



A DIGITAL TWIN FOR ROOT CAUSE ANALYSIS AND PRODUCT QUALITY MONITORING

A. Detzner and M. Eigner

Abstract

Mass customization and increasing product complexity require new methods to ensure a continuously high product quality. In the case of product failures it has to be determined what distinguishes flawed products. The data generated by cybertronic products over their lifecycle offers new possibilities to find such distinctions. To manage this data for individual product instances the concept of a Digital Twin has been proposed. This paper introduces the elements of a Digital Twin for root cause analysis and product quality monitoring and suggests a data structure that enables data analytics.

Keywords: product data management (PDM), big data analysis, product lifecycle management (PLM), digital twin, quality management

1. Introduction

Manufacturing companies today face an increasing demand for customization, shorter time-to-market and intense global competition. They tackle these challenges with a rising number of product derivatives, diverse global supplier networks and local production as well as local procurement. The customization options of a premium automobile allow the staggering number of 10^{27} possible product variants (Zagel, 2006). In conjunction with the choices of the manufacturer such as different suppliers or production sites, the overall complexity of product development process is constantly increasing.

In the light of this development, ensuring a trouble-free functionality and high quality becomes a major challenge. Because even with advanced methods for simulation and virtual testing not all product variants can be validated. Hence despite great effort to identify potential problems early in product development, product failure in the field cannot be completely avoided.

In case of a product failure in the field, quality engineers aim to identify the root cause that lead to the failure. Given the complexity described above, the answer to this question is rarely obvious. Hence a detailed analysis is conducted to determine what product attributes differentiate products that fail from those that do not. These differences are crucial for understanding the technical root cause and subsequently improving current products and processes as well as future product generations. However given the vast number of product attributes this analysis can be very time consuming and heavily relies on the experience of quality engineers.

The Internet of Things and Services (IoT), cybertronic products (Eigner et al., 2015a) and smart factories will make even more product attributes available to the manufacturer. An internet connection enables machines, products and components to communicate and collect data along the entire product lifecycle. At the same time the storage of massive amounts of data has become logistically and economically feasible and algorithmic breakthroughs such as the MapReduce Algorithm (Dean and Ghemawat, 2008) have made analyzing this Big Data possible. Collecting, storing and analyzing data, from detailed

manufacturing information to individual user profiles, offers the opportunity for in-depth analysis of technical issues and a data-driven continuous improvement process.

But with the large amount of data that is going to be generated, enterprises need IT-architectures that ensure the availability of the right data, in the right quality, at the right location (Eigner et al., 2016b). In System Lifecycle Management the concept of a Digital Twin has been proposed to manage and store the heterogeneous data generated over the entire product lifecycle. However it has to be determined what the right data for root cause analysis and product quality monitoring is, by evaluating all available product attributes. Additionally the large number of product attributes makes a manual data analysis increasingly difficult. Therefore anomalous product attributes have to be determined with the aid of Big Data analytics.

In this paper we propose the Digital Twin as a framework for product quality monitoring and root cause analysis. Chapter two gives an overview of different frameworks proposed to manage data over the entire product lifecycle and introduces different interpretations of the Digital Twin concept. Chapter three provides a detailed collection of the elements required in a Digital Twin for product quality monitoring. In chapter four an analysis regarding the problems of data analytics on large amounts of data is conducted. Based on that analysis, a hierarchical data structure that enables Big Data analytics is proposed. The chapter concludes with pointing out the shortcomings of current data management processes and proposes a way to manage data requirements in the future.

2. State of the art

Eigner et al. (2015b) summarize that by 2020 investments of 500 billion Euros into cybertronic products are to be expected. Since these products will generate massive amounts of data, there is an imminent need for cross-enterprise data strategies and IT-architectures to utilize Big Data analytics along the Product Lifecycle. Li et al. (2015) conducted a meta-study to analyze the great potential of Big Data analytics in Product Lifecycle Management (PLM). They present several examples where Big Data analytics could be employed, such as providing accurate and personalized product services as well as improving the quality and output of production and design.

Different frameworks have been proposed to incorporate product feedback data into existing PLM architectures. Lee et al. (2013) point out, that merely collecting data is not sufficient, as context and meaning is crucial, for the data to be understood by the right personnel. Their research focuses on the manufacturing phase and proposes an architecture for prognostics and health management of manufacturing machinery. They later refine this idea, proposing a 5C architecture for cybertronic products where they identify connection, conversion, cyber, cognition and configure as the core steps of an architecture for IoT products. In their architecture, data analytics tasks such as prediction and similarity identification are performed at the cyber level. Previous levels ensure data availability and quality, while subsequent layers enable manufacturers to turn the data into action (Lee et al., 2015).

Abramovici and Lindner (2011) propose a framework for constant improvement of future product generations. Their framework also consists of different layers including an Operative Information Layer where raw information is stored, an aggregation layer where the harmonization of heterogeneous data takes place and diagnosis layer to utilize the information. Abramovici et al. (2016) later point out that architectural-, component- and usage-information are all necessary for managing cybertronic products. They state that the management of such information in a cross-enterprise network poses a major challenge. Furthermore they point out that continuous and transparent information flows are a prerequisite for efficient collaboration over the entire product lifecycle.

Similar layered architectures are proposed to utilize PLM data as well as product use information for Big Data analytics by Kassner et al. (2015) and Zhang et al. (2017) or to support root cause analysis by integrating maintenance information (Madenas, 2014). Dienst et al. (2014) conclude based on the work of Abramovici et al. (2011) and Uhr et al. (2012) that current PLM systems are not suitable to integrate feedback data due to the large amount of data and the difficulty of adapting existing PLM solutions.

In addition to these general frameworks, the concept of a Digital Twin has gained traction as a promising approach to manage product related data across the entire product lifecycle. The Digital Twin is commonly defined as a virtual mapping of a physical product. This mapping includes structure and

properties defined during product development as well as the current configuration of the product and usage information (Muggeo and Pfenning, 2015). While the classical digital model or digital master in a PLM system is not related to a specific order, a Digital Twin is always associated to a singular, physical product instance (Eigner et al., 2016a). Since historically PLM systems are not capable of displaying individual product instances this concept requires a new architecture for PLM data. A unique serial number serves as the link between a product instance master data and its lifecycle feedback data (Dienst et al., 2014).

The concept of a Digital Twin was first proposed by Grieves in a 2002 University of Michigan presentation as a “conceptual idea for PLM” which introduced the idea that each physical object has a digital counterpart and both are synchronized by a data flow between them (Grieves and Vickers, 2017). Starting out as a sparse CAD object, current interpretations of the Digital Twin concept often aim to design, test and simulate entirely based on virtual versions of a system (Grieves and Vickers, 2017) or focus on the aspect of continuous product improvement.

The Digital Twin as a conceptual model was quickly adopted by the aerospace and astronautics industry (Grieves and Vickers, 2017). The prominent reason being that parts are constantly replaced during the exceptionally long lifecycle of an aircraft of up to 50 years. The Digital Twin allows to break with the traditional differentiation between static and dynamic product information and represents a current, as-maintained state of an aircraft (Rios et al., 2015). An application example is given by Tuegel et al. (2011) where a Digital Twin containing product information and load data from sensors is employed for accurate fatigue prediction of individual aircrafts. Glaessgen and Stargel (2012) see future generations of NASA and U.S. Air Force vehicles utilizing the Digital Twin concept to integrate diverse information about the physical product into high fidelity simulations. They define the Digital Twin as an “integrated multi-physics, multiscale, and probabilistic simulation of a complex product [that] uses the best available physical models, sensor updates etc. to mirror the life of its corresponding twin”.

Tao et al. (2018) point out that up-to-date several definitions of Digital Twin have been proposed. Their interpretation focuses on the Digital Twin as the basis for simulation studies and assumes a permanent real time reflection between digital twin and physical counterpart. The scope of research at the Siemens cooperation is focused on using the Digital Twin for simulation applications as well. Boschert and Rosen (2016) point out the potential of a Digital Twin to manage the growing complexity of cybertronic products and to reduce the time-to-market for new products. Their vision of a Digital Twin includes all information that might be useful in later lifecycle phases, though they immediately point out that this approach is not feasible due to the immense volume of the diverse and unstructured data. Based on their expert knowledge they propose that the elements contained in a Digital Twin have to be chosen based on specific questions. Grieves and Vickers (2017) elaborate what data can be contained in a Digital Twin. Depending on the use case they state that a complete 3D model of the physical instance and its components, a Bill of Materials that lists current and replaced components and a Bill of Processes that lists the operations that were performed during manufacturing can be included. In addition the results of any measurements and tests on the instance, a service record that describes past services and all available sensor data from the product instance should be added to the Digital Twin.

3. Elements of a Digital Twin for product quality monitoring

Recent interpretations of the Digital Twin concept have moved away from the idea that all available data can be included. Instead a selection of data has to be performed depending on the use case. This chapter proposes the elements of a Digital Twin for the purpose of quality monitoring and root cause analysis in the case of product failure in the field.

The primary purpose of a Digital Twin in this use case is to consolidate product attributes in one place that might distinguish products that fail from those that do not. The necessary product attributes can be grouped into product structure, service data, usage phase data, manufacturing data and logistics data. According to the Digital Twin concept, a unique set of attributes derived from this data needs to be stored for each individual product instance.

3.1. Product structure

The product structure is the core element of the Digital Twin. Information about the individual parts of a product are crucial for two reasons. First product failures in the field can often be related to certain parts built into a product or are introduced with changes to a part. Second a large portion of other information, such as assembly data, depends on the individual product structure of a product instance.

In the automotive domain PLM systems manage a configurable Bill of Materials (BOM) that includes all parts for all possible choices of a customer. A tree-like *generic product structure* is defined as a common setup for all vehicles. Attached to the top nodes of this structure, position variants indicate which parts are available for a product instance. Figure 1 shows an excerpt of a fictional generic product structure, options and items as an example. Based on the options chosen in a specific customer order a manufacturing BOM is created that includes all parts for this product instance. The order information is archived to allow a retroactive creation of the *as-planned* BOM. In the automotive domain, for parts critical to customer safety, unique serial numbers are archived linked to the unique vehicle identification number (VIN). In addition the serial numbers of electronic control units (ECU) built into a product instance are also retained to ensure the compatibility of later software updates. These serial numbers are often not archived in the PLM systems but in supplemental databases.

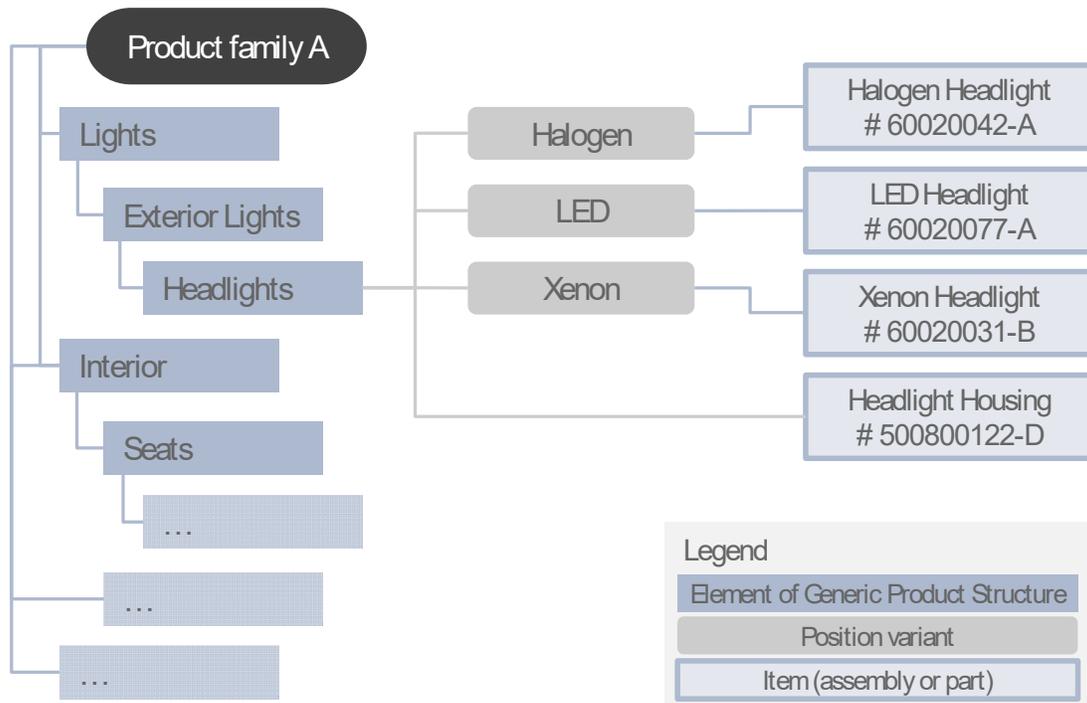


Figure 1. Generic product structure with options and items in the automotive industry (Eigner, 2017)

The data-driven analysis of product failures in the field is primarily based on the order information. To fully enable the desired data analysis insights, a Digital Twin must include a Bill of Materials (BOM) for each product instance. This BOM has to include the part numbers including revision index and if available the unique serial numbers of each part built into a product instance. Currently software updates are often traced by a change in the revision index of the ECU. To clearly differentiate between hardware and software changes the version numbers of software have to be stored independently from the physical part.

Despite the advantages of maintaining a complete *as-built* BOM in a Digital Twin, tracking every single part is logistically difficult and economically not feasible. Hence for each part of the generic product

structure it has to be determined if a traceability by unique serial number is necessary. In any case the part numbers given in the *as-planned* BOM should be archived for each product. If the number of the part built into a product instance cannot be determined with certainty, a fuzzy approach is recommended: All alternative parts are included in the Digital Twin and assigned a probability of occurring in that product instance. This approach enables Big Data analytics and is preferable to the alternative of not storing the part information at all.

3.2. Service data

While the classic representation of the BOM is static, a Digital Twin needs to evolve with the current state of a product. To track the physical configuration of a product, all part changes conducted during service and repair need to be recorded in the Digital Twin, thus creating an *as-maintained* BOM. Again it is economically not justifiable to track every single part. Hence the same requirements regarding traceability as for the *as-built* BOM apply. In addition to the change of physical parts, software updates need to be documented as well.

For product quality monitoring it is crucial that the Digital Twin representation is not limited to the current state of the product. Historic information about previous software versions or parts built into a product instance can be essential to determine the root cause of a problem. In fact a repair can provoke further problems. Therefore the Digital Twin needs to include detailed information about all actions performed during service and repair. During the warranty period this information is available in the accounting systems of original equipment manufacturers (OEMs). In conjunction with the dates of a repair, an analysis regarding previous repairs is possible.

Reports on the repair are often accompanied by a problem description given by the customer or the service personnel. These texts can be tremendously useful to determine the nature of a problem. But unstructured data requires special analytical methods to be used. Regardless text information needs to be part of the Digital Twin.

Despite the importance of service data, adding this information to the Digital Twin brings two major challenges. First the general classification of a problem as well as the individual tasks performed during a repair are manually registered in the accounting system by the workshop personnel. Hence the data is prone to human error. Nevertheless this classification is fundamental to distinguish between unrelated problems in order to reduce the complexity of the analysis. Even though some amount of mislabelled repairs cannot be prevented, incentives for a correct documentation can improve the situation.

The second challenge is that not all service is carried out at repair shops authorized by the OEM. In such cases no data about the repair is available. Since non-professional repairs can result in subsequent problems, the missing data is a significant problem. In addition undocumented repairs often lead to implausible sensor data. To counteract this problem, previously undocumented repairs should be retroactively added to the Digital Twin when discovered. For cybertronic products the service data can be enriched by sensor data. For instance an airbag activation indicates a crash even if no repair of a corresponding crash damage was recorded.

3.3. Usage phase data

Traditionally the reach of automotive OEMs ended at the point of sale of their product. The data infrequently collected during service and repair being the only exception. The ability to collect data during the usage phase of a product is one of the radical changes introduced by the Internet of Things. Real time data from a product instance enables completely new, service-oriented business models such as car sharing and predictive maintenance. As illustrated in the introduction, root cause analysis aims to determine the product attributes that distinguish products that fail. Naturally this analysis requires a reference group of products without failure to determine the normal distribution of product attributes. Prior to cybertronic products, data was only collected during repairs and therefore limited to faulty products. Hence for product quality monitoring a vital advantage is the availability of data from products without failure.

For product quality monitoring three kinds of information are valuable: Product condition information in the form of fault indicators, usage information and environmental information. Product quality

monitoring does not require a real time reflection of the physical product in the Digital Twin. Instead regular updates on sensor data and product condition are sufficient.

3.3.1. Product condition information

In the automotive domain the fault indicators are called Diagnostic Trouble Codes (DTC). Given certain sensor conditions or software states a DTC is saved in one of the vehicles ECUs. The conditions are defined during product development. Therefore DTCs are aggregated information based on expected issues and the know-how of product developers. Accompanied by a unique timestamp these fault indicators make up the core element of the usage phase data contained in a Digital Twin.

Limiting the product condition information to fault indicators is a concession to the volume of data in a Digital Twin. While high-frequency traces of all on-board operations would be useful for diagnosis, the resulting amount of data would rapidly exceed all other data in the Digital Twin, rendering it very difficult to collect and store this data for each product instance. Nevertheless on-board product architectures should be able to relay this data for individual product instances and limited time frames. In addition the Digital Twin must be able to evolve by adding new failure indicators. The reflection between Digital Twin and product instance means that new failure indicators need to be introduced to the physical products via remote software updates if authorized by the customer.

3.3.2. Usage information

Usage information describes data related to customer behavior. Since customer behavior has a dramatic impact on the wear and tear of a product, this information is essential for product quality monitoring. The central elements of usage information in the automotive domain are the mileage and the number of months in service. While this information is already useful, a Digital Twin would greatly benefit from more detailed data. Accumulated usage information about trip distances, speeds and idle times, would be invaluable for determining the influencing factors of a technical problem. In addition the usage frequency of individual product functions can be employed to estimate the attrition and promote a data-driven product development.

For all usage information the protection of individual customer privacy is of paramount importance. Since more than one customer can be associated to a product instance this poses a challenge for data analytics. If two product users have diverging privacy settings, the usage information can be incomplete and possibly misleading.

3.3.3. Environmental information

The third category of usage phase data is environment information. This can include weather conditions such as temperature and humidity from on-board sensors or information deducted from third-party services based on the GPS-location of a product. Data from this category can be expected to grow in the future due to cybertronic products becoming more aware of their surroundings. In the automotive domain more advanced driver assist systems and the transition to autonomous driving require explicit information about the vehicles environment. Hence information such as road conditions will become available.

Data derived from autonomous driving systems is highly unstructured. As illustrated for the product condition information, this unstructured data is not suitable to be added to a Digital Twin for product quality monitoring. Hence equivalently to failure indicators, aggregated product attributes have to be derived. The consistency of these attributes across different product variants and product generations is a core requirement for a Digital Twin architecture.

3.4. Assembly and logistics data

All information generated during the manufacturing and distribution lifecycle phases of a product can be summarized as assembly and logistics data. This includes general information about a product instance such as production date and production site. It also includes detailed information about the assembly process because changes in a manufacturing process can introduce errors which in turn can cause product failures.

A Digital Twin needs to be linked to a Bill of Processes (BOP) which lists all actions performed during the assembly of a product instance. Traditionally each process in a BOP includes the parts involved in that process. For the representation in a Digital Twin this representation would have to be reversed: Each process is assigned to a part. That way changes to the manufacturing process of a certain part can be retraced. This reversal is necessary to enable data analytics despite the massive number of manufacturing processes.

Assembly data also includes machine sensor data. This data can be associated to certain parts such as a torque logged during bolt fastening or consist of general machine operating parameters such as internal temperature during the assembly of a product instance. Furthermore non-standard processes and events such as rework during manufacturing have to be archived in the Digital Twin.

Logistics data includes the means of transportation and storage locations as well as the duration of transports and storage for the final product as well as for all parts that make up the product instance. It can also incorporate environmental conditions measured during storage and transport.

4. Digital Twin architecture for product quality management

In addition to product quality monitoring and root cause analysis the concept of a Digital Twin shows a wide range of possible applications. Despite the fact that the potential of data-driven analysis along the lifecycle is well known, the implications of this trend have not translated into the product development process of many traditional manufacturing companies yet. Until recently the purpose of data was to answer specific questions mostly limited to individual product lifecycle phases. Therefore current ways to collect, store and manage data were not developed for the purpose of large scale analytics. As a result the data often lacks availability, quality and persistency and a structure that enables data analytics.

4.1. Analysis

In the previous chapter it was described that up-to-date data-driven analysis of product failures in the automotive domain is primarily based on order information. The introduction of a Digital Twin allows to break with this tradition. This is vital because order information is not sufficient to compare products, as options are neither unique nor static. The same option can have different implications on the manufacturing BOM based on other information in that order. For instance an option might not appear in orders for a country if the option is mandatory in that country. Equivalently when an option becomes standard equipment for a product it will disappear from orders of that product. To cater to the demands of global customers and to manage various legal requirements, companies often employ packages of options that differ from market to market. While this helps to reduce the complexity of managing diverse products, it results in ambiguous information that is of limited use for data analytics.

A Digital Twin for product quality monitoring and root cause analysis must enable the user to identify problems introduced by defective parts, part changes, process changes and user behaviour. The core element is an *as-maintained* Bill of Materials including software versions for each product instance. Parts that can be changed easily by the customer can be excluded from the Digital Twin. As described in the previous chapter the Digital Twin is enriched with data collected during service and repair, with sensor data from the usage phase and assembly as well as logistics data.

The complete Digital Twin contains a vast amount of data. Although a fine data granularity is advantageous when searching for product attributes that distinguish product instances that fail from those that do not, analyzing extremely detailed data is not without problems. If an analysis is conducted based on too many variables, the large number of variables will inevitably lead to spurious correlations. In fact it can be shown that these spurious correlations vastly outnumber the meaningful ones when the number of variables is large enough (Calude and Longo, 2016). This phenomenon is depicted in Figure 2 showing a strong correlation between two completely unrelated variables. This effect is amplified by the relatively low number of product failures in the field, because it increases the chance for a random product attribute to be correlated with the occurrence of a failure by chance alone.

Alternatively the threshold for a statistically significant correlation between product attributes and product failures has to be set very high. With this approach many potentially interesting correlations will be missed, as the relationship between product attribute and failure is rarely deterministic. This

means product failures are often pandered to by certain product attributes but these product attributes are no definite precursor of product failure.

A large number of variables also complicates the interpretability of the results. If the technical root cause of a problem can be traced back to a certain engineering design concept, the product instance will be distinguishable by a large number of part numbers, manufacturing data and different measurement values without their implicit commonality being directly clear. Equivalently if the occurrence of a failure is limited to a geographic region, all local suppliers would appear to be related to the problem without their common denominator being evident.

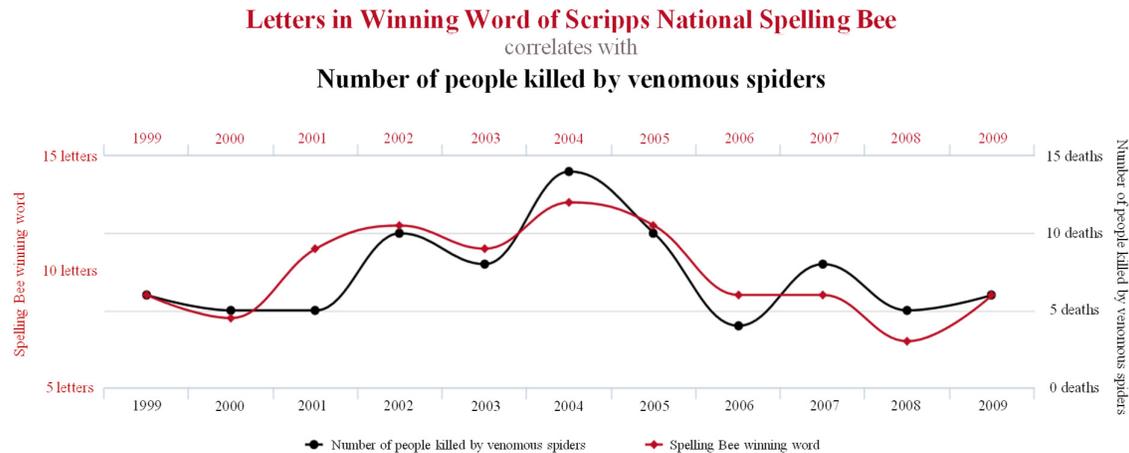


Figure 2. Spurious correlation of random variables (Vigen, 2017)

4.2. Hierarchical structure

The core element of a Digital Twin for root cause analysis and product quality monitoring is an *as-maintained* Bill of Materials of each product instance with all part numbers including revision index. Parts that can be changed easily by the customer or bulk material can be excluded. These core elements need to be recorded when manufacturing and stored as a complete list for each product instance. Data collection during assembly is crucial to account for the delay between product data changes and changes of manufactured products, while parts in inventory are used up.

The analysis in the previous subchapter shows that data analytics on the complete list of information about a product instance is not going to be successful. Hence a Digital Twin needs aggregation of the data. Order information and options are not sufficient as aggregation levels. Hence the Digital Twin requires a novel data-structure that enables Big Data analytics by permitting a piecemeal refinement of the observed data.

We propose a data aggregation layer for the Digital Twin that allows to organize the data hierarchically. For items in the BOM the elements of the aggregation layer will often be the position variant or options in the automotive domain. If the inclusion of parts is not dependent on position variants, they can also be aggregated by a top-node of the generic product structure. The elements of the aggregation layer are also core elements of the Digital Twin and need to be stored for each product instance.

Each part in the Digital Twin is assigned to at least one element in the aggregation layer. These dependencies allow to refine search parameters based on prior information and enable a stepwise data analytics approach. For example if an option has been identified as related to a problem, a focused search can compare different revision indices of dependent parts instead of analyzing the entire BOM of all failed products.

Whenever reasonable, the Digital Twin also needs to include unique serial numbers for each part. Based on the serial number further information can be accessed for a part such as supplier information, manufacturing process data or logistics data. For this part-dependent information the part number serves as an aggregating element, because a comparison, for example of environmental conditions during transport, is only meaningful for equal or similar parts.

General measurement values such as the environment temperature recorded by a product instance should be added to the core elements of the Digital Twin. Equivalent to the aggregation done by position variants for BOM items, this data can be aggregated into statistical values such as average, minimum and maximum temperature. General measurement values can also be dependent on other elements of the aggregation layer like a position variant. For example tracking the burn time can be considered useful for halogen headlights but might not be necessary for LED headlights, because of their long expected service life. Figure 3 shows the possible aggregation and refinement of fictitious data in a Digital Twin for product quality monitoring and root cause analysis.

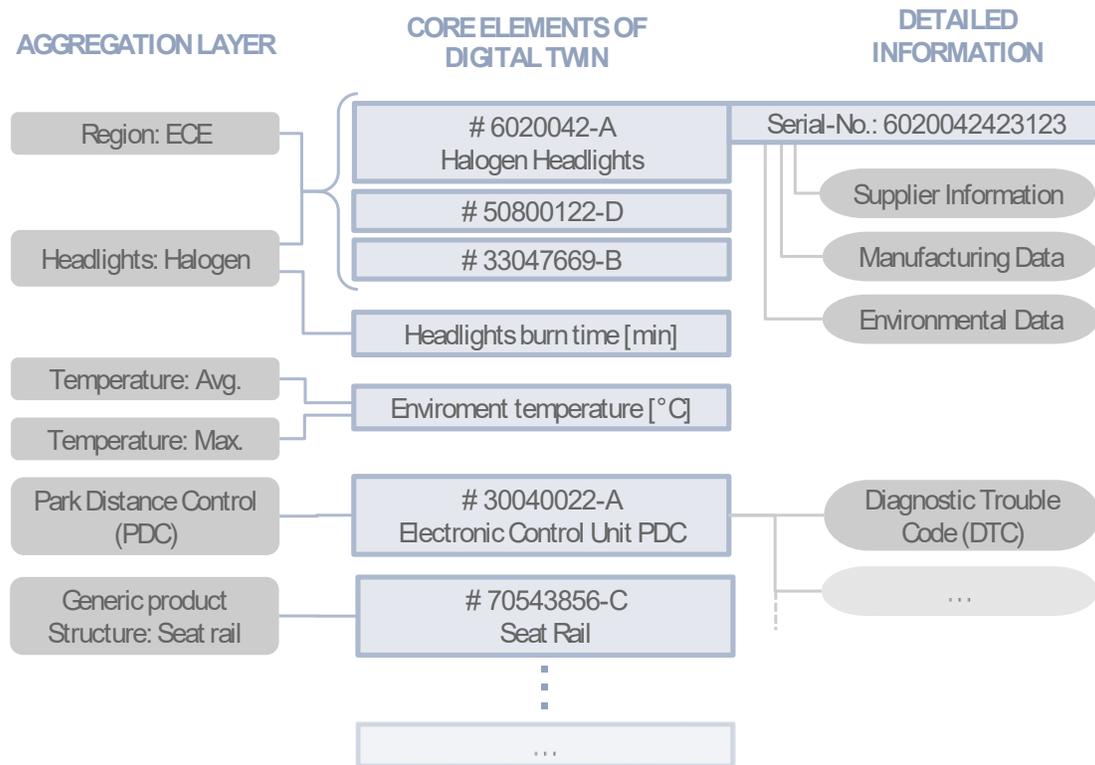


Figure 3. Aggregation and refinement of data in a Digital Twin

A data analysis will normally start at the uppermost aggregation layer. If a product attribute from this aggregation layer shows a correlation with the problem, the analysis can be refined on the subset of products with that attribute. In a manual analysis the user should be able to delve into specific product attributes at will.

To enable Big Data analytics the name or identifier of an element should remain constant over product generations and for similar products. This requirement is necessary to compare engineering concepts or measurement values between different products. Based on the example in Figure 3 the *Diagnostic Trouble Code (DTC)* is only defined for an Electronic Control Unit (ECU) with a particular part-number and revision index. If an equivalent DTC were defined for all Park Distance Control ECUs, the DTC could be added to a higher aggregation layer thus becoming a core element of the Digital Twin.

4.3. Requirements management for data

Data is often collected through various channels and stored in scattered legacy systems. As a result information is divided in silos across enterprise divisions and product lifecycle phases. To enable Big Data analytics, product information has to be consistent across different products and product generations. The generic product structure from PLM systems provides a common framework for product structure data. For usage phase data no equivalent framework has been established. As a result many traditional manufacturing companies lack a cross-enterprise strategy to manage product feedback

data. The collection of usage phase data is often based on requirements during product development and neglects the potential of data analytics in later product lifecycle phases.

Consolidating all data requirements in a Digital Twin architecture can remedy these shortcomings. We propose a meta-data model to manage the elements of a Digital Twin. The meta-data model defines which information is collected and stored for each product instance as well as additional parameters such as units and frequency. Once a product attribute or measurement value is added to the meta-data model, it becomes a requirement for all future products. This ensures comparability between products and product generations.

The meta-data model is not static but can be expanded. When new sensor data is recorded in subsequent product generations or more detailed data becomes available the model is extended. The decision which product attributes and measurement values are added to the Digital Twin has to be made in a cross-enterprise, cross-discipline effort. These data requirements then have to be validated like other product requirements to prevent implausible measurement values and inconsistencies.

For each product instance the meta-data model is a template that defines the elements of the Digital Twin. Utilizing the Digital Twin as a Single Source of Truth minimizes redundant data and potentially conflicting information from heterogeneous systems. Since the same meta-data model is valid for all products, elements of the Digital Twin can be empty. Therefore each datum in the Digital Twin needs a clear distinction between a zero value, a missing value, indicating a transmission or storage issue, or a non-applicable element.

A software solution has to be employed to manage the elements of the meta-data model. The software could also be used to manage the shareholders of each element. This ensures that changes of a datum can be communicated and coordinated between all shareholders. It also establishes transparency for all users of a datum and allows to share the costs associated with collecting and storing the data. To prevent redundant data a central element of the software would be the capacity to search the meta-model. This way users can determine if a certain piece of information is already available.

5. Conclusion

Traditional manufacturing companies collect data through various channels and store it in scattered legacy systems. As a result information is divided in silos across enterprise divisions and product lifecycle phases. Consolidating these islands of information in a Digital Twin lays the foundation for a more efficient product quality monitoring and root cause analysis.

The elements of the Digital Twin proposed in this research serve as a blueprint for data to be collected for each product instance. The proposed hierarchical data structure ensures that the data is suitable for Big Data analytics. This is crucial since cybertronic systems and the Internet of Things will make significantly more product instance specific data available to the manufacturers. Further the notion of spurious correlations between random variables illustrates the challenges of data analytics in large datasets and exemplifies the need for a hierarchical structure of the elements in the Digital Twin.

In consecutive research the authors intent to determine the most promising approaches to select product attributes which best distinguish products that fail from those that do not.

References

- Abramovici, M. and Lindner, A. (2011), "Providing product use knowledge for the design of improved product generations", *CIRP Annals - Manufacturing Technology*, Vol. 60 No. 1, pp. 211–214. <https://doi.org/10.1016/j.cirp.2011.03.103>
- Abramovici, M., Göbel, J.C. and Dang, H.B. (2016), "Semantic data management for the development and continuous reconfiguration of smart products and systems", *CIRP Annals - Manufacturing Technology*, Vol. 65 No. 1, pp. 185–188. <https://doi.org/10.1016/j.cirp.2016.04.051>
- Abramovici, M., Lindner, A., Walde and Dienst, S. (2011), "Decision support for improving the design of hydraulic systems by leading feedback into product development", *Proceedings of the 18th International Conference on Engineering Design (ICED'11)*, Vol. 9: Design Methods and Tools, Lyngby/Copenhagen, Denmark, August 15-19, 2011, The Design Society, Glasgow, pp. 1–10.
- Boschert, S. and Rosen, R. (2016), "Digital Twin - The Simulation Aspect", In: Hehenberger, P. and Bradley, D. (Eds.), *Mechatronic Futures*, Springer, Cham, pp. 59–74. https://doi.org/10.1007/978-3-319-32156-1_5

- Calude, C.S. and Longo, G. (2016), “The Deluge of Spurious Correlations in Big Data”, *Foundations of Science*, Vol. 59 No. 7, pp. 595–612. <https://doi.org/10.1007/s10699-016-9489-4>
- Dean, J. and Ghemawat, S. (2008), “MapReduce”, *Communications of the ACM*, Vol. 51 No. 1, pp. 107–113. <https://doi.org/10.1145/1327452.1327492>
- Dienst, S., Fathi, M., Abramovici, M. and Lindner, A. (2014), “Development of a knowledge-based feedback assistance system of product use information for product improvement”, *International Journal of Product Development*, Vol. 19 No. 4, pp. 191–210. <https://doi.org/10.1504/IJPD.2014.063027>
- Eigner, M. (2017), “Variant Design as Part of an Integrated and Interdisciplinary Product Configuration along the Product Lifecycle”, *Smart Automotive Variant CON 2017, Berlin, Germany, November 29-30, 2017*.
- Eigner, M., August, U. and Schmich, M. (2016a), White Paper: Smarte Produkte erfordern ein Umdenken bei Produktstrukturen und Prozessen, Siemens PLM Software, Köln, Germany.
- Eigner, M., Dickopf, T., Schulte, T. and Schneider, M. (2015a), “mecPro² - Entwurf einer Beschreibungssystematik zur Entwicklung cybertronischer Systeme mit SysML”, In: Schulze, S. and Muggeo, C. (Eds.), *Tag des Systems Engineering: Verteiltes Arbeiten mit ganzheitlicher Kontrolle*, Carl Hanser Verlag, Munich, pp. 163–172. <https://doi.org/10.3139/9783446447288.017>
- Eigner, M., Faißt, K.-G., Apostolov, H. and Schäfer, P. (2015b), “Kurzer Begriff und Nutzen des System Lifecycle Management”, *ZWF Zeitschrift für wirtschaftlichen Fabrikbetrieb*, Vol. 110 No. 7-8, pp. 475–478. <https://doi.org/10.3139/104.111364>
- Eigner, M., Muggeo, C., Apostolov, H. and Schäfer, P. (2016b), “Kern des System Lifecycle Management”, *ZWF Zeitschrift für wirtschaftlichen Fabrikbetrieb*, Vol. 111 No. 1-2, pp. 63–68. <https://doi.org/10.3139/104.111473>
- Glaessgen, E.H. and Stargel, D.S. (2012), “The Digital Twin Paradigm for Future NASA and US Air Force Vehicles”, *53rd AIAA/ASME/ASCE/AHS/ASC Structures, Structural Dynamics and Materials Conference, Structures, Structural Dynamics, and Materials and Co-located Conferences, Honolulu, Hawaii, April 23-26, 2012*. <https://doi.org/10.2514/6.2012-1818>
- Grieves, M. and Vickers, J. (2017), “Digital twin: Mitigating unpredictable, undesirable emergent behavior in complex systems”, In: Kahlen J., Flumerfelt S. and Alves A. (Eds.), *Transdisciplinary Perspectives on Complex Systems*, Springer, Cham, pp. 85–113. https://doi.org/10.1007/978-3-319-38756-7_4
- Kassner, L., Gröger, C., Mitschang, B. and Westkämper, E. (2015), “Product Life Cycle Analytics – Next Generation Data Analytics on Structured and Unstructured Data”, *Procedia CIRP*, Vol. 33, pp. 35–40. <https://doi.org/10.1016/j.procir.2015.06.008>
- Lee, J., Bagheri, B. and Kao, H.-A. (2015), “A cyber-physical systems architecture for industry 4.0-based manufacturing systems”, *Manufacturing Letters*, Vol. 3, pp. 18–23. <https://doi.org/10.1016/j.mfglet.2014.12.001>
- Lee, J., Lapira, E., Bagheri, B. and Kao, H.-A. (2013), “Recent advances and trends in predictive manufacturing systems in big data environment”, *Manufacturing Letters*, Vol. 1 No. 1, pp. 38–41. <https://doi.org/10.1016/j.mfglet.2013.09.005>
- Li, J., Tao, F., Cheng, Y. and Zhao, L. (2015), “Big data in product lifecycle management”, *The International Journal of Advanced Manufacturing Technology*, Vol. 81 No. 1-4, pp. 667–684. <https://doi.org/10.1007/s00170-015-7151-x>
- Madenas, N. (2014), Integrating product lifecycle management systems with maintenance information across the supply chain for root cause analysis, PhD thesis, Cranfield University.
- Muggeo, C. and Pfenning, M. (2015), “Die Rolle von MBSE und PLM im Industrial Internet”, In: Schulze, S. and Muggeo, C. (Eds.), *Tag des Systems Engineering: Verteiltes Arbeiten mit ganzheitlicher Kontrolle*, Carl Hanser Verlag, Munich, pp. 279–287. <https://doi.org/10.3139/9783446447288.028>
- Rios, J., Hernandez, J.C., Oliva, M. and Mas, F. (2015), “Product Avatar as Digital Counterpart of a Physical Individual Product: Literature Review and Implications in an Aircraft”, In: Curran, R., Wognum, N., Borsato, M., Stjepandić, J. and Verhagen, W.J.C. (Eds.), *Transdisciplinary Lifecycle Analysis of Systems, Series: Advances in Transdisciplinary Engineering, Vol. 2*, IOS Press, pp. 657–666. <https://doi.org/10.3233/978-1-61499-544-9-657>
- Tao, F., Cheng, J., Qi, Q., Zhang, M., Zhang, H. and Sui, F. (2018), “Digital twin-driven product design, manufacturing and service with big data”, *The International Journal of Advanced Manufacturing Technology*, Vol. 94 No. 9-12, pp. 3563-3576. <https://doi.org/10.1007/s00170-017-0233-1>
- Tuegel, E.J., Ingrassia, A.R., Eason, T.G. and Spottswood, S.M. (2011), “Reengineering Aircraft Structural Life Prediction Using a Digital Twin”, *International Journal of Aerospace Engineering*, Vol. 2011 No. 3. <https://doi.org/10.1155/2011/154798>
- Uhr, P., Dienst, S., Klahold, A. and Fathi, M. (2012), “Kontextbasierte Bereitstellung von Textdokumenten im Produktverbesserungsprozess”, *Werkstattstechnik online*, Vol. 7-8, pp. 528–534.

- Vigen, T. (2017), *Spurious Correlations*. [online] Tylervigen.com. Available at: <http://www.tylervigen.com/spurious-correlations> (accessed 09.06.2017).
- Zagel, M. (2006), Übergreifendes Konzept zur Strukturierung variantenreicher Produkte und Vorgehensweise zur iterativen Produktstruktur-Optimierung, PhD thesis, Technische Universität Kaiserslautern.
- Zhang, Y., Ren, S., Liu, Y., Sakao, T. and Huisinigh, D. (2017), "A framework for Big Data driven product lifecycle management", *Journal of Cleaner Production*, Vol. 159, pp. 229–240. <https://doi.org/10.1016/j.jclepro.2017.04.172>

Alexander Detzner, M.Sc.
Technische Universität Kaiserslautern, Institute for Virtual Product Engineering
Gottlieb-Daimler Straße 44, 67663 Kaiserslautern, Germany
Email: alexander.detzner@bmw.de