Universal Data Space for Vehicle Development: Managing and Orchestrating Workflows via Policy-driven Data Transfer in Global Enterprises

Nikhil Mitapalli, Nadine Riske, Christopher Woll, Sebastian Fink

GNS Systems GmbH, Braunschweig, Germany

1 Introduction

The car of the future thinks ahead - and, based on the analysis and processing of data, can do far more than humans alone behind the wheel. Radar, lidar sensors as well as cameras, already collect and evaluate large amounts of information in real time and detect hazards such as black ice, objects such as stationary cars or the ends of traffic jams. If desired, the automobile of tomorrow will even drive autonomously. But it will be some time before all offline vehicles have disappeared from the roads and most of all cars are navigating through traffic completely autonomously. A study by the Prognos research institute for the ADAC [1] shows that autonomous driving is not expected to become established until 2040. By the time the first driverless cars are on the roads in Germany, both the volume and variety of data will have exploded. According to estimates by the international market research and consulting firm IDC [2], the volume of data worldwide will grow to as much as 143 Zbytes by 2024. And an end to the flood of data is not in view. In its latest study, IDC forecasts a globally generated data volume of around 284 Zbytes by 2027 [3].

Companies in all industries are therefore faced with the major challenge of mastering this data from the rapidly advancing digitalization today, as well as in the coming years. For example, the independent management and technology consultancy BearingPoint states that the effective use of data will be the key to navigating volatile markets to ensure long-term success [4]. Companies improve their competitive position through more efficient and intelligent use of the data they collect, through more targeted collection of new data, and through the application of advanced intelligent evaluation methods, cf. the Luenendonk study "From Data Silos to Data Streams" [5] from 2022. The goal should therefore be to bring data from a wide variety of sources and processes together and process it efficiently so that it can be used effectively. Whereas data management used to be a process that took place primarily in the background, today it is more a matter of making data usable for BI and analytics, as is explained, for example, in the BARC article "Data Management Survey 23" [6]. The professional and technical relevance of data management is constantly increasing, and with it the relevance of data management in corporate strategy. Data has become a strategic asset that should be efficiently maintained as a business-relevant resource.

The timely and innovative use of information of all kinds often requires company-wide collaboration, the elimination of data silos and the provision of new tools to create uniform access to the data. However, this assumes a clear data strategy and comprehensive data governance, states BearingPoint in its study [4]. With the change to a data-driven corporate culture, the demands on data management are therefore increasing: The focus here is, for example, on transparency through clear data classifications, traceable data flows and stringent reporting concepts. It is also necessary to ensure high data quality through appropriate control mechanisms and clearly regulated responsibilities and processes in governance, summarizes Luenendonk in professional cooperation with KPMG [5]. As a comprehensive and flexible solution for the management of data, the Global Autonomous Data Space (GADS) starts exactly there, because with GADS seamless data orchestration in the company is made possible. The powerful and scalable platform is capable of handling large volumes of data and ensures seamless access to data across different locations and applications. GADS can thus significantly help to increase flexibility in data usage and manage the complexity of highly distributed data landscapes.

2 Current situation in the automotive industry

The automotive industry is currently undergoing an enormous transformation, from a traditional car manufacturer to a modern mobility company. Increasingly data-based product development is a logical consequence of this transformation. It requires a new mindset regarding data use and allows agility in the form of continuous releases (CI/CD), as can be read in the publication "The data-driven Enterprise

of 2025" [7] by MC Kinsey. For several years now, it can be observed that data has become the driver of major technological innovations. Under market pressure and ever shorter and faster development cycles, automotive companies are responding with highly networked development and production processes whose efficiency depends on the ability to use data worldwide. For example, the development of new driving functions in the context of autonomous driving is simply not possible without an extensive data set.

The automotive sector is thus becoming significantly more digital, data-driven and more complex. The introduction of XiL simulation and a stronger focus on systems engineering will further increase the associated complexity in the coming years. The efficient analysis and evaluation of large heterogeneous data sets, for example with the help of suitable AI methods, already makes it possible today to recognize patterns and regularities. Based on this, solutions can be developed for optimized processes and workflows in development and subsequent series production - before the start in product development of new vehicles until later in the ongoing operation of the vehicle on the road. The data-driven approach therefore provides engineering with the necessary basis to make efficient use of extensive data from a variety of sources. Data sets must shed their "Cinderella image" [8] and get out of the passive role of pure archiving. It is highly negligent to accumulate a treasure trove of data and not make use of it. McKinsey & Company, for example, notes in its study for the Consumer Electronics Show in Las Vegas 2021 [9] that the use of data from cars alone could generate up to an average of $310 in additional annual revenue per vehicle for mobility companies across the entire ecosystem (automakers, suppliers and service providers, through to insurance, infrastructure and tech companies). Despite the great economic potential, data from cars has not yet been used on a large scale. McKinsey sees the reasons primarily in the differentiation of the company's own offering and the resulting lack of customer interest, a lack of a powerful organizational unit in the companies, and a limited ecosystem for driving data to date.

Companies are faced with the challenge of leveraging their data treasures and generating added value with the help of the right data strategy. To ensure that the future analysis of data leads to valid results, a viable data strategy ideally includes the modernization of the existing data stock as well as the integration of new data sources. A unified data environment is therefore at the beginning of all activities. It not only solves the aforementioned challenges in dealing with the further increasing volumes of data. Rather, this data environment forms a solid basis for meeting these new requirements in engineering, ever shorter development cycles and volatile markets in a future-oriented manner. Development engineers already require unrestricted access to all data for their simulations, as well as barrier-free access to tools and integrated development and software environments. In addition, new developments in the storage and processing of large amounts of data, in particular cloud and edge computing, are considered catalysts for data-driven value creation and promote the emergence of data- and cloud-based ecosystems, see [10]. In the key trends of the future - electric mobility, networking, autonomous driving and the automation of processes and production - software solutions, data and platforms will be at the center of future value creation.

3 Global Automotive Data Space - Interconnected data environment and optimized workflow processes

With the integrated data platform in engineering, GNS Systems is pursuing a holistic approach to the management and exchange of data. The GADS breaks down data silos, isolated solutions and redundancies and offers consistent access to the growing number of data types and systems. The solution allows large volumes of data to be streamed, quickly processed, stored and analyzed for the development of new vehicle models. Through the platform, companies can combine data in the cloud with on-prem resources, process it, and prepare or use it for Big Data analytics. According to the policy-based provisioning, transfer and access rights management of globally uploaded data, an optimal data flow is ensured at any point in the development process. Whether in data garages, at different development or production sites, GADS ensures seamless access and sharing of data across applications. Using appropriate policies, GADS orchestrates data flow across various locations up to different companies. The responsibility for optimal and secure operation lies with the data governance team.

3.1 Allowing seamless data flow

In accordance with the Digital Thread, data can be tracked over the entire life cycle of a product. This is made possible by archiving safety-relevant data and applications for at least 10 years after End of Life

(EoL). The opportunity to efficiently use large amounts of data from associated applications and components is therefore a decisive advantage for optimizing the development process in the labyrinth of data from different sources and for further development for future vehicle generations. GADS significantly increases speed, agility and efficiency in product development.

This already makes the essential features obvious: GADS enables the collection, distribution and management of large, unstructured data volumes in real time. GADS always ensures data quality, preserves information integrity and complies with existing legal requirements. As a learning platform, GADS therefore contributes to the continuous optimization of storage and computation allocations across the product lifecycle, supply chain, and user-specific data usage, see Figure 1.

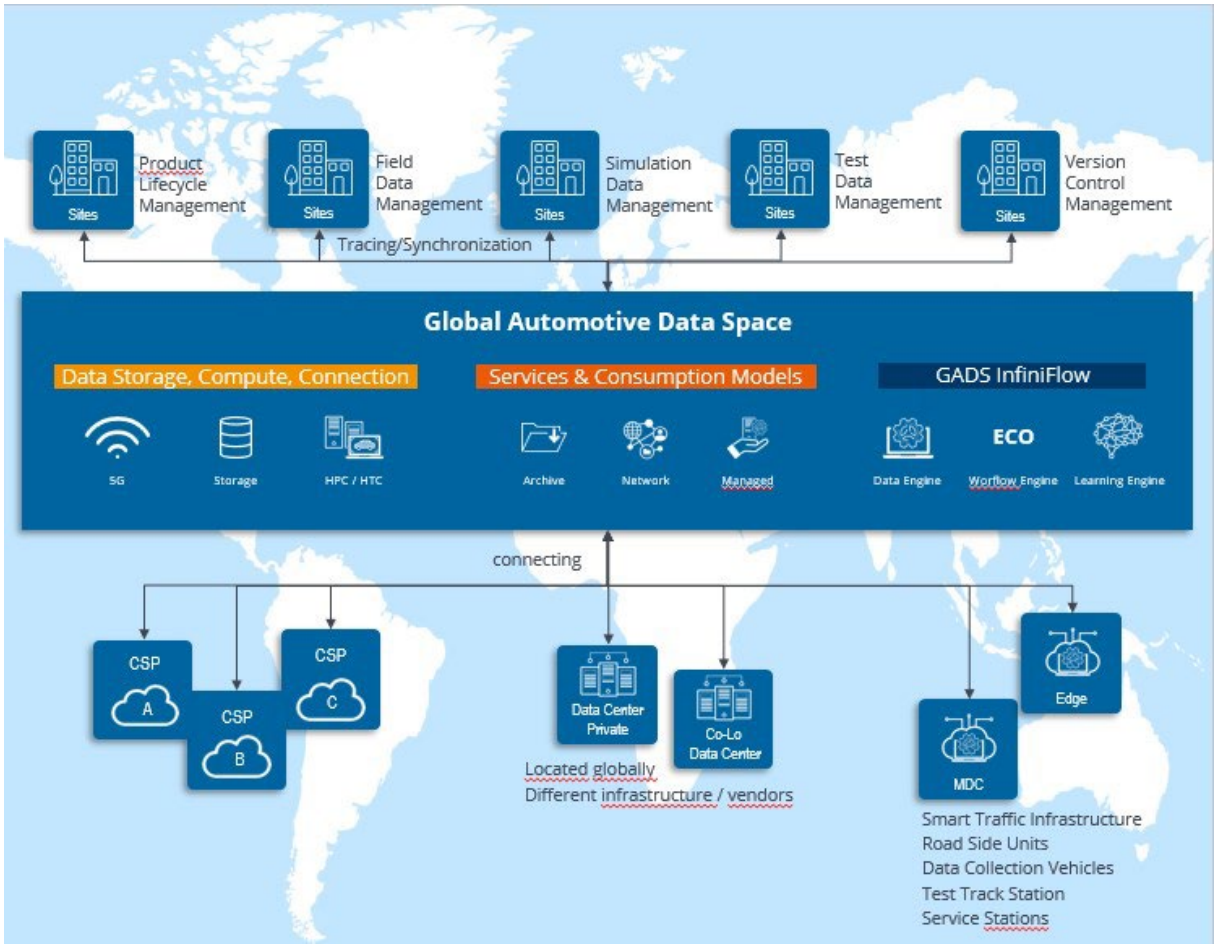

Fig.1: The motivation behind GADS: Traceability between areas and organizations must be ensured in accordance with IP and legal requirements. This requires use case and policy-driven orchestration of data and optimization of data flow.

3.2 Supporting workflows with consistent data

The data management, analysis and transfer of globally distributed data is a major benefit for the user. This is an essential factor for robust, efficient and effective data workflows. Taking the extensive Ansys product portfolio as an example, this can be illustrated as follows: GADS supports developers in using even complex data products/models for their development task with Ansys tools such as for example ANSYS Mechanical, LS-DYNA and Ansys Fluent. The connection of data, information and applications via the unified, consistent data platform adds significant value to the user. For example, the platform enables design engineers worldwide to access data with domain-specific, intelligent semantic data models. Data from different locations, such as from a structural analysis at headquarters in Germany, from a high-speed crash analysis in Japan, and from a low-speed crash analysis at a development center in Sweden, can be easily managed and orchestrated via GADS, as depicted in Figure 2.

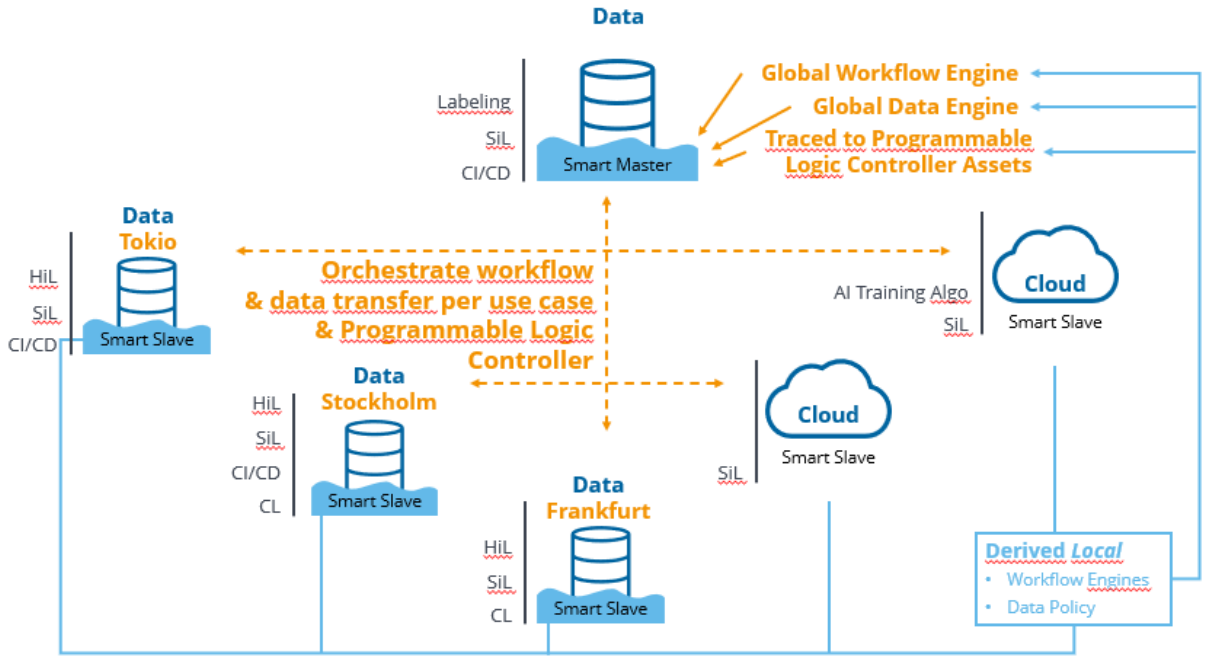

Fig.2: The process and workflow behind GADS: By connecting different systems (e.g., test benches) with a uniform, consistent data platform such as GADS, data transparency, data redundancy and data democratization as well as data management become easier. The platform enables automotive engineers worldwide to access data with domain-specific, intelligent semantic data models.

Besides the example given above by applying GADS in the CAE environment, the second application will be addressed to autonomous driving where GADS could also generate a large benefit. By the fact that there is a need to drive billions of kilometers to be able to make a statement if the autonomous driving car is as safe as the human driven one. A further example is given in [11], where it is listed that an autonomous driving car generates during an eight-hour driving period roughly 50-70 terabytes of data. It is obvious that this enormous data set needs not only huge storage capacities, but they also require meaningful preparation and evaluation. Here comes GADS in with the possibility to globally upload, store, search and process the data across edge, core and cloud. Once, when the data are uploaded and prepared by policies, various groups like development engineers, data scientists or researchers can perform further analytics and/or visualizations using the prepared data.

These options for using data in a variety of ways, which are shown here, contribute to an increase in efficiency in the development process. Centralized data management provides developers with the necessary basis for using complex data relationships to implement innovations across system boundaries. The cross-system, global search for data as well as the identification and consideration of data relationships significantly accelerates workflows in product development.

3.3 Organization-specific rules create maximum data transparency

By closing the gap between communication, orchestration and compliance, GADS connects the processes with its unified platform approach. For this purpose, GADS supports different storage types. Data can be uploaded, moved or downloaded from different sources or destinations. Especially for the development of autonomous vehicles, user-defined policies form a powerful tool that offers maximum flexibility and efficiency to the developer. Multiple applications can use the same custom policies. At the same time, an application can execute actions of multiple custom policies, as shown schematically in Figure 3.

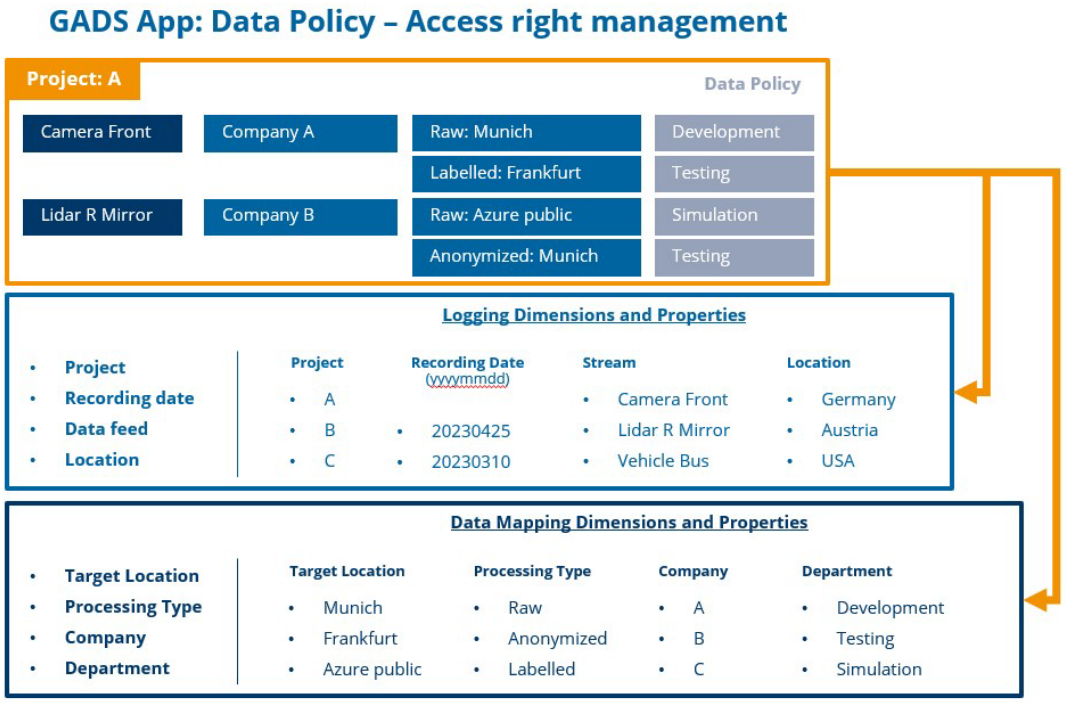

Fig.3: User-defined policies and GADS: The storage, transfer and management of access rights for globally uploaded data is the whole time legally compliant and meets the highest requirements in the automotive industry. With case-optimized access to global data, it is also easy to perform data integrity checks and obtain reliable results for autonomous driving simulation.

This works as follows: Already during the upload of the data they get their tags. These tags ensure that the data can be found again within the system, for example by means of filtering. For example, searching for a CAD file that is part of the chassis is easy using the keywords CAD and the corresponding chassis component. However, tags are also used to combine the data with the selected policy so that it can be executed in accordance with the Data Flow Definition. The Data Flow Definition thus contains all relevant metadata to enable the execution of many different tasks in vehicle development. The totality of the data flow definitions forms the data policy that defines the entire process flow.

In practice, it looks like this: For example, a policy contains a Data Flow definition with a "Move" action. Once the data is uploaded to a location, the selected Data Flow definition with the "Move" action performs the action according to the corresponding policy. The data is then moved to the previously specified destination. This creates maximum data transparency. The management of data flows becomes easier for the user, as the references within the various data are clearer. The use of a data product owner with a front end helps the user to monitor the data flow and control it according to the defined guidelines.

3.4 GADS supports the product quality

In practice in the automotive industry, vehicle development is based on a standardized architecture in a flexible manufacturing context so that the same basic parts can be used for millions of vehicles. Systems for traceability of individual components across all stages of the manufacturing process enable optimized supply chains and shorter lead times. This principle can also be transferred to GADS: The close networking of data via the platform enables greater transparency of data flows. Companies can thus better monitor and compare production lines. At the same time, they receive important information via the uniform database to find out which production steps take longer than expected and why. High data integrity is thus a crucial factor in accelerating and optimizing processes in product development. Automating processes in the virtual product development of modern vehicles also helps to reduce development times and costs. Maximum standardization - from version control and optimally scalable computing capacities to automated integration test systems - significantly ensures the quality of virtual vehicle development. Integrated AI-based analyses of the data sets additionally increase efficiency by identifying valuable information and patterns that can then be used profitably.

4 Conclusion

Today, companies already have a huge amount of data that needs to be used efficiently. In the future, there will be considerably more. Data-driven companies will dominate numerous markets or open new ones with their products and services. To be able to establish efficient data analysis within the company and beyond its boundaries, a clear strategy is required. This must include not only internal data sources, but also external information. Access to all data must be subject to clear standards. This not only ensures the most homogeneous access possible for all parties involved, but also optimal control of the data in terms of security, governance and compliance.

The GADS unified data environment provides developers with a coherent strategy that addresses data challenges (for example, data storage and management, optimal data access, traceability) in a future-oriented approach. The underlying concept of a single data space connects different CAE tools, ensures process traceability and increases confidence in the digital data treasure.

GADS captures the specified metadata of the processes and orchestrates them across different locations by following individually definable policies. Developers in the industry benefit from the enormous process security and high flexibility offered by a single system for managing data. GADS is an efficient environment for information and provides a unified interface for retrieving data. This streamlines the process of bringing together data from a physical system with applications that use such data. The result is a streamlined data-driven development process. GADS creates the foundation to connect to topics such as data analytics, smart data and predictive analytics, but also to the fields of artificial intelligence and machine learning.

Many systems, people and locations are already involved in the production of automobiles today. The combination and exchange of data is essential to produce the optimal automobile for the market. GADS supports the developers by providing them with easy access to the necessary data in often globally distributed teams. GADS thus makes a significant contribution to data democratization and generates added value for the data within its value chain.

5 Literature

[1] ADAC: Autonomes Fahren: So fahren wir in Zukunft, VÖ 31.08.2023 unter https://www.adac.de/rund-ums-fahrzeug/ausstattung-technik-zubehoer/autonomes-fahren/technik-vernetzung/aktuelle-technik/ (26.09.2023)

[2] IDC: Data Driven Intelligence, 2021 unter https://www.it-daily.net/speicherguide/datacenter-spg/idc-ungebremstes-datenwachstum-mangelhafte-datenqualitaet-25154 (31.08.2022)

[3] Statista: Volumen der jährlich generierten/replizierten digitalen Datenmenge weltweit von 2010 bis 2022 und Prognose bis 2027 unter https://de.statista.com/statistik/daten/studie/267974/umfrage/prognose-zum-weltweit-generierten-datenvolumen/; Quelle: IDC (01.09.2023)

[4] BearingPoint: Does your data underpin resilience – or undermine it?; Zürich, 11. August 2022

[5] Lünendonk/KPMG: Von Datensilos zu Datenströmen - Der Wandel von Banken und Versicherungen zu datengetriebenen Unternehmen, Mindelheim, 2022

[6] BARC: The Data Management Survey 23, 2022

[7] MC Kinsey: The data-driven enterprise of 2025, Januar 2022 unterhttps://www.mckinsey.com/business-functions/quantumblack/our-insights/the-data-driven-enterprise-of-2025 (31.08.2022)

[8] CIO/CSO/Computerwoche: Data-Driven Enterprise 2023, 2023

[9] McKinsey: Der ungenutzte Datenschatz: Durch Fahrdaten jährlich mehrere Hundert US-Dollar Wertpotential pro Auto möglich; Pressemitteilung zur Consumer Electronics Show in Las Vegas 2021 unter https://www.mckinsey.de/news/presse/2021-01-10-car-data (01.09.2023)

[10] e-mobil BW GmbH – Landesagentur für neue Mobilitätslösungen und Automotive Baden-Württemberg: Digitalisierung in der Mobilitätswirtschaft - Erfolgsfaktoren der Daten- und Plattformökonomie, Stuttgart, Februar 2023

[11] Frank Kraemer: AI and big data management for autonomous driving (AD). In: 20. Internationales Stuttgarter Symposium. Automobil und Motorentechnik, Springer Vieweg 2020