Marketplace

Data governance in product development

AI MARKETPLACE

Data governance in product development

Publisher:
Heinz Nixdorf Institute, Paderborn University, Germany

Authors:
Julian Knaup (inIT)
Robert Bakschik (inIT)
Alexander Dicks (inIT)
Gernot Boege (fiware)
Dennis Wendland (fiware)
Bernd Stöhr (HNI)

Contents

1 Introduction 4

2 State of the art 6

3 Data governance 8

3.1 Data handling 9

3.2 Data availability 10

3.3 Data security 11

3.4 Data quality 12

3.5 Data culture 13

4 Data governance in practice 14

5 Data governance in the ecosystem 15

5.1 Data ecosystems 15

5.2 AI-based vehicle diagnostics 16

6 Summary 20

Bibliography 21

Imprint 25

1 Introduction

Data is the oil of the 21st century. This analogy illustrates the immense potential for added value and the relevance of data for companies. A quick glance at the 100 largest publicly traded companies by market capitalisation reveals that four of the top five companies (Apple, Microsoft, Alphabet and Amazon) engage in data-driven business models. They achieve a market capitalisation of more than 8.6 trillion US dollars [PWC22]. Data is not consumed on usage as opposed to oil, which is. Thus data can be utilized by multiple parties, even repeatedly. Additionally, the global oil reserves are in steady decline whereas the amount of data generated is growing rapidly as shown by the tenfold increase in annual data generation from 2012 to 2020 [SSD22].

In addition to being intrinsically valuable data also drives innovation for the manufacturing industry and therefore plays a principal role in product creation [DÖK+21]. Its significance within the product development cycles (see Figure 1) is increasing, especially due to digitalisation.

Nowadays the area of product creation encounters faces an increasing number of customers with demand for individualised solutions and rapidly evolving product requirements. This creates a dynamic business environment in which companies encounter an ever mounting number of uncertainties [Fre22]. Acceleration of corresponding product development cycles becomes essential to keeping up with shorter product life cycles. For this purpose, a consistent use of available data can generate diverse added values. An analysis based on artificial intelligence (AI) methods of e.g. customer reviews can support strategic product planning and service development by automating and identifying potential at an early stage. The reuse of existing product data by means of common parts management can simplify product development and manufacturing improvements. However, a prerequisite for the automated or AI-supported use of this data is the availability of said data in a particular quality [BMK20], for which rules and responsibilities are needed. In terms

Data offers enormous potential for value creation that must be managed in order to be exploited.

Data governance is a collection of policies, standards, and responsibilities for using data profitably.

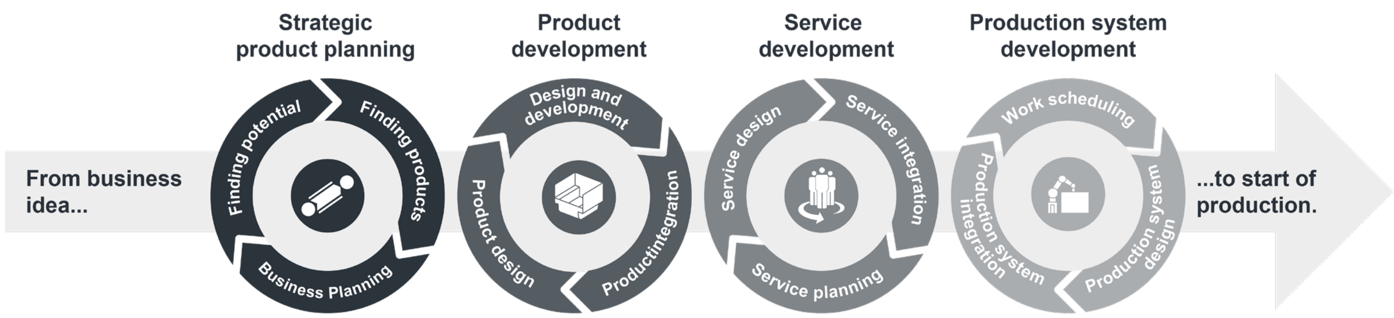

Figure 1: Product development cycle [DÖK+21]

of data the collection of guidelines, standards and responsibilities is called data governance.

This white paper aims to familiarise the reader with the most important concepts of data governance. In chap. 2 an overview of the state of the art is given, followed by a discussion of the most important concepts of data governance in chap. 3. In chap. 4 a method for determining the level of data governance maturity is presented. Finally, chap. 5 the focus on the generation of added value through cross-company data exchange is presented on the basis of a use case, before a summary is given.

2 State of the art

The growing importance of data governance is reflected in the literature [Dim]. Publications of various kinds are being produced time and again, highlighting different aspects of data governance. Some works deal with case studies to illustrate the importance of data governance [WRMR22], while others aim to evaluate and summarise the current literature [Al19]. The focus of many papers [Fre22,Mah21, Mah21+,Mah21*,EHS17] is primarily on a basic description of data governance.

The majority of examples in the academic and popular literature are either generic or focused on large companies [Al19,TRI20]. Occasionally, data governance implementations in small and medium-sized enterprises have also been studied [BC12]. Even less frequently, data governance is examined in the context of product development. Publications can be found discussing AI methods in product development, but in general data governance is considered a side note [WLL21].

Comparing the definitions of data governance in publications, it quickly becomes apparent that a uniform definition is still to be achieved. The same conclusion is reached by [Al19] and [ABH18]. In [OJS+16] data governance is defined as an „organisational capability that aims to manage data as an asset, regulating rights and responsibilities and providing methods and tools for this purpose“, while in [EHS17] data governance is understood as the „exercise of authority and control (planning, monitoring and enforcement) over the management of data assets“. The definition of data governance is often adapted to the context – a more economic view in [OJS+16] and a more abstract or general view in [EHS17].

There is no clear definition of data governance so far. While the general tenor is the same, details vary according to the focus of the publication.

The term data governance is separated to varying degrees from two related terms in different publications. In [ABH18] data governance is separated from IT governance - the structures and processes that govern the use of information technology to meet business objectives - despite overlaps in content. In [EHS17] and [Mah21], distinctions of varying degrees are made between data governance and data management. According to [EHS17], data governance is a fundamental subcategory of data management, which promotes the interaction of all other subcategories (see Figure 2). In contrast, [Mah21] sees the two concepts as equal. According to [Mah21], data governance regulates the „what“ - i.e. the establishment of guidelines, processes and strategies - while data management deals with the „how“ - i.e. the actual implementation in everyday business.

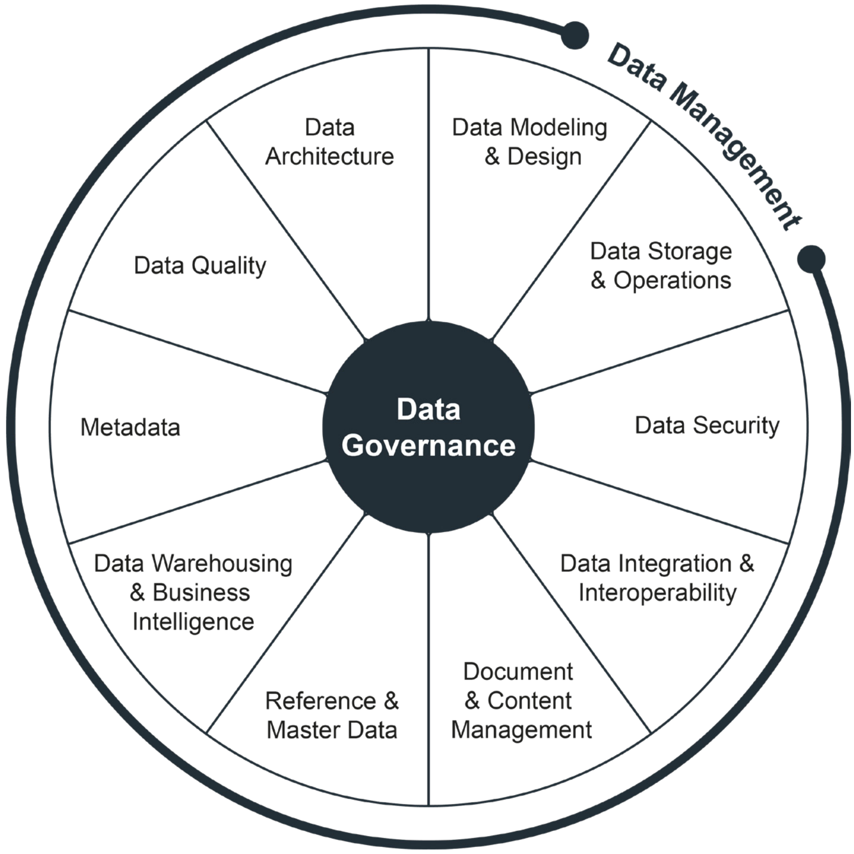

Figure 2 : Subcategories of data management according to [EHS17] with data governance as the core element

This work explicitly does not aim to provide a definition of data governance itself. Rather, the reader should be given the opportunity to get a grasp on an introductory model of data governance and to use this knowledge to lay the foundation for a functioning data governance concept in their own company.

3 Data governance

For companies data governance is not a mere end in itself, it helps to maximise the company's success. The financial oversight of neglecting data governance becomes evident through the realisation that data governance aims to minimise risk and optimise processes [EHS17]. Here risk minimisation is to be understood as compliance with legal requirements (e.g. protection of personal data), safeguarding proprietary knowledge and data as well as securing production lines from exploitation of external vulnerabilities. Process optimisation is a broad concept which applies to the entire company. Through increased data quality and availability, decisions can be made based on data, processes can be accelerated and the use of machine learning can generate insights from large amounts of data that would otherwise have remained hidden.

In the following, an entry-level model for data governance is presented to make companies aware of risk minimisation and process optimisation. Figure 1 shows the basic structure of this model. While the outer ring is described in detail in the following subchapters, the inner ring represents the foundation. According to [Mah21+], the core elements of the inner ring can be defined as follows:

People means all persons working with data. This includes the daily users and processors of the data, as well as all higher-level planning and decision-making bodies. If you share your data with partners outside the company, they must be considered as part of this structure for every change you want to implement.

Processes are the procedures for handling data. This includes the goals the introduction of data governance should achieve, as well as rules, guidelines and standards according to which the data should be handled.

Technology is not only the physical IT infrastructure for data processing, but also the software component. In general, this means any tool that facilitates the handling of data.

People, processes and technology - all three must build on each other and be available. For each category in the outer ring, it should be remembered that adjustments must be made in all three foundational categories to establish a robust data governance program.

The goal of data governance is to minimize risks and optimize processes when handling data.

Data governance deals with three core elements: People, Processes and Technology.

Figure 3: Data governance with the core elements of technology, people and processes.

3.1 Data handling

Administration consists of administrators and administrated. In the context of data governance, we speak of data stewards and data users, among others. However, instead of listing and describing these roles, an overview of possible tasks needed to be covered is given here. These tasks serve to ensure a smooth process of gaining insights from data.

Initially, instances are needed that establish usage rules based on usage requirements. Usage requirements should be established company-wide and, where possible, should also be standardised across different departments. „Standardised“ here refers to the readability of data to avoid situations where, for example, a different file type makes the data from department A practically unreadable for department B. This control instance also indirectly includes IT staff who are responsible for setting up and maintaining the infrastructure. Furthermore, the administration of the data repositories falls under the umbrella of the control instance, as it is accountable to the company management for all data-related processes.

After this control instance is established, the task of data collection follows. In compliance with the previously established rules, data must be collected from various sources. Here it does not matter whether it is surveys, customer messages or sensor data. Responsibility and accountability must be established for the collection of every type of data.

The acquisition is followed by the processing of the data. Here, an access hierarchy should be established. Since most users themselves have little to do with the processing or preparation of the data, their access should be restricted, but not completely prohibited. Every change to the data must be made traceable – the keyword here is meta data, i.e. data about data.

Data management regulates the control, collection and management of data.

Finally, the actual users or consumers of the, now processed, data should be named in this list. The users/consumers often mix with those responsible for processing/preparing the data. As before the following advice applies: an access hierarchy should be created that govern who is allowed to make changes and how these changes are documented.

These basic tasks may overlap depending on the size of the company. It is even considered desirable that the supervisory authority has close contact with representatives from data collection, processing and use or that representatives from these very areas are part of the supervisory board. Against the background of the legal basis at the time of writing this paper, the appointment of a controller for data protection must be mentioned here. Due to the importance of data protection, the involvement of said controller in the supervisory body is sensible, but not mandatory.

Therefore, when changes are made while handling data during its life cycle, they should be documented in such a way that it is clear who made the change and in what role. The next chapter discusses how the data can be made accessible to ensure the efficient use of the data by the employees of a company in the context of their tasks described in this chapter.

3.2 Data availability

Data must be readily available for its users in order to allow the use of AI applications and well-informed business decisions. The key to efficient data use in the age of Big Data is transparency. Whether heterogeneous data from the data lake or pre-processed data from the data warehouse, metadata management in the form of a data catalogue provides an overview of the data inventory and enables exploration and targeted filtered searches based on meta information. In this way, the required data can be retrieved as needed, availability is increased and the data lake does not devolve into a data swamp.

Especially for non-public data, the access rights of different user groups must be defined. In particular, confidentiality and possible contractual agreements must be taken into account. Besides the necessity of permissions to read, alter, create or delete data there must be a system in place to authenticate the user to prevent unauthorised access.

However, data exchange and access is not limited to internal use; data can be made accessible across company boundaries. In this way, companies can monetise their data or generate further added value through cooperative usage models. The AI marketplace as an ecosystem for AI in product development offers companies a platform to make their data available to AI experts. Through intelligent match making, companies with data challenges and AI solution providers are brought together. In general, companies should consider licensing and usage agreements when sharing their data beyond their own organisation. Maintaining data sovereignty can be ensured through solutions such as Data Spaces, which is discussed in more detail in Chap. 5.

Furthermore, the question arises as to the form in which data is made accessible. For companies one major reason for hesitancy is the possible disclosure of sensitive data [GTGG21]. This can be personal data of customers, but also production data that unintentionally provides information about the manufacturing process of a product. To counteract this, data records can be anonymised or at least pseudo-anonymised . In [GTGG21], the authors present a method for an anonymously learning AI which does not allow to draw clear conclusions one the original learning data based on the models generated.

Data availability describes the access infrastructure and rights inside and outside the company.

3.3 Data security

With the EU-wide adoption of the General Data Protection Regulation (GDPR) in mid-2018, companies that process data of EU citizens have the obligation to deal with the issue of data security. In the event of non-compliance, penalties of up to 20 million euros or 4% of global turnover, whichever is higher, are imposed in accordance with Article 83 (5) GDPR. Other vulnerabilities in data security may also cost companies dearly. In 2017, e.g., a cyberattack with the ransomware NotPetya encrypted company data of Mærsk, the largest container shipping company in the world, and rendered it unusable [Spi22]. The damage amounted to an estimated amount of 200 to 300 million euros and could only be limited by a lucky coincidence. Similar existential threats can occur when product development data is stolen or made unusable for one‘s own company. Data governance with a coherent and adaptable data security concept can minimise the risk of such an incident.

The establishment of rules and procedures, as well as the application of mechanisms and tools aimed at defining, establishing, controlling and improving information security is referred to as an Information Security Management System (ISMS) [ISO18]. The ISO/IEC 27000 series of standards is a collection of standards and recommendations for action on information security and supports the establishment of an ISMS. Companies can also have their ISMS voluntarily certified according to the ISO/IEC 27001 standard in order to strengthen the trust of business partners and thus secure an advantage in competition. Although certification does not automatically result in compliance with the GDPR, it does facilitate implementation. Further helpful guidelines can be taken from the IT-Grundschutz of the Federal Office for Information Security (BSI). The IT-Grundschutz Compendium and the IT-Grundschutz Profiles are particularly suitable for setting up an ISMS in small and medium-sized enterprises (SMEs) [BSI20].

In the attack on Mærsk in June 2017 with NotPetya, the EternalBlue exploit in Windows computers was utilised [BSI18], while a security patch from Microsoft was already available at that time. To reduce vulnerabilities, therefore it is advisable to keep software, operating systems and their patches up to date. However, since the exposure of a single computer cannot always be prevented, the BSI recommends further technical options, such as network segmentation and the restriction of local administrator accounts, to avoid a major network outage [BSI18]. Development data in particular often represent the know-how of current and former employees and are thus assets that require special protection. Backing up this data regularly to have a backup in case of emergency can be essential for companies. Therefore, development data should also be protected from internal misuse through authentication procedures. In addition to technical solutions, employees play an important role in the various processes. Training can significantly improve security awareness, so that phishing e-mails, for

Data security is not only required for customer data by the GDPR, but is strictly advisable for all data.

It is essential to discuss the importance of IT security with all employees in the company and to provide appropriate instructions for action.

example, can be counteracted. Technical security solutions may provide support by warning against harmful attachments and URLs.

Due to advancing digitalisation and networking, the question is not if there will be a confrontation with malware and cyberattacks, but when. In addition to the legal obligations, there is consequently a second risk if the topic of data security is ignored. Companies are therefore well advised to address this topic proactively in order to prevent damage.

3.4 Data quality

In [EHS17], various approaches to assessing data quality are discussed. These approaches are largely based on the work of Richard Y. Wang from the year 1996 [WS96]. This original work identified 16 dimensions (properties) that significantly influence data quality. The dimensions mentioned in [WS96] were selected through a survey and mix objective and subjective evaluation factors. Over the years, many variations of this system have been developed, with the main goal of providing a simple, general and objective way to assess data quality. Here, a compact model is presented that was developed in [DI12].

First of all, attention must be paid to the completeness of the data. This is primarily about completeness in the sense of: Are there empty data points? Whether these non-empty data points have a valid entry or not is determined later (see validity).

After the check for completeness comes the check for uniqueness. Only uniquely identifiable data can be used profitably. In the worst case, ambiguity can cause additional costs. For example, if two of your customers are stored under the same identifier, they may be confused for one another. This confusion can lead to additional costs in delivery or even cause lasting damage to customer confidence. To avoid such cases - on a small or large scale - it is essential to check whether data can be clearly identified.

After the uniqueness of the data has been ensured, the check for timeliness follows. Timeliness means that data should be revised promptly when a change is made. In a warehouse, for example, when material is withdrawn, the changes in stock should be entered immediately so that restocking the warehouse in a timely manner is possible.

Next, the validity of the data is checked. This is about conformity with the usage requirements and rules addressed in chapter 3.1. For example, a mixture of date formats should be avoided. Cases in which data do not adhere to the valid ranges also indicate errors (for example, a 220-litre barrel that is declared in the database as containing 270 litres).

Closely interwoven with the validity of the data is the absence of errors. Make sure that your data reflects reality. Here it is useful to look back at the example of the warehouse. If an error - of any kind

Data quality is defined by a varying number of quality dimensions. This introductory model includes the quality dimensions: Completeness, Uniqueness, Timeliness, Validity, Error-free, and Consistency.

- creeps in here, it can lead to serious delays in the company's operations. The potential for error ranges from inconvenience due to missing office materials to production downtime due to missing parts.
The last quality dimension is the consistency of the data. If several data sets have been recorded for one object, they should be consistent with each other. As an example, we again look at databases relating to a warehouse. A data record appears problematic if, for example, a barrel that contains 170 litres according to the database only contains 90 litres after 50 litres have been removed.
These six basic quality attributes form the basis for good data quality management. As mentioned at the outset, there are other approaches to addressing data quality, but as a foundation for a beginning data governance programme, these ideas are enough to prevent decision paralysis and get the proverbial ball rolling.

3.5 Data culture

After the previous sections have laid out the foundations and guidelines for functioning data governance, this final section of the chapter gives the basis for a sustainable transition to the profitable use of data - the establishment of a data culture. While the term seems ambiguous at first, it is defined as the behaviour of all persons in the company when dealing with data. The aim is to convey to all employees that compliance with the guidelines makes it easier to deal with data, even if it seems tedious in the beginning.
The first step towards a functioning data culture is to abandon ways of thinking à la „it has always been done this way". Of course, caution is needed when changing direction within a company, but thanks to the many experiences of companies that have relied on data for years, solutions to many problems are already available [TRI20].
This shift in thinking must be applied to all company levels. In order to build a healthy data culture, it is necessary to shift from a hierarchical, experience-based to a data-based decision-making process. Every member of the company should be regularly trained in the handling of data and the corresponding infrastructure so that relevant experts can be involved in the decision-making process and the proper handling of data becomes second nature [Sou22].
Once these basics are in place, with the appropriate research, it is possible to delve as deeply as desired into the topic of data culture and discover further possible improvements. And that is exactly what good data culture is all about. Regular self-monitoring and self-improvement of procedures, technical infrastructure and company processes should become a habit.

Data culture describes the behavior of employees in dealing with data.

4 Data governance in practice

Data governance is an important building block for data preparation and thus for the use of AI methods. For a practicable implementation of the methods described in chapter 3 a gradual approach to data governance is necessary. First, the initial situation in the respective company must be determined through a status analysis. In order for companies to gain an initial overview of their current status in the area of data governance, the AI marketplace offers the Data Governance Check on its homepage[1] as an orientation aid. The Data Governance Check is a questionnaire with 15 multiple-choice questions covering the three core elements, technology, people and processes. After completing the check, the user automatically receives a test result. The test result is graded on a five-stage maturity model (see Figure 4). Depending on the maturity level achieved, adequate recommendations for action are sent to support the user in reaching the next higher level. Further services can be initiated by the AI marketplace on the basis of an individual consultation.

Establishing data governance is a step-by-step process in which the AI Marketplace is there to advise you.

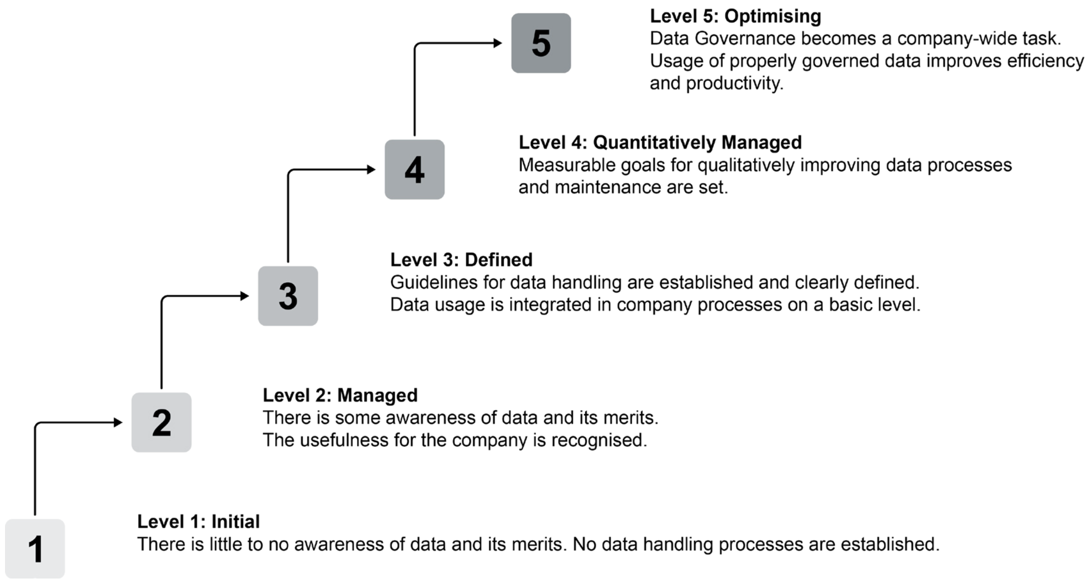

Figure 4: Data governance maturity model. Based on [SSL+15].

[1]https://ki-marktplatz.com/angebote/data-governance-check/

5 Data governance in the ecosystem

The previous chapters mainly focussed on internal company data. However, in order for data to develop its full potential, it must be made available across company and industry boundaries in ecosystems. These data ecosystems or data spaces promote the creation of value from data and thus enable new services, products, business models and innovations for companies [IDSA21].

Finally, data governance should not only be established within the company itself, but also in a data ecosystem that also includes business partners and customers.

5.1 Data ecosystems

Whether in business, politics, media or society - digital platforms have become an integral part of today‘s digitised world. These platforms enable access to products, services, digital content, information and data for everyone.

When it comes to bringing supply and demand together as efficiently as possible, digital platforms are clearly superior to traditional business models. They have evolved from technical tools to complex ecosystems. Digital platforms can create new markets and revitalise existing ones. In highly fragmented markets, they can bring together people or companies that would otherwise never have come into contact with each other.

In short, digital platforms, such as the AI Marketplace, hold enormous potential and countless opportunities. Digital platforms, applications in the fields of AI, IoT or Big Data, projects such as Gaia-X or AI4EU - all these application areas and endeavours are hardly conceivable without Data Spaces. [IDSA21]

From a data governance perspective, Data Spaces are to be seen as a data integration concept and are based on a distributed data architecture [OJS+16], i.e. data is physically held at the producer and central integration is omitted. Decentralisation serves the requirements of a distributed data infrastructure according to Gaia-X [BMW20], the European cloud infrastructure project. Data spaces in particular address the points of data sovereignty and data traceability [OJS+16] and ensure interoperability through semantic integration instead of schema matching (common database). Semantic integration here means the use of a uniform vocabulary, but not the semantic description of the data at domain level, because a data standard such as ECLASS for product master data does not yet exist in the area of product creation.

Data Spaces are the key to success in the future data economy and the prerequisite for the European Data Strategy. By being able to share data with unknown partners and data endpoints while maintaining data sovereignty, companies can help shape and leverage the true value of their data in a thriving ecosystem. [Nag21]

5.2 AI-based vehicle diagnostics

A pilot project within the framework of the joint project AI Marketplace is intended to exemplify the added value of data exchange across company boundaries. The use case of the industry partner Hella Gutmann Solutions GmbH focuses on AI-based vehicle diagnostics, which is presented below using 3 scenarios.
Scenario 1 represents the actual state of the vehicle diagnostics (see Figure 5). The end customer brings the vehicle to the garage for maintenance or in the event of a mafunction. The garage personnel then connect the vehicle to a Hella diagnostic device which reads the error codes presented by the control units. Using historical readout data, AI models were trained to carry out a diagnosis. Based on the recommendations of the diagnosis, the garage personnel undertake maintenance and repair work.

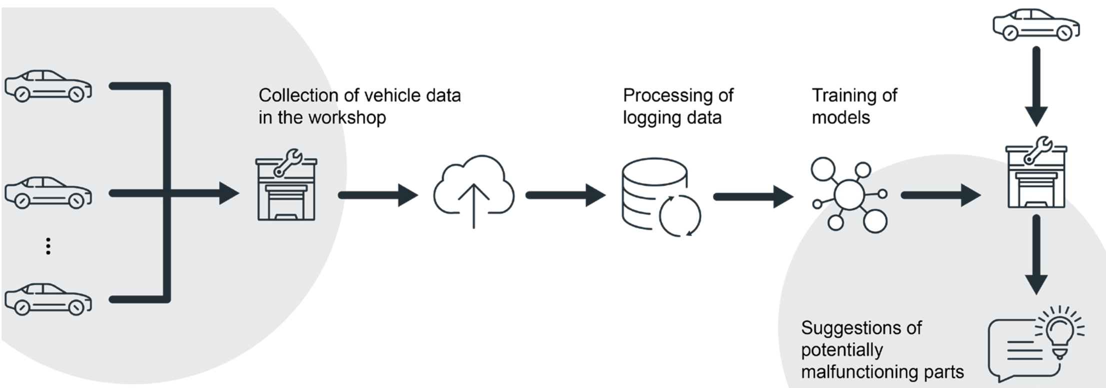

Figure 5: As-is state of AI-based vehicle diagnostics

This procedure is intransparent, characterised by manual activities and offers potential for improvement in the customer experience. In order to optimise the customer‘s workshop visit, the cross-company data exchange must be more interconnected so that the diagnosis is carried out predictively and workshop appointments are allocated according to need. To ensure an optimised process, as shown in Figure 5 can be achieved, two changes are required. Firstly, the vehicles must be enabled to send their data during normal use, and secondly, a data governance concept for data exchange is needed.

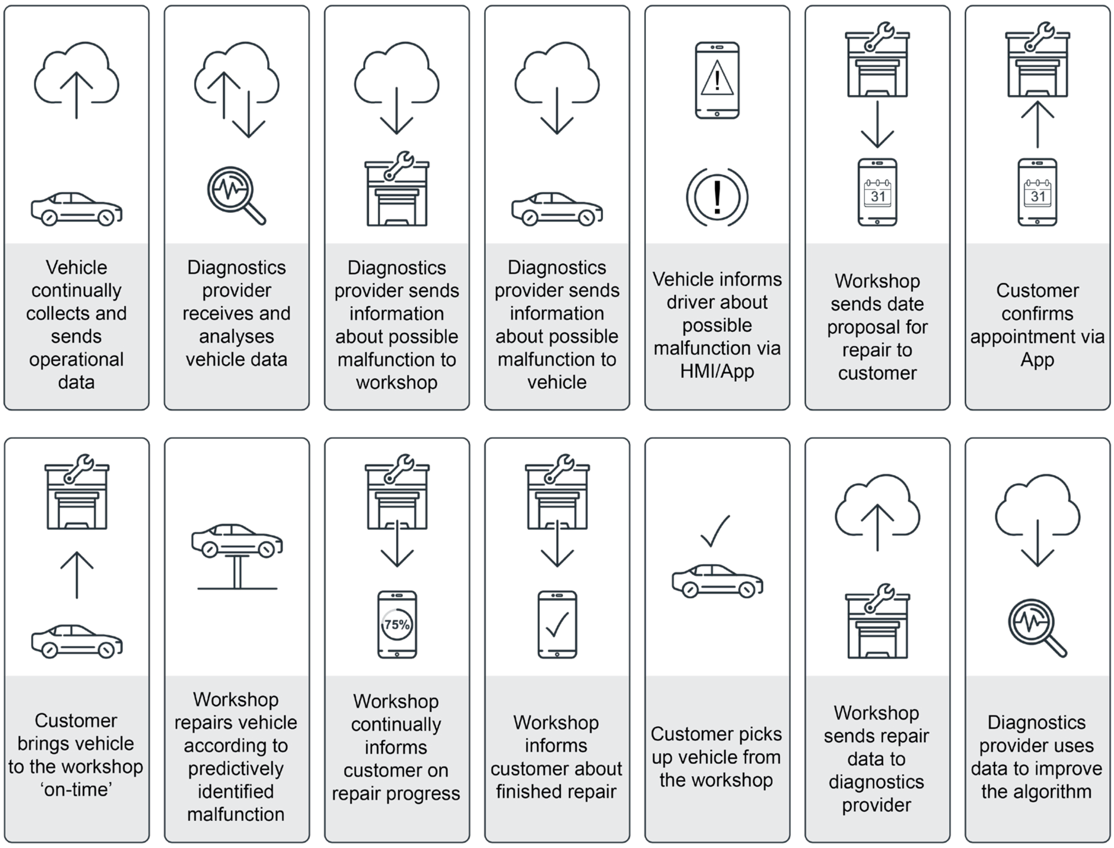

Figure 6: Process of optimised, AI-based vehicle diagnostics

Scenario 2 is an intermediate step that adds value by sharing data but does not improve the customer experience (see Figure 6). The vehicles send their operating data to the Hella diagnostic platform, which in turn creates aggregated diagnostic data. The diagnostic data is then offered via microservice (light grey arrows) or manually (red arrows) on the AI marketplace as a CSV file. As an ecosystem for product development data, the AI marketplace acts as an intermediary to bring together the data provider Hella with other players who can derive potential benefit from the data. The AI marketplace is the central platform on which the various players register to gain access to the data marketplace. This physical centring is also currently practised by data platforms such as the AWS (Amazon Web Services) Marketplace, among others.[2]

[2] https://aws.amazon.com/de/data-exchange/

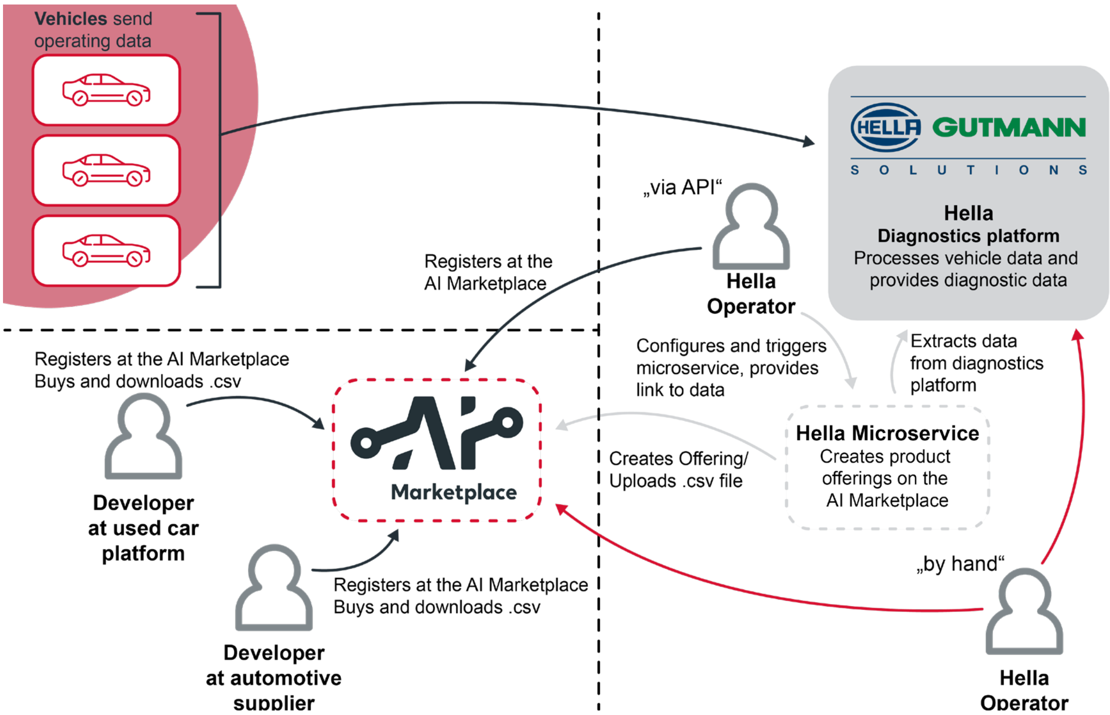

Figure 7: AI marketplace as a central data platform

In scenario 3, data spaces are used as a data integration concept. Here, the AI marketplace is also the central component, but it acts as an intermediary in that the data is linked (dynamically) and not held locally (statically). Since historically different initiatives have grown in the area of data spaces, there are also different, technologically incompatible implementations. These differentiate themselves, among other things, in access authorisation management. In the case of the AI Marketplace, both the i4Trust[4] and the reference implementation IDSA were focused on offering data in parallel in order to reach a maximum number of customers. Through the use of Data Spaces, other participants, such as the workshop, can subsequently be integrated in order to optimise the process flow according to Figure 5 can be achieved. More detailed information on Data Spaces can be found in the reference architecture model [OSTL19,IDS22] of the International Data Spaces Association[4] (IDSA).

[3] https://i4trust.org/
[4] https://internationaldataspaces.org/

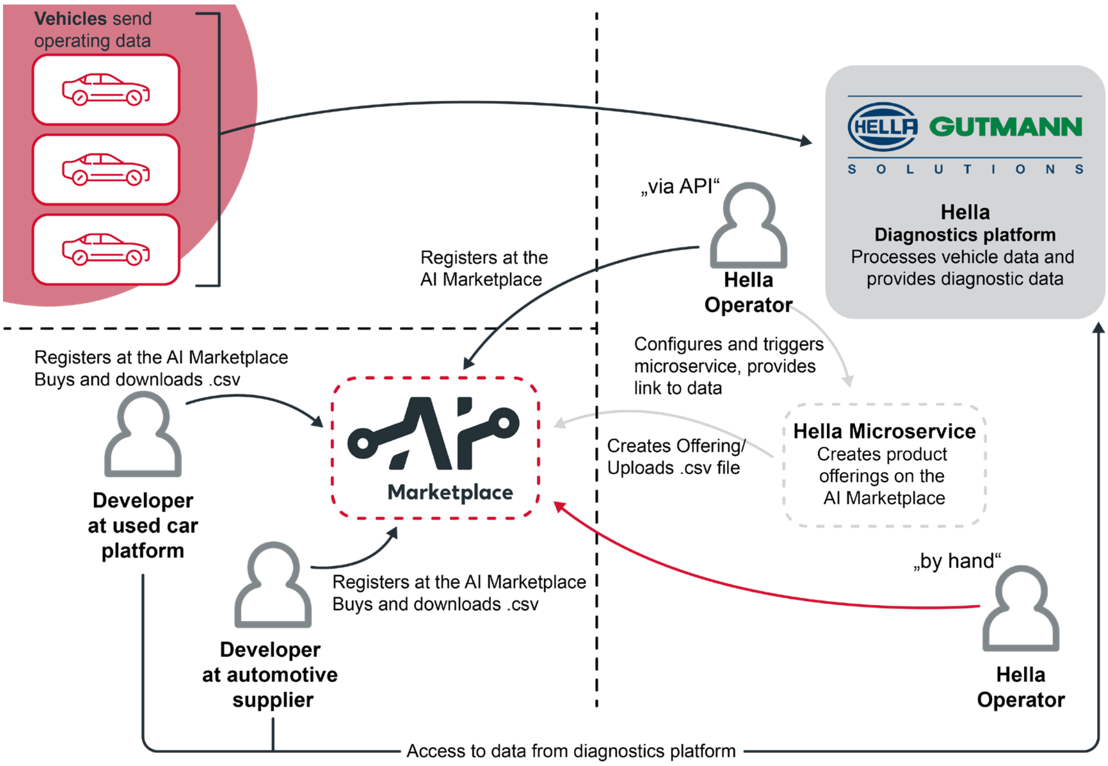

Figure 8: AI marketplace in the Data Space

Other data marketplaces that do not operate in the area of product creation already use Data Spaces. Advaneo[5] , for example, relies on IDS solutions for this purpose. It is thus becoming increasingly attractive for companies to share their data across their own company boundaries on data marketplaces, as they can both achieve a monetary advantage and ensure that data sovereignty is maintained. To ensure interoperability between individual Data Space technologies, the Data Spaces Support Centre has been founded by the European Commission. Here, the standardisation tasks for Data Space components are driven forward.

[5]https://www.advaneo.de/

6 Summary

A prerequisite for the use of AI algorithms to raise potential in product creation and in general is that sufficient data of a suitable quality is available. Data governance is therefore a central concept that supports the applicability of AI methods in companies. For a successful concept, points such as responsibilities, data security and data culture must be considered in addition to availability and data quality. Despite non-uniform definitions of the term, data governance, for example, can be summarised from an economic perspective as the „organisational capability that aims to manage data as an economic asset and regulates rights and responsibilities for this purpose and provides methods and tools" [OJS+16]. In this context, data management is not limited to internal company use, but can take place across company boundaries. With the use case of AI-based vehicle diagnostics, the AI marketplace as an ecosystem for AI in product development shows how added value can be generated for companies by sharing their data. In the process, successive expansion scenarios are shown that lead to implementation by means of Data Spaces. With Data Spaces, companies can simultaneously sell their data profitably without having to give up their data sovereignty.

Bibliography

[PWC22] PricewaterhouseCoopers International Limited, „Global Top 100 companies - by market capitalisation“, 2022. [Online]. Available: https://www.pwc.de/de/content/4af3e644-aee9-41bf-ad27-16427b52670f/pwc-global-top-100-companies-by-market-capitalisation-2022.pdf [Accessed date: 04.10.22]

[SSD22] Statista Search Department, „Volume of digital data generated/replicated annually worldwide in 2012 and 2020 and forecast for 2025“, Statista, 2021. [Online]. Available: https://de.statista.com/statistik/daten/studie/267974/umfrage/ forecast-worldwide-generated-data-volume/ [Accessed date: 04.10.22]

[DÖK+21] R. Dumitrescu, L. Özcan, P. Ködding, M. Foullois and R. Bernijazov, „Artificial intelligence in product creation,“ 2021 [Online]. Available: https://ki-marktplatz.com/wp-content/uploads/2021/10/Whitepaper-KI-Marktplatz_2021_10_13_FINAL.pdf [Access date: 04.10.22].

[Fre22] W. Frenz. Handbook Industry 4.0. Springer Nature, 2022.

[BMK20] P. Brous, M. Janssen and R. Krans, „Data Governance as Success Factor for Data Science“, In: Hattingh, M., Matthee, M., Smuts, H., Pappas, I., Dwivedi, Y., Mäntymäki, M. (eds) Responsible Design, Implementation and Use of Information and Communication Technology, Springer, 2020. https://doi.org/10.1007/978-3-030-44999-5_36

[Dim] Dimensions. Search term: „data governance“. [Online.] Available: https://app.dimensions.ai/discover/publication?search_mode=content &search_text=%22data%20 governance%22&search_type=kws&search_field=full_search [Accessed 26.10.22].

[WRMR22] D. Whittard, F. Ritchie, R. Musker and M. Rose, „Measuring the value of data governance in agricultural investments: A case study“, Experimental Agriculture, 58:e8, 2022.

[Al19] M. Al-Ruithe et.al, „A systematic literature review of data governance and cloud data governance“, Personal and Ubiquitous Computing, 23:839-859, 2019.

[Mah21] R. Mahanti, Data Governance and Compliance: Evolving to Our Current High Stakes Environment, Springer Nature, 2021.

[Mah21+]	R. Mahanti, Data Governance and Data Management: Contextualizing Data Governance Drivers, Technologies, and Tools, Springer Nature, 2021.
[Mah21*]	R. Mahanti, Data Governance Success: Growing and Sustaining Data Governance, Springer Nature, 2021.
[EHS17]	DAMA, S. Early, D. Henderson, L. Sebastian-Coleman (Eds.), The DAMA Guide to the Data Management Body of Knowledge (DAMA-DM BOK), 2nd Ed., Technics Publications, 2017.
[TRI20]	MIT Technology Review Insights, „The global AI agenda: promise, reality, and a future of data sharing“, MIT Technology Review, 26 March, 2020. [Online]. Available: https://www.technologyreview.com/2020/03/26/950287/the-global-ai-agenda-promise-reality-and-a-future-of-data-sharing/ [Accessed date: 04.10.22]
[BC12]	C. Begg and T. Caira, „Exploring the SME Quandary: Data Governance in Practise in the Small to Medium-Sized Enterprise Sector“, The Electronic Journal Information Systems Evaluation, 15(1):03-13, 2012.
[WLL21]	L. Wang, Z. Liu, A. Liu et al, „Artificial intelligence in product lifecycle management“, The International Journal of Advanced Manufacturing Technology, 114:771-796, 2021.
[ABH18]	M. Al-Ruithe, E. Benkhelifa and K. Hameed, „Data governance taxonomy: cloud versus non-cloud“, Sustainability, 10(1):95, 2018.
[OJS+16]	B. Otto, J. Jürjens, J. Schon, S. Auer, N. Menz and S. Wenzel, „Industrial Data Space“, Fraunhofer-Gesellschaft, 2016 [Online]. Available: https://www.fraunhofer.de/content/dam/zv/de/Forschungsfelder/ industrial-data-space/Industrial-Data-Space_whitepaper.pdf [Accessed: 04.10.22].
[GTGG21]	A. Girka, V. Terziyan, M. Gavriushenko, and A. Gontarenko, „Anonymization as homeomorphic data space transformation for privacy-preserving deep learning,“ Procedia Computer Science, 180:867-876, 2021. Available: https://doi.org/10.1016/j.procs.2021.01.337
[Spi22]	„Hackers cripple companies worldwide“, Spiegel,26 June, 2017. [Online.] Available: https://www.spiegel.de/netzwelt/web/maersk-hacker-legen-computer-von-groesster-reederei-der-welt-lahm-a-1154696.html [Accessed date: 04.10.22]

[ISO18] International Standard ISO/IEC 27000, 5th Ed., 2018. Available: https://standards.iso.org/ittf/PubliclyAvailableStandards/c073906_ISO_IEC_27000_2018_E.zip [Access date: 04.10.22].

[BSI20] „Informationssicherheit mit System - Der IT-Grundschutz des BSI“, Federal Office for Information Security, 2020. [Online]. Available: https://www.bsi.bund.de/SharedDocs/Downloads/DE/BSI/Grundschutz/ sonstiges/Informationssicherheit_mit_System.pdf?__blob=publicationFile&v=3 [Access date: 04.10.22]

[BSI18] „The State of IT Security in Germany 2017“, Federal Office for Information Security, 2018 [Online]. Available: https://www.bsi.bund.de/SharedDocs/Downloads/DE/BSI/ Publikationen/Lageberichte/Lagebericht2017.pdf?__blob=publicationFile&v=1 [Access date: 04.10.22]

[WS96] R. Wang and D. Strong, „Beyond Accuracy: What Data Quality Means to Data Consumers“, Journal of Management Information Systems, 12(4):5-33, 1996.

[DI12] Dama International, The six primary dimensions for data quality assessment, 2012.

[Sou22] P. Southekal, „Data Culture: What It Is And How To Make It Work,“ Forbes, 27 June, 2022. [Online.] Available: https://www.forbes.com/sites/forbestechcouncil/2022/06/27/data-culture-what-it-is-and-how-to-make-it-work/ [Accessed date: 04.10.22]

[SSL+15] A. Sathi, N. Sathi, C. Loza, M. Thomas, R. Harken and T. Eunice, „Explore the advanced analytics platform, Part 8: The Information Governance Model“, IBM Corporation developerWorks, 2015.

[IDSA21] International Data Spaces Association, „International Data Spaces - Enabling Data Economy“, 2021. [Online] Available: https://internationaldataspaces.org/wp-content/uploads/dlm_uploads/IDSA-brochure-International-Data-Spaces-Enabling-Data-Economy.pdf [Accessed: 04.10.22]

[Nag21] L. Nagel, „Let‘s Look Back on 2021, When Data Spaces Became Reality“, 2021. [Online] Available: https://internationaldataspaces.org/lets-look-back-on-2021-when-data-spaces-became-reality/ [Accessed: 04.10.22]

[BMW20] Federal Ministry for Economic Affairs and Energy (BMWi), „GAIA-X Technical Architecture, 2020. [Online].
Available: https://www.data-infrastructure.eu/GAIAX/Redaktion/EN/Publications/gaia-x-technical-architecture.pdf?__blob=publicationFile&v=5

[OSTL19] B. Otto, S. Steinbuß, A. Teuscher and S. Lohmann, „International Data Spaces Association: Reference Architecture Model", 2019. [Online].
Available: https://internationaldataspaces.org/wp-content/uploads/IDS-Reference-Architecture-Model-3.0-2019.pdf
[Accessed 04.10.22].

[IDS22] International Data Spaces Association, IDS RAM 4.0, GitHub Repository, 2022. [Online.]
Available: https://github.com/International-Data-Spaces-Association/IDS-RAM_4_0 [Accessed: 04.10.22].

Imprint

Year of publication:
2023

Place of publication:
Paderborn

Publisher:
Heinz Nixdorf Institut, Universität Paderborn

Authors:
Julian Knaup (inIT)
Robert Bakschik (inIT)
Alexander Dicks (inIT)
Gernot Boege (fiware)
Dennis Wendland (fiware)
Bernd Stöhr (HNI)

Design:
Lena Heller

HEINZ NIXDORF INSTITUT
UNIVERSITÄT PADERBORN

Heinz Nixdorf Institut
Universität Paderborn
Fürstenallee 11
33102 Paderborn
Telefon +49 (0) 5251 | 60 62 67
Telefax +49 (0) 5251 | 60 62 68
www.hni.uni-paderborn.de

Supported by:

Federal Ministry
for Economic Affairs
and Climate Action

on the basis of a decision
by the German Bundestag