the same time as their physical counterparts, namely during production – or even during the design phase of a product. They accompany the product through production and use until it is decommissioned. They serve as a central point for organizing and structuring access to all data during this entire period – which can last decades, depending on the product.

Asset Administration Shell (AAS)

The Asset Administration Shell is the implementation of the digital twin concept in Industry 4.0, with a strong focus on cross-manufacturer interoperability. The AAS is designed as an open-source solution and is intended to establish itself worldwide as an industry-neutral standard. The Industrial Digital Twin Association (IDTA) was founded specifically for this purpose.



5.

Rethink to implement

Implementing such a comprehensive approach requires a technology stack that sometimes involves handling and pre-processing data in a way that differs from the previous method. However, companies must prepare themselves for the digitalization journey. This includes making optimal use of scaling effects for a fast time to market (or “time-to-proof-of-value” for applications in internal optimization). Semantic data structuring is also the basis for software solutions that are built on merging data from different phases of the product lifecycle or the value chain. The lack of a coherent data strategy makes it difficult to implement a dynamic and iterative backflow, for example of information obtained from field data, into the engineering or production phase. The goal is therefore to include all dimensions in the data strategy: both vertically (from the individual data source to the application) and horizontally (from the production phase to recycling). This opens up all possibilities and reduces integration and scaling efforts.

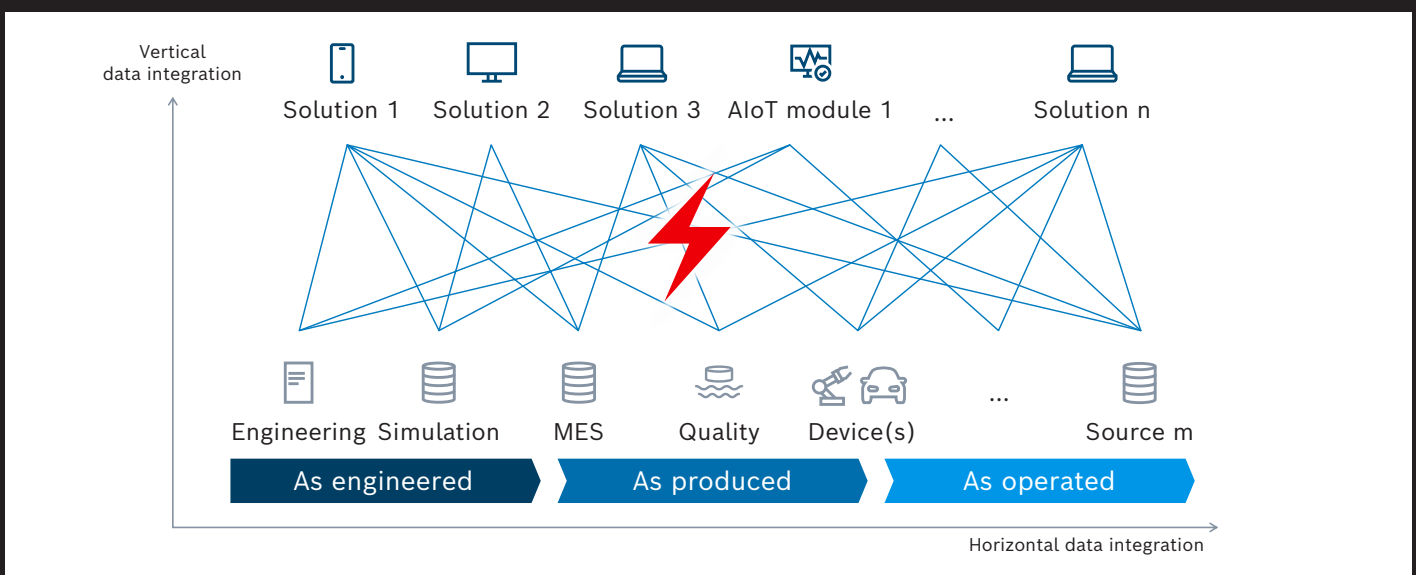


Figure 1. Challenges of data integration throughout the entire product life cycle

The data strategy, and consequently semantic homogenization, does not have to be rolled out company-wide overnight. It should, however, define the framework conditions and set a uniform course across all existing data silos. The new approach is introduced step by step, starting with individual projects that are initially more complex to implement due to the unfamiliar approach, for example for a specific product group or lifecycle phase. Each additional project that follows this strategy expands the network of mutually usable and readable data. The implementation of new data sources becomes easier, while the added value and the number of solutions increases exponentially. If this strategy is implemented consistently, the speed of innovation increases significantly over time, while the resources and costs involved continue to decrease.

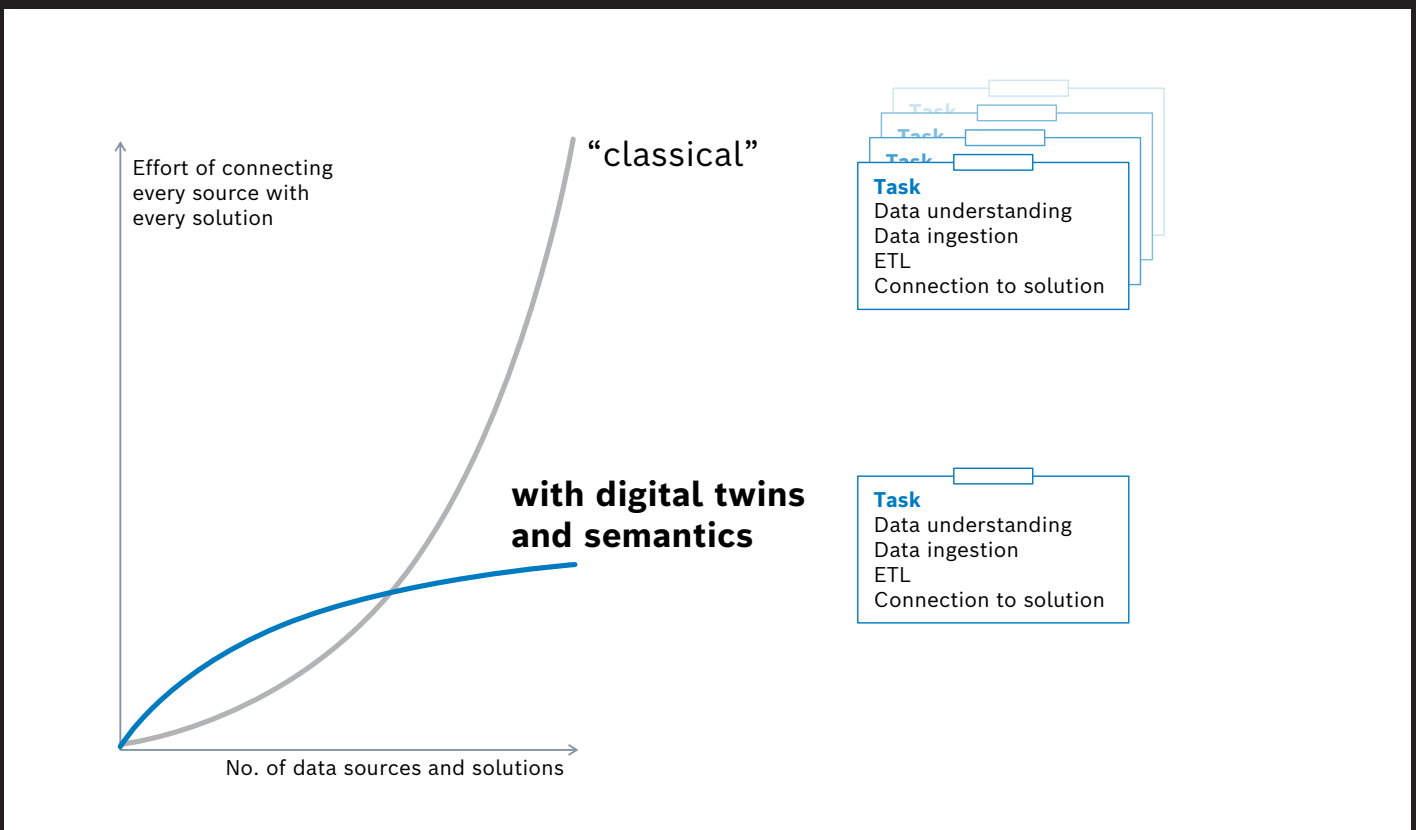


Figure 2. The initial effort for the first use cases, which is greater than with a traditional approach, decreases with each additional connected data source or solution.

Let's take an update manager as an exemplary lighthouse project that can access the digital twins of thousands of products in the field. Another product group can now be added with little effort, or the solution can be transferred to another data pool that has the same semantic structure. The number of connected products grows, and another solution can be added, such as a tool that analyzes the error rate of updates and generates new insights. In the case of new legal requirements regarding update tracking, the corresponding interface to access this information can also be implemented quickly and easily.

6.

User-oriented solution: Bosch Semantic Stack

Bosch recognized the industry's need for a holistic approach to data structuring at an early stage. The experts benefit from their experience as leading users and leading suppliers in the production of industrial components and know exactly how to develop practical solutions. The aim is to establish a comprehensive technology stack that makes all data along the entire product lifecycle transparent, understandable, and easily retrievable. The approach must be universal to integrate seamlessly into existing IT architectures and to map company specifications. This is facilitated by the integration of numerous open source components and common standards, as well as cross-company collaboration in many manufacturing industry initiatives.

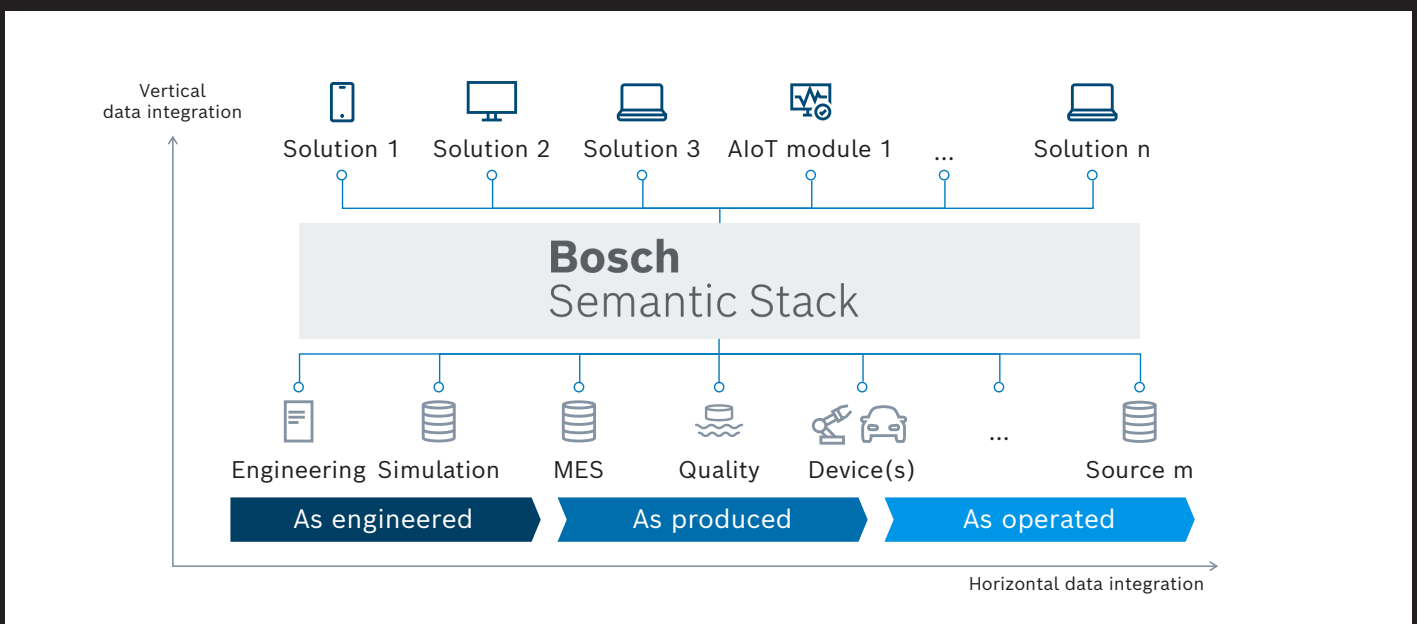


Figure 3. Bosch Semantic Stack as solution for data description and homogenization

Innovative solutions in the field of data description and homogenization, primarily digital twin technology, knowledge graphs (networks between the individual twins), and data lake house

storage architectures, were brought together and combined in the Bosch Semantic Stack portfolio. All three are based both on the Semantic Stack tools provided by Bosch and on open source approaches. They are designed to grow in an open ecosystem. Bosch Semantic Stack is not a self-contained product. Instead, it is in constant, dynamic development together with its users and their real-life challenges. It provides the semantic framework for developing software applications tailored precisely to the respective requirements based on all existing product data from the entire lifecycle. In the long term, Bosch Semantic Stack enables ever faster development of solutions, simple implementation of data sources, and free scaling within a company.

6.1 Blueprint architecture for value-adding solutions

The Bosch Semantic Stack tools and the generated artifacts result in a blueprint architecture. It enables the use of product data across the entire lifecycle – adapted to a specific use case or a particular application. The schematic diagram illustrates the layers between the connected data sources from development, production, and operation, as well as the user interfaces. This underlines the crucial role of tooling and open source components, both for the integration of data and for its semantic structuring. Once the architecture has been implemented, any number of data sources can be connected and value-added solutions can be set up in an ever shorter time. An already connected data pool is available to new applications, which expands the scope of possible functionalities and makes it possible to integrate more and more areas of the lifecycle, enabling even complex solutions to be set up quickly.

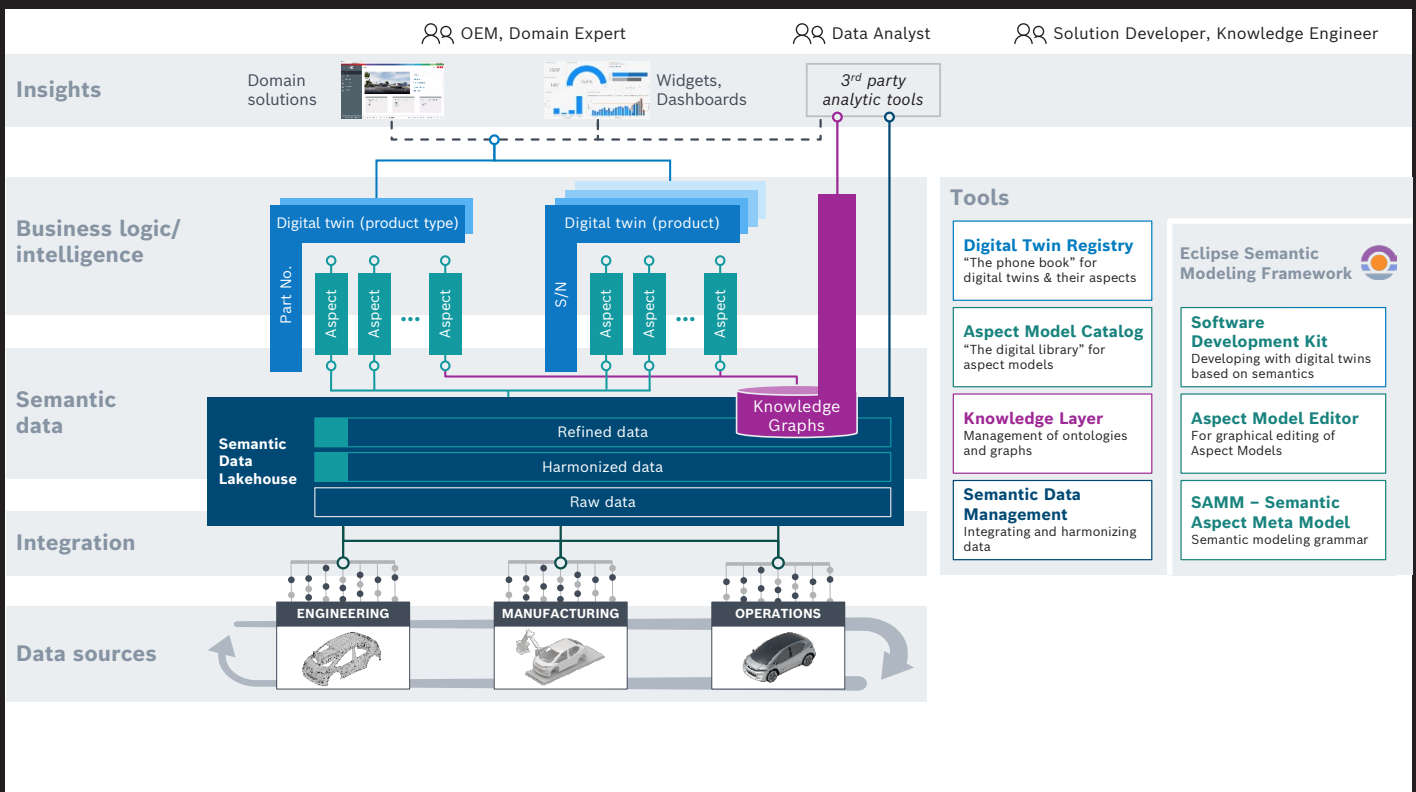


Figure 4: Blueprint architecture for vertical and horizontal data homogenization with Bosch Semantic Stack

6.2 Practical use cases

Semantic data structuring often involves a lot of theory – which is necessary, since it is initially about generally valid, formal descriptions. The technology stack, whether Bosch Semantic Stack or another approach, provides the basis for creating digital twins, aspects, and semantic descriptions, so-called aspect models. Practical use cases are needed for concrete implementation, which can run through the exemplary process. The introduction of semantic data structuring should also take place in this way: start with a use case that generates immediate added value and enables the workforce to understand the purpose of the transformation.

The full potential of this new data strategy will unfold in subsequent projects when the increasingly dense network of homogenized data makes it possible to answer increasingly complex questions. Bosch Semantic Stack allows any digital twin to act as an administration shell. This industry standard developed by the IDTA is easy to use for external developers and software thanks to fully available open source documentation. The following examples show how this translates into practice.

PLM-Portal: starting a comprehensive digitalization strategy

Bosch Rexroth manufactures innovative, industrial equipment that is often ordered very individually, starting from a batch size of one. While the variety of products is a great advantage for customers, staff faces the challenge of finding information about a specific product, which may have been manufactured several years ago, within a reasonable time frame. Of course, every company wants to avoid customers having to wait for an answer, especially if there is an urgent reason for the request.

To overcome this challenge, the master data of each product should be consolidated on a single platform, the PLM portal. This is where the digital twin technology of the Bosch Semantic Stack portfolio comes into play: a digital twin is created during product development and accompanies the physical products throughout their entire lifecycle. Thanks to the globally unique Global Thing ID (GLTID), the data of millions of products can be clearly identified and assigned to the right asset. New products are added daily to the 270 million already recorded. The individual digital twins are also continuously expanded, for example by integrating additional field data. This creates an ever more complete picture of the respective product, with a previously unattainable level of detail and consistency.

What may initially “only” sound like a very comprehensive product directory has its true potential in the wealth of new, easy-to-implement applications and use cases. Thanks to the PLM Portal, all data packages are semantically structured in the same way, i.e. the acquisition of additional product-related knowledge is no longer an obstacle. While all data had to be laboriously compiled and processed to develop an additional function in the past, a homogenized data directory is now available for comprehensive use. This also makes it possible to identify dependencies, trends, or cross-connections that were previously hidden.

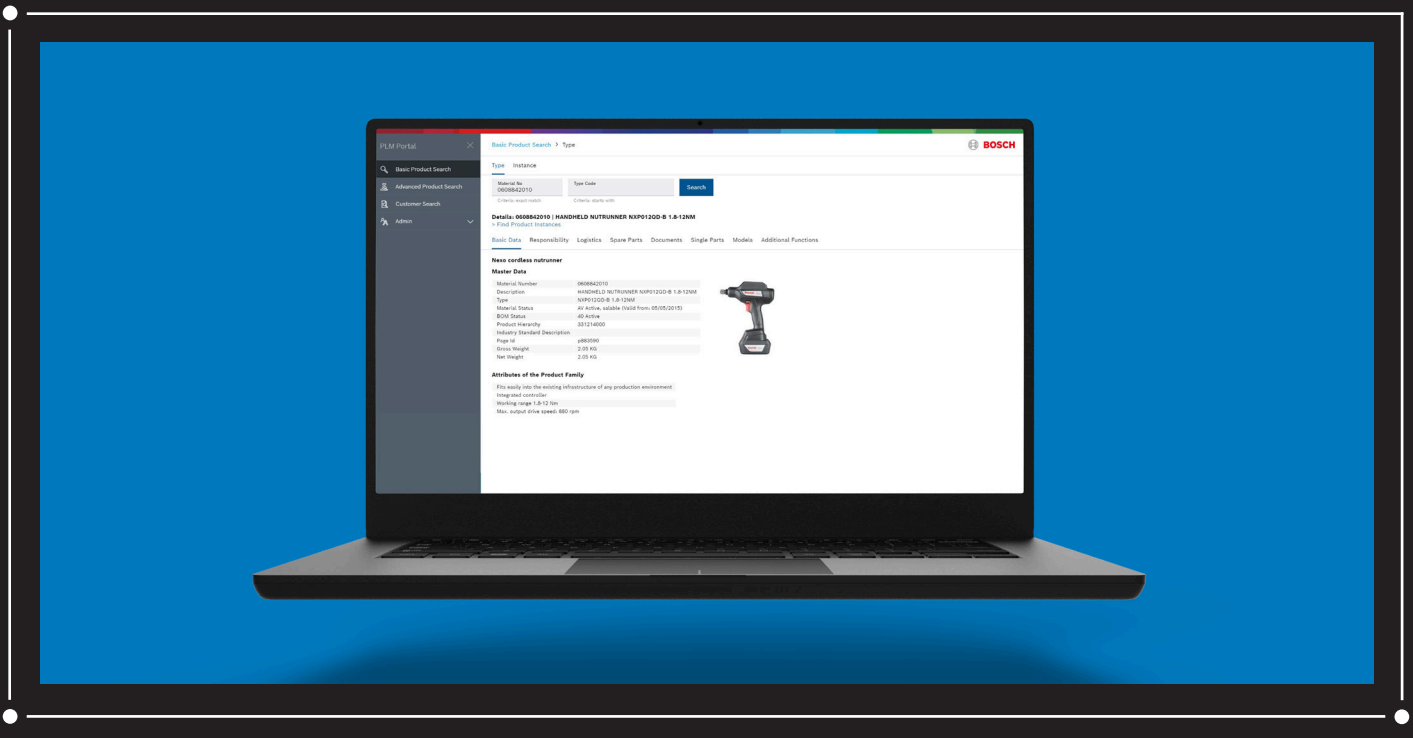


Figure 5: All product information at a glance with the PLM Portal

Product carbon footprint (PCF) calculation: simple creation of new use case

Calculating the product carbon footprint (PCF) poses major challenges for companies that are just embarking on the digitalization journey. Data collection along the entire product lifecycle requires the support of all relevant departments, combined with external data from individual component and material manufacturers, as well as database values. Compiling and consolidating this data is a time-consuming process that begins anew with every product. However, time is pressing: for example, PCF calculation will be mandatory for batteries from the automotive sector from 2027 as part of the European Union's Green Deal. In the long term, a mandatory digital product passport could be introduced for all sectors, which would require a consistent database from production to recycling – right down to component level.

This is where the semantic data homogenization effort in the PLM portal pays off: all required data is available, so that the need for an automated calculation of the CO₂ footprint can be met quickly. Such information is not only important for Bosch Rexroth, but above all for customers. They can conveniently access the PCF data of components installed in their systems via the respective digital twin. As already mentioned, Bosch Semantic Stack is not a purely in-house approach. Thanks to the option of an administration shell, it is open for use cases within a digital ecosystem. The PCF data can therefore be integrated into cumulative calculations for complex machines, making the overall calculation much easier.

Digital Service Assistant (DSA) app: implementing customer requirements quickly

Specific customer requirements were also the reason for a further use case. Having to use a service department for every telephone or electronic inquiry about a component is a time-consuming process. When it comes to globally distributed products, this is even more so when language or time barriers impede fast processing. Here too, Bosch Rexroth was able to create a concrete solution that relies on the database created using Bosch Semantic Stack: an application that is independent of the PLM portal and can be accessed directly on mobile customer devices.

The Digital Service Assistant (DSA) app provides customers with the support and data needed by means of a simple request to the digital product twin. This is a simple, standardized process for all new and existing products: customers scan their products with the app using a QR code or the type plate. All relevant information, including service and order data, is immediately available at a glance. Repair or maintenance requests can also be processed via the app, so that Bosch Rexroth service technicians know immediately which specific component is affected.

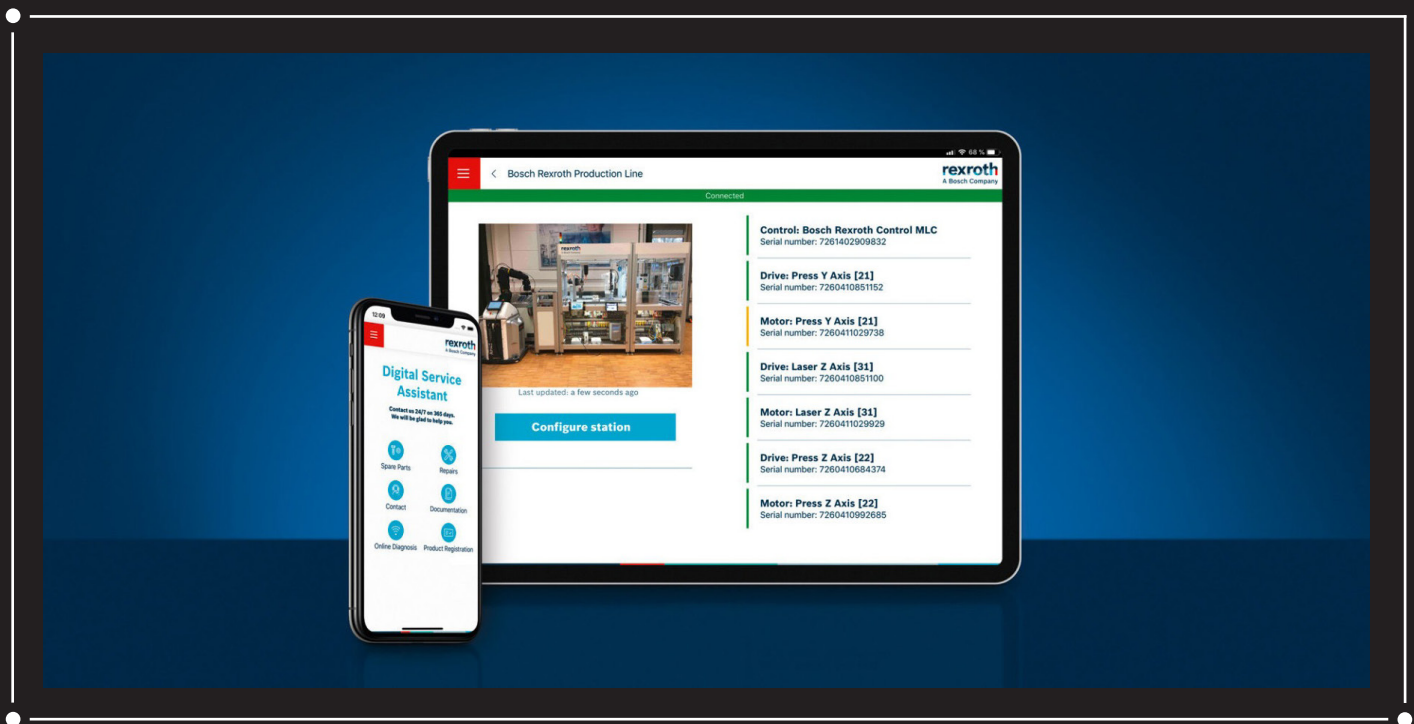


Figure 6: Reuse of the twin data in the Digital Service Assistant (DSA) app

7.

Digital twins as future enablers for Agentic AI: efficient data interaction without complexity

The evolution of data interaction in corporate environments has undergone a significant change. Complex SQL queries originally dominated the consolidation of data from heterogeneous database tables. This was followed by the introduction of graphical user interfaces (UIs), which enabled simplified data extraction via input screens. The next disruptive development stage is Artificial Intelligence (AI). This technology promises to partially autonomously solve complex tasks in proactive interaction with human users and to integrate seamlessly into corporate processes.

The role of semantics and digital twins as a data foundation for AI

The success of AI applications is inextricably linked to the quality and semantic preparation of the underlying data. Digital twins provide the ideal basis for this. Through their ability to map precise, context-rich and semantically prepared data, they create the conditions for the generation of reliable, traceable and high-quality results by AI systems. The more detailed and comprehensive the data is described in the digital twin, the more effectively AI models can make well-founded decisions.

A „Talk To Your Data AI Agent“ is an example of such an application that builds on this synergy. Instead of complicated query syntax or tedious navigation through disparate UIs, this agent enables users to query data in natural language and obtain precise answers almost in real time. Users communicate with the AI agent via a voice or chat interface. The AI translates the natural language query and breaks it down internally into many sub-queries. Instead of having to access a large number of diverging systems, all relevant information about a product - bundled in a digital twin - can be found centrally. This bundled information includes aspects such as master data, malfunctions or detailed maintenance histories, for example.

Reducing complexity through Agentic AI

The power of the „Talk To Your Data AI Agent“ manifests itself in the ability to process even highly complex requests efficiently. The following question could be asked of human colleagues: Can you find the batch number, the production time and the processed part ID for the product instance which is linked to the quality complaint notification number 230006798943? The product family is a battery; I do not know the exact name.“

This query is complex in nature and requires the correlation of multiple data points across different dimensions. However, since all of this relevant data for the specific product instance is stored in the digital twin, the AI can answer this question precisely. Based on architectures such as the Bosch Semantic Stack, the agent understands the intention of the users, identifies the respective context (in this case, quality complaint, batch, product), questions the structured digital twin data and presents actionable insights.

In the background, the AI agent acts in a highly automated manner: It creates sophisticated query pipelines for various systems, navigates intelligently through complex aspect models and finds the relevant data sources. It generates easy-to-understand responses almost in real time. This eliminates the need for associates to formulate nested database queries or to click through confusing lists, reports or tickets. Associates do not need to understand the internal data structure or the exact storage location of the information. They simply ask their question and receive a well-founded answer almost immediately.

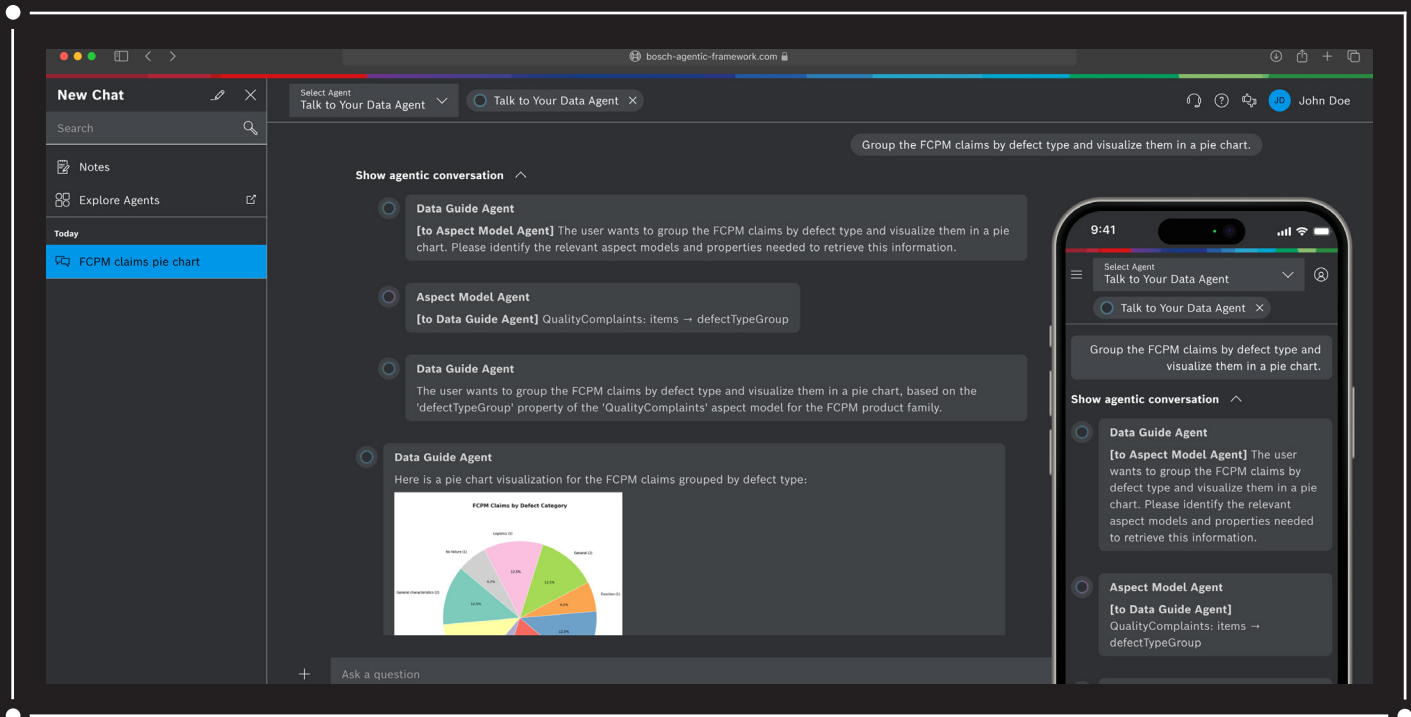


Figure 7: Interaction with the „Talk To Your Data AI Agent“

The strategic added value of Agentic AI

The significant added value of Agentic AI lies in its ability to fundamentally simplify human interaction with large, complex data volumes. It supports associates in mastering complex tasks by democratizing access to relevant information and abstracting the complexity of the data query. This leads to a considerable increase in operational efficiency, enables more informed decisions and makes a significant contribution to the agility of corporate processes. Semantics, digital twins and Agentic AI thus form a powerful trio for an intuitive data economy.

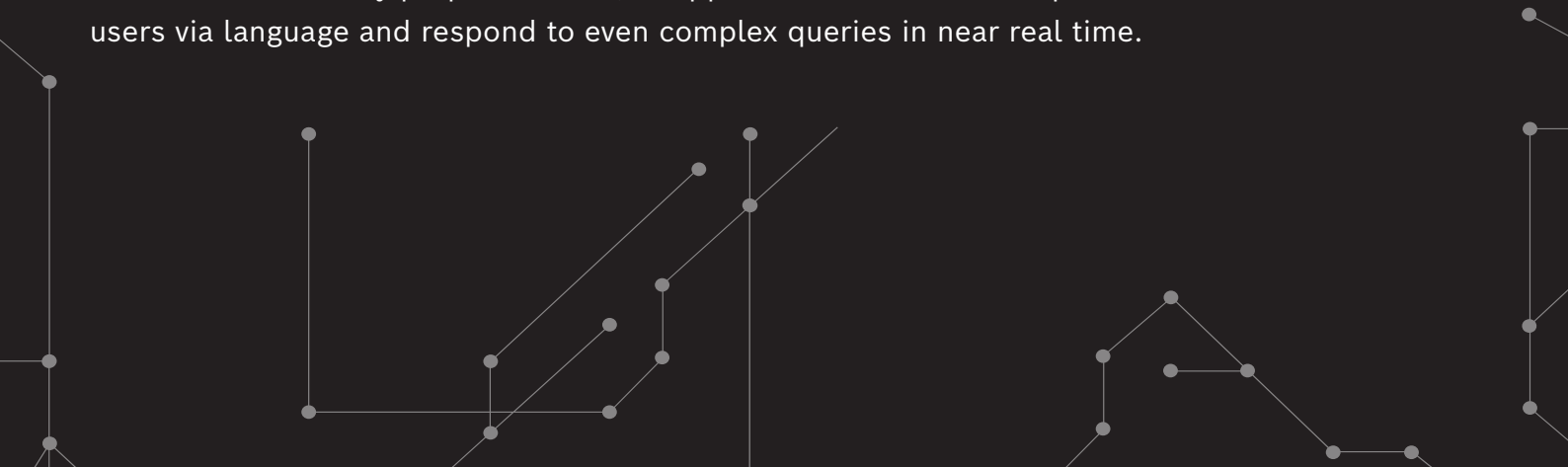
8.

Conclusion: the time is now

These practical examples show that a coherent data strategy not only allows solutions to be developed for specific problems, as in the case of the PLM portal. It also enables the fast and easy implementation of additional use cases, be it due to regulations such as the PCF calculation or customer requirements such as the DSA app. One case focuses on functionality within the existing database interface, while the other involves an independent app. Both use the same digital twins; only the APIs or aspects differ depending on the question. This scenario can be extended by an infinite number of use cases.

For manufacturing companies, this results in maximum flexibility and low development costs in the long term. The main part of traditional software programming, which is required to make data usable in the first place, can be omitted. This cost-efficient flexibility is badly needed: on the one hand, there is a high demand for environmental protection and safety, which is reflected in ever new regulations that often differ from region to region. On the other hand, customers more often require easy-to-implement data connections to components to create their own digital solutions or have increasingly greater information requirements.

A coherent data strategy is the best basis for AI applications. The quality of decisions made by AI models largely depends on how detailed and comprehensive the data in the digital twin is described and semantically prepared. Then, AI applications can be developed that interact with human users via language and respond to even complex queries in near real time.



Against the background of a constantly growing IT ecosystem, it is becoming increasingly important to homogenize data across company boundaries on the basis of a comprehensive data strategy and to agree on common standards. This would not be feasible for many companies without technology stacks such as Bosch Semantic Stack, which were developed with interoperability and the use of open source in mind from the outset. Ultimately, all players within the manufacturing industry are united by one challenge: they must remain competitive on a global scale with innovative products. And this is exactly why now is the best time to implement a coherent data strategy based on semantics and digital twins.



Make the first move!

The ideal data strategy with digital twins and semantics

Bosch Connected Industry is at your side with the Bosch Semantic Stack portfolio, comprehensive domain knowledge and pinpoint consulting. Together, we harness the valuable information from the various phases of your product life cycle – creating not only greater transparency along the value chain, but immediate added value.

We help you with holistic product optimization and ensure fast and scalable success with approach. You can also rely on us in the further course of data strategy and your data-driven projects: from existing blueprints to integration expertise to comprehensive training offerings, our experts are ready with in-depth knowledge of Bosch Semantic Stack and other topics.

www.bosch-semantic-stack.com

Robert Bosch Manufacturing Solutions GmbH

PO Box 30 02 20
70442 Stuttgart
Germany
+49 (711) 811 10 900

www.bosch-connected-industry.com
© Robert Bosch Manufacturing Solutions GmbH
June 2025

Bosch and the device mark are registered trademarks of Robert Bosch GmbH, Germany.

This document is a schematic representation and not an operating manual.