

Chapter 7

Data Quality Management for Interoperability



Josip Stjepandić and Wjatscheslaw Korol

7.1 Introduction

Data becomes an even more important constituent of our daily lives and an essential asset, driver and fuel for mature and new industries. While data is becoming a crucial importance as a strategical resource, the quality of the data that is used during business processes will impact the business outcome today and tomorrow [1]. As a fundament of the digital economy, data must have a high quality. For the data quality, the same semantic understanding is transferred from the production of physical goods to the management of data [2]. The quality of data, defined by the objectively measurable degree to which the data properties fulfill requirements, can have a tremendous impact on the businesses [3]. Subsequently, data can be seen as a raw material for the production of information products through an information production process [4]. This also fits the definition of information as the “processed data”. Since information products reproduce itself almost recursively, the terms data and information can only be distinguished in the basic process chains and used synonymously further [5]. Data quality is usually assumed [6]. It becomes a topic once the data quality falls below a certain threshold [7].

Research and industry reports continuously discover that huge efforts are spent to improve the quality of the data being used in many applications, sometimes even only to understand the quality of auxiliary data, in order to preserve a proper function of information systems. Data quality is defined as a context-dependent, multidimensional property and expresses the fitness for use of certain data for a user in a specific context. The inherent context dependence of the data quality emphasizes that the

J. Stjepandić (✉) · W. Korol
PROSTEP AG, Dolivostr. 11, 64293 Darmstadt, Germany
e-mail: josip.stjepandic@prostep.com

W. Korol
e-mail: wjatscheslaw.korol@prostep.com

requirements for data elements depend on the intended use and only the user can decide for himself whether the data objects are usable for him or not. The context can include, for example, the data-using business process, country affiliation, applicable regulations, time of data usage, data-processing application or the business process role of the data user [8].

Considering the variety of business views, use cases, properties, or simply the specificities of the systems being evaluated, the quantitative assessment of the data quality can become an extremely difficult task which can hardly provide clear results [9]. In general, the importance of data quality increases the more complex the business processes become and the more applications and interfaces have to interact [10]. Data translation mostly impairs the data quality [11] and, therefore, cause for several issues.

In addition, there are many data quality dimensions which, in combination, express the data quality (Fig. 7.1): timeliness, credibility, reliability, interpretability, operability and sufficiency accompanied with data quality attributes. The quality of a data object can be measured by checking certain properties of the values of the data elements it contains [13]. If the data elements of a data object have all the required properties, its quality is perfect. Properties of data (e.g. certain value is mandatory within a range, completeness of a data element) can be formulated as business rules and thus checked. There are structural (how must a data object be structured?) and operational business rules (how do values for individual data elements have to be set?) [8].

Data Quality Management is a subordinated area of data management, which, as part of company-wide information management, aims to make optimal use of data in the company. Data quality is never an end in itself. As a typical supporting topic, it gets its meaning through a process chain, mostly as a pre-requisite for a highly performant interoperability [14, 15]. Subsequently, the term digital thread refers to a communication framework [16]. Companies can only offer digital services, open up new business opportunities or make processes between companies more efficient, if data on customers and products, but also context information on whereabouts, preferences, and billing are available in high quality [1].

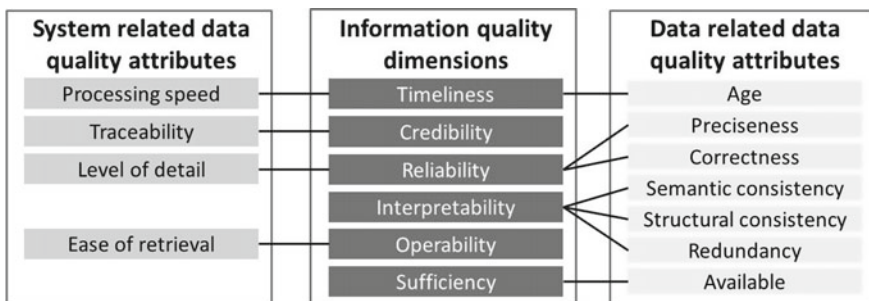


Fig. 7.1 Data and information quality model [12]

Like other business assets, data also has its own lifecycle that needs to be managed in a suitable way to continuously ensure its purpose and need. The lifecycle of a data object begins with its planning and specification, through the creation during the execution of business processes, and ends with their archiving or deletion [8]. It is important to say that the life cycle of the data is always longer than the life cycle of the product concerned due to the regulatory rules. This fact has a crucial impact to Digital Twin because all components must be adjusted with one another in order to enable seamless interaction. A high data quality is a pre-requisite for this.

The given definition of the data lifecycle corresponds to the understanding of the product lifecycle, which also begins with the first product requirement and not with the current product representation. Established reference models of Product Lifecycle Management (PLM) describe several process stages for planning and implementing the entire product life cycle [17]: planning the product portfolio, designing the product, planning the production process, supplying the end customer with the product, providing service and support services, and, finally, product disposal and recycling. In era of model-based definition and processes, the data management including data quality management is a function within a powerful commercial PLM system [17, 18]. Thus, one of the aims of PLM is the provision of data in a sufficient quality for downstream processes.

The structure of this chapter reflects this aim. In Sect. 7.2, Digital Thread and its supporting concept are briefly introduced. Data quality classification with respect to its dimensions and related standards is discussed in Sect. 7.3. Subsequently, achievements related to the issue of data quality metrics in manufacturing industry is introduced in Sect. 7.4. Section 7.5 showcases the achievements of data quality in industrial applications of design and manufacturing for various industries. A discussion chapter in Sect. 7.6 gives insight into benefits and gaps of current applications of data quality as well as future directives. Finally, an outlook is given in Sect. 7.7 with respect to the future importance of data quality from business process perspective.

7.2 Digital Thread

The fundamental vision of integration in manufacturing industry supposes a seamless flow of information in all product lifecycle stages from the first product idea until disposal and recycling. This vision is expressed by three layers as illustrated in Fig. 7.2 which can be seen as structural constituents of Digital Twin.

Data integration is a widely requested ‘digital data’ lever for digital transformation. It describes a product (born digital) holistically with (1) domain-specific application models for example, mechanical, software, simulation or cost models. It demands cohesive communication in the (2) supply chain based on business data streams with partners, in joint ventures and across factory plants [19]. It finally realizes (3) a fusion between up- and downstream in the entire lifecycle, where digital aspects of the product solely are used as engineering, manufacturing and service

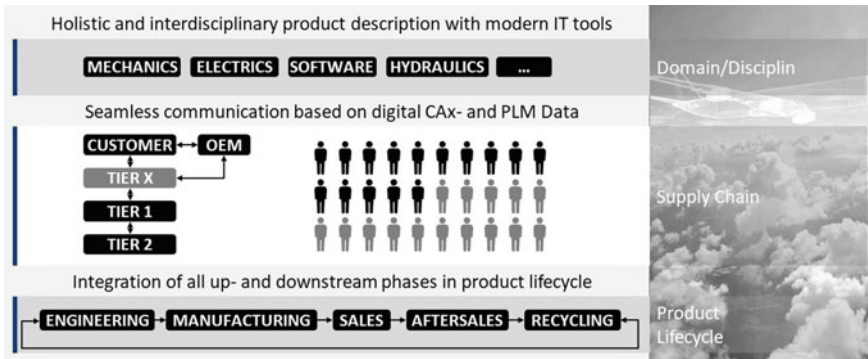


Fig. 7.2 The integration vision in the automotive industry [20]

bridges [20]. Although all three layers have its particular importance for Digital Twin, the product lifecycle layer (3) prevails due to the volume and complexity.

The connection between the real asset and the development and planning models that describe its history is known as a digital thread [16]. Like a data highway, it connects the information of a real product instance across processes and IT systems. On the one hand, this enables all data from the lifecycle of the product instance or the real asset to be brought together and thus forms the basis for the creation of Digital Twins. Without digital thread, the Digital Twins could be recreated manually, but it would be difficult or impossible to keep them up to date. On the other hand, the traceability along the digital thread makes it possible to track and monitor decisions in development and production and to identify potential for optimization with the help of the operating data [21].

Product data is an umbrella term that includes many different types of information: PDM, CAX, planning, inspection data [17]. Usually only certain data is necessary at sequences (gates) for assessment in downstream processes [18]. Large and comprehensive functionalities, such as provided with modern CAX systems are not mandatory. In fact, the number of consumers of CAD data in extended enterprise is about a factor of ten at least higher than data creators in engineering. The use of powerful CAX systems however in downstream processes such as purchase, production, assembly and quality assurance needs to be scrutinized in the course of an efficient product creation [22]. Certain level of data quality and level of detail (filter) is always supposed.

7.3 Data Quality Classification

To properly place data quality, clear classification criteria are necessary. Classification can be based on several criteria (environment, organization, purpose, status). For

our purpose, its classification relative to the afore-mentioned dimensions (Fig. 7.1) needs to be understood.

7.3.1 Data Quality Dimensions

Ensuring the timeliness of data processing requires the ability to acquire, transfer, process, transform and use the data in required time. It comprises the temporal capability of the virtual representation of entities from the real world, based on the data quality attributes age and processing speed. Age refers to the time passed since a change of the measured value in context of change. Processing speed is a system-related attribute measuring the interval from data collection to information provision [12]. Timeliness expresses the time expectation for accessibility and availability of data. As the value of data can rapidly decrease over time, the computing architecture needs to perform all the calculation and communication almost on the fly with the data that was recently provided.

Data credibility expresses the level of certitude to which the good faith of a source of data can be relied upon to facilitate its reuse in order that the data really represents is what the data is supposed to represent, and vice versa [23]. In other words, data credibility indicates the confidence of the Digital Twin users supported by the trustworthiness of the provided data. It is based upon the consistency with other evidence. It fosters the willingness to rely on the data with the goal to increase the intensity of data usage [12].

Data traceability is the ability to track a data construct back to the construct it was derived from as a more concrete instantiation. This can be ensured by metadata that track information provenance, for instance implemented in form of a data pedigree (effects of attributional qualities of a source). A pedigree is a list of ancestors with some attribution of purity of the lineage [24].

Data reliability is the degree to which prior historical reports from a source have been consistent with fact. Reliability includes a notion of dependability, that the data will be produced, and attain some level of accuracy and precision [24]. The data quality attribute correctness binarily differentiates data considered ‘correct’ or ‘incorrect’. This requires that all data values for a business attribute must be correct and representative of the attributes. Preciseness is a data-related attribute measuring inaccuracy on data item level, while level of detail is determined by system limitations. While this is trivial for numeric values, other data types need to be translated into a numeric representation that allows for deviation measures, or special distance metrics need to be applied [12].

The information provided from Digital Twins must be interpretable for the users. Interpretability can be assessed at two different levels: by examining models (heuristic approach) or representations (mainly with user-based surveys). In the former case, simple measures can be used to compare several models of the same type, such as the number of rules and terms in decision rules or the number of nodes

in decision trees. If models differ, then this comparison is not that obvious and other heuristics have been proposed [25].

The provided data should be free of unintentional duplicates, as expressed by the data quality attribute redundancy. In contrast, an intentional redundancy can be useful and improve process reliability [4]. Besides, defined rules need to be satisfied, as expressed by semantic consistency and structural consistency. Semantic consistency describes consistency of the meaning of data, achieved through unified definitions, labels assigned to real-world objects, and vocabulary [12]. Semantic consistency is important for the mathematical data quality of CAD data. Structural consistency refers to technical specifications of structure and format and impacts the organizational quality of CAD data [4].

Data operability describes a level of data record ability to be used directly, without additional processing (translation, filtering): how the information consumer interacts with the Digital Twin. The quality attribute ease of retrieval is assessed on a scale from 'inaccessible to user' to 'machine readable and ready as input for analysis software' [12]. It belongs to the specific data operability challenges such as semantic duplicates, data fusion or information extraction.

Data sufficiency is related to the amount of information provided to fulfil a certain purpose, e.g. a complex assembly feature. Insufficient data yield unstable models. The attribute availability of a dataset describes whether mandatory data items are available, e.g. in case of data replication [12].

From the engineering perspective, the data quality dimensions can be roughly classified in two groups: technical and organizational. Technical dimensions are primarily related to the capabilities of product and organizational to the capabilities of process. In order to check the data quality by appropriate tools, technical dimensions are checked for singular entities. In contrast, organizational dimensions are checked for data structures.

7.3.2 Related Standards

The primary purpose of standards is to define compliance clauses in a way that vendors can claim compliance to differentiate their offering from those that are not compliant. Compliance with an international, industry or proprietary standard is often a legal pre-requisite for a supplier to get a contract with his customer [19]. While a specific Digital Twin standard, which would define the necessary data quality requirements, neither exists yet nor can be expected in the next future, the existing standards related to data quality will be shown here.

ISO 8000 is a set of data quality management standards developed by ISO TC 184/SC4/W G13 [26]. The committee's mission is the development of standards for the exchange of complex data in an application neutral form to provide data portability and long-term data preservation in an environment where the life cycle of software applications used to capture and manage data is but a fraction of the life cycle, if the data itself lie in focus of this committee. ISO 8000 defines which characteristics of

data are relevant to data quality, specifies requirements applicable to those characteristics, and provides guidelines for improving data quality. It deals with Master Data, Transactional Data, Referenced Data and Engineering Data. ISO 8000 standards can be applied to manufacturing processes defined in IEC 62,264 (Enterprise-control system integration). Manufacturing processes defined in IEC 62,264 are restructured according to processes of ISO 8000-61. Each process consists of purpose, outcomes and activities. Achievement of each process can be confirmed with work products [27].

For sake of Digital Twin, the data quality management process reference model of ISO 8000-61 [28] and process assessment of ISO 8000-62 [29] are of a particular importance. ISO 8000-61 specifies the processes required for data quality management. Each process is defined by the purpose, outcomes and activities that are to be applied for planning, controlling, assuring and improving data quality. It comprises also the data-related support and resource provision. The processes are used as a reference model in assessing and improving data quality management. The implementation cycle is based on the 'Plan, Do, Check, Act'-cycle defined in ISO 9001. Based on 8000-61, ISO 8000-62 identifies those elements of the maturity model that exist in other standards and specifies additional elements of the maturity model. ISO 8000-62 provides guidance on assessing the maturity level of an organization and derives organizational process maturity level rating from process profiles. The assessing of the organizational maturity level for data quality management conveys how well the organization is fulfilling the requirements identified by the process reference model specified by ISO 8000-61 [27]. ISO 8000-62 specifies six maturity levels and process profiles to indicate when organizations have achieved each of the maturity levels. ECCMA (Electronic Commerce Code Management Association) has developed a series of compliance certificates for individuals, organizations and their software applications and data services [29].

IEC 62,264-1:2013 defines the functions of an enterprise involved with manufacturing and the information flows between the functions that cross the enterprise-control interface in order to improve integration regardless of the degree of automation. Global acting companies are very interested in it because it unifies and merges different IT methods and enables robust, easy-care integration solutions to be achieved in the long term. This standard is important for both manufacturers and users and system integrators. It offers a uniform terminology for corporate IT and control systems as well as a number of concepts and models for the integration of corporate functions. The technical solution is determined by the uniform modeling of the interfaces between the corporate functions and control functions. The main concepts are object modeling and the modeling languages UML and XML [30].

Among the industry standards for data quality, the guidelines that have been developed by the German and international automotive industry since 1980 have achieved a widespread use. These activities were driven by increasing complexity of product models which are described by thousands of CAx and PDM tools in the automotive supply chain. The difficult data exchange was caused by not powerful interfaces and low data quality [31].

The main purpose of the widespread VDA (German automotive association) recommendation 4955/2 is to improve the collaboration of project partners during the product development phase [32]. This guidance helps to reduce remastering times and costs of CAD processes through the exchange of information and experience, the definition of cross-company, common data quality criteria, the provision of both CAD system-neutral test programs and repair aids as well for the frequently used systems or system pairings. The recommendation is divided into the following areas:

- geometrical data quality,
- organizational data quality,
- recommendation for agreement among data exchange parties, and
- recommendation for a proper extent of CAD models.

This recommendation is written in such a level of detail like an implementation guide for the development of check tools. Although the fulfillment of quality criteria is checked using proprietary system functions, the criteria are formulated neutrally in order to preserve comparability and system independence of the results. The translation and check software usually are certified by VDA against this recommendation, whereby the various check tools are compared by using different test models from practice in order to guarantee the reliability and consistency of the test results. An example for consistency issue in a CAD model with a virtual gap in solid that really does not exist is shown in Fig. 7.3.

Similar standards were established in further countries with automotive industry. Then, these have founded the association SASIG (Strategic Automotive product data Standards Industry Group) which has taken the further development of data quality standards. The result was “Product Data Quality Guidelines for the Global Automotive Industry” [33], which has adopted and extended the previous works of VDA. The last version of this document entails CAD, CAE, PDM and inspection data. It contains some suggestions for project management, communication and know-how for better CAD model quality. However, these suggestions are generic and cannot be applied directly. A direct implementation in a check tool is not known yet [33].

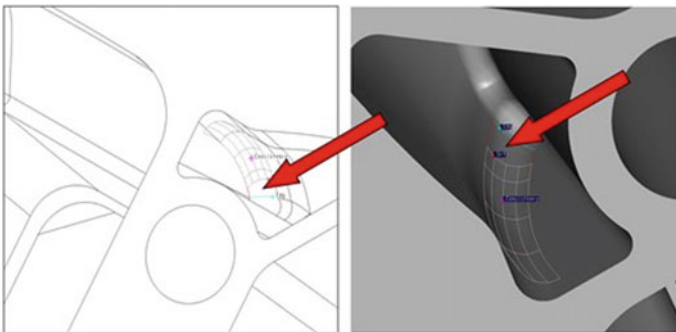


Fig. 7.3 Exemplary data quality problem: consistency [31]

7.4 Data Quality Metrics

A metric, or Key Performance Indicator (KPI), is a quantifiable attribute of an entity or activity that helps to describe its performance [18]. This can be measured to help, manage and improve the entity or activity. In term of data quality, it represents a set of attributes which describe the data quality in a sufficient way. From a quality perspective, two moments have importance during the data's lifetime: the moment it is created and the moment it is used. The quality of data is fixed at the moment of creation or change. However, data quality is usually not assessed until the moment of use. If the quality looks low, users typically attempt by working around the data or correcting errors themselves. Therefore, quality of product models needs to be continuously controlled in engineering workflow, especially in systems based on downstream data.

Model quality impacts not only the model accuracy and modifiability but also the changeability of the whole engineering systems. Careful and thorough model verification facilitates effecting product model quality. Verifying product models and designs manually is a tedious and time-consuming process [19]. By automating parts of the verification process, e.g. by using intelligent templates for check tools [34], benefits can be achieved in the time frame and end results of the verification.

The metrics presented below belongs primarily to CAD data as the main input for Digital Twin. A 'one size fits all' set of metrics is not a solution and that assessing data quality is an on-going effort that requires awareness of the fundamental principles underlying the development of subjective and objective data quality metrics. The main dimensions for the metrics are shown in Fig. 7.4. The dimensions contain almost universally agreed model quality dimensions described in Sect. 7.3.1. These dimensions should be the basic principles when assessing product model quality. In addition, often used quality dimensions are accessibility or reachability of data. Considering the downstream processes, modifiability and reusability are the paramount dimensions. However, for all-around measurement for quality of configurable product models, even more additional dimensions are needed [35].

Simplicity of product models defines the topological structure built of a few simple and understandable elements. Simple models facilitate consistency in the modeling

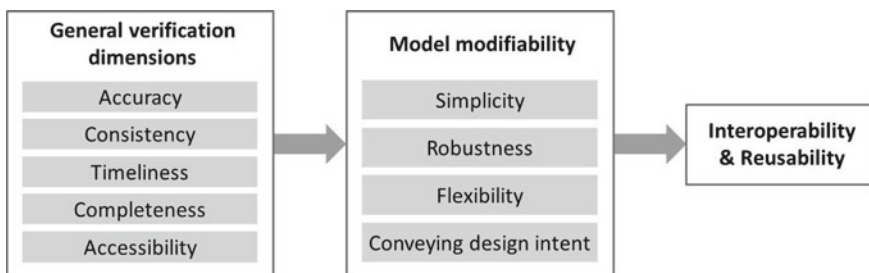


Fig. 7.4 Model verification metrics [35]

system in order to reduce causes of possible instability. Model instability would cause huge additional efforts to repair the model. Robustness of a model describes the resistance to error while being modified. It is a prime indicator for overall model quality as it is a result of minimizing error and quality issues from models [35]. Basically, the approach to create robust models means to create simple models with as simple features as possible. Referencing increases the robustness if detail feature shows to the basic feature [31].

Flexibility expresses the range of reachable states depending on the time and cost required to change state. Flexible systems are built for a set of reachable states which are predefined during the engineering process [35]. Interoperability describes how a master model can be accurately transferred from one format to another, e.g. CAD to CAD or point cloud to CAD. Interoperability can be efficiently supported to create and follow common methodology for modeling in all systems. Interoperability is a prerequisite for seamless downstream processes in Digital Twin [31].

Reusability describes good model reuse derived from the structures and references of the model. Reuse can be performed at different levels from utilizing library component—as practiced in our solution—to using existing designs of similar properties. As the modification to the models in the case of reuse is frequent, the quality of reusable models must be higher than normal as they need to reliably allow for modifications while maintaining the original design intent. Design intent and rationale are always needed for a model to be reusable [35]. Therefore, Digital Twin requires a native model in similar quality as a manually generated CAD model following a method. Conveying design intent is important for some newer engineering processes such as Model-based Design (MBD), Model-based Engineering (MBE) and Knowledge-based Engineering (KBE). Simplicity helps to convey the intent of the original design.

On a morphological level the model quality can be quite effectively quantified even with CAD environment native tools. Different type of geometry checks and identification of topological errors is usually even built-in the software to make sure the software works as it should [35]. Such functions are often collected in a specific data quality check module. Such modules can be easily controlled by using pre-defined profiles for specific purposes [34]. The most important metric on data quality is the distribution of error rate during a period, e.g. project duration [4].

For product data in PDM different rules of data quality are used which mostly refer to structural data quality dimensions. This is caused by configurable product data structures which are the fundament of PDM. Because of this, the consistency of data in the system and planning of the data structures is crucial. Furthermore, the consistency of data between different systems in the modeling environment is one of the most important aspects in modern multi-environment engineering systems. Many key performance indicators (KPI) have been established for product data in PDM. To these, primarily belong consistency, completeness and timeliness [17]. These metrics can be used to determine the quality level of data sets, but they are not as suited for evaluating and verifying the quality level of single product data instances. Furthermore, some metrics or parameter thresholds need to be used to correctly evaluate if product data is incomplete or not [35].

7.5 Practical Examples

The framework of the practical application is usually formed by the standard steps according to ISO 9000: quality planning, control, testing, and improvement. The focus is set here to control and testing. Practical management of data quality consist of methods to ensure required data quality and their implementation in process chains—primarily by deploying suitable data quality check tools. Good product data quality means providing the right data to the right task at the right time [33]. Data provision is realized by workflows in modern PDM systems. Data quality methods (‘Design for Quality’) help users in usage of IT systems (e.g. CAD) to ensure the quality of their work from the perspective of various stakeholders. It consists of approaches for manual techniques for assessment of CAD models, identification and repair of typical model errors, as well as application of modules and tools for interactive improvement of models [31].

The quality assurance should occur in each phase when data is changed. It gets its particular importance during the detail design, when both the amount of new generated data and the frequency of changes are high. Therefore, the practical data quality assurance is focused either on the data creation period or its phase of exploitation. In this section, three practical examples are presented which belong to Digital Twin: the use of a commercial tool in the design phase with focus on the methods and training (Sect. 7.5.1) [36], the conception and implementation of a knowledge-based check tool for downstream application in manufacturing (Sect. 7.5.2) [37], and the control of CAD data migration process by using a CAD quality tool (Sect. 7.5.3) [4].

7.5.1 Design

The study presents the use of a standard, commercial module of SolidWorks (SolidWorks Design Checker—SWDC) in a representative case study of modern Model Quality Testing tools (MQT) [36]. SolidWorks is a leading CAD system, widely used in manufacturing industry which is also applied as authoring system for factory models in our DigiTwin project [37]. SWDC can identify and sometimes repair data errors that could affect the simplification, interoperability, and reusability of CAD models.

This study has investigated the usefulness of this check tool as an assessment mechanism both for instructors and for self-evaluation. SWDC integrates the following modules: build checks, check active document, check again existing file, and learn check wizard. By mapping the Build Check requirement of SWDC against the CAD quality criteria available in the literature, the main conclusions can be drawn that SWDC only covers lowest semantic level quality criteria, and is designed for intensive use to maintain consistency across documents [36].

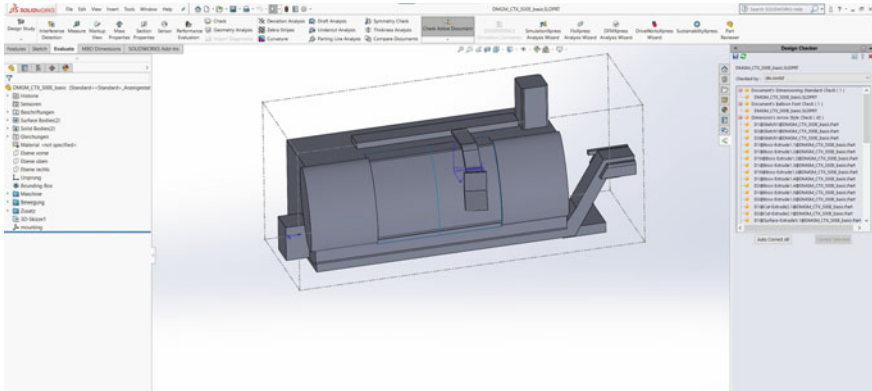


Fig. 7.5 Usage of SolidWorks Design Checker—SWDC in context of digital twin

This study provides an exhaustive insight how to map requirements to quality dimensions. Issues arise on detection of constraints that are repetitive but not incompatible. In general, the structure of the application is designed for intensive use and to maintain consistency across vast amounts of documents. Most criteria implemented in SWDC are aimed at verifying settings, and thus intended to ensure the semantic correctness of the CAD model [36].

Two additional observations are included: SWDC will repair certain errors, but others (not all) will only be identified, and SWDC partially overlaps with the built-in checking capabilities of the CAD application, which can sometimes perform better than the MQT [36]. The findings include the insight that the product data quality can be achieved by design rather than by MQT tools. This is practised in our development of Digital Twin (Fig. 7.5) [37].

Finally, although MQT is considered solved by a number of scholars, it remains an open practical problem, as new quantitative metrics require the design of new application programming interfaces that transform current MQT tools into mechanisms to assess higher semantic level quality aspects [36].

7.5.2 *Manufacturing*

Different model users in various stages of the manufacturing process have varied requirements in terms of model quality. To handle this diversity of models, a study proposes a knowledge-based MBD part model quality analysis system and its implementation technologies to analyze and test the quality of models from the perspective of different model-used stages. Alternatively, such a system would need to create partial models by using entity filters. It is fully integrated into CAD system Siemens NX and, therefore, uses system functions for its operations. The system decomposes

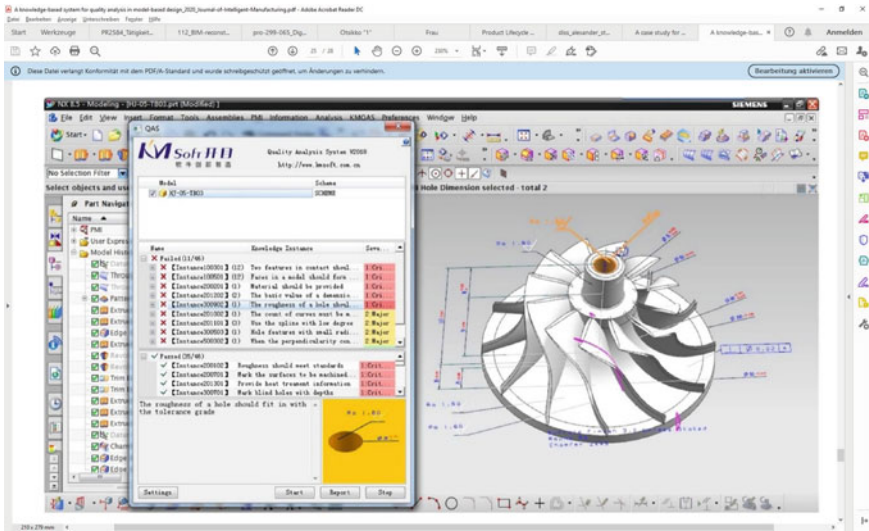


Fig. 7.6 Data quality analysis in model-based design [38]

the MBD part model into model definition instances (MDI) as well as the relationship between them for model quality verification with the help of relevant model quality knowledge. The model quality knowledge from different model-used stages enables the system to analyze quality defects from the respective perspective of model-used stages in MBE [37]. After the analysis is finished, model quality issues and corresponding modification suggestions are written into report.

Based on a knowledge-management system with its own data quality data model, this tool acquires new model quality knowledge into the knowledge base. The rule-based framework of knowledge representation facilitates the functional extension of this system. The applicability and expressive power of the rule-based framework is enhanced by object chain rules and parameter table rules (Fig. 7.6). In practice, with the help of the system, the knowledge derived from various model-used stages in MBE is collected, stored, and reused in downstream processes [38]. The unknown model defects are represented by new knowledge objects.

Apart of the detection of the model quality defects, this system provides modification suggestions in the design stage. However, the quality defects mostly need to be addressed manually, which results in a lot of repetitive work. Hence, new quality defects may occur for the wrong modification operations. Moreover, only the qualitative analysis is provided in this system [38]. The most important drawback is the missing comparing functionality. It is difficult for the analysis results of different models to be compared directly. Besides, for different quality analysis requirements from various models, the analysis scheme has to be created manually [38]. It is tedious in case of an assembly model which contains of a large variety of part models.

For a frequent use in generation of a Digital Twin three improvements should be made: the procedure or workflow to automatic quality defects modifications based

on knowledge, the technology for the quantitative analysis of model quality, and the approach to automatic generating quality analysis scheme according to the model, as well as relevant knowledge [39].

7.5.3 Data Migration

Data migration refers to a large system change which includes the translation of a huge quantity of data which were often collected during decades within a short time period. Such migrations happen every ten to fifteen years and set a huge challenge for the organization and the users. Remastering of a factory with all the equipment, e.g. by using the method presented in Chap. 6, can be understood as a migration of the factory representation, too. This is valid in particular, if the object recognition does not recognize all objects which subsequently must be built by a singular feature recognition or manual remastering. Generation of Digital Twin must preserve an option to remove a system by another.

In such a case, a huge change arises in the customer process of almost each internal and external supplier, because they are forced to keep the current process running and to ramp-up the new process. This procedure includes many CAD translation steps which are principally not beneficial for good data quality. The challenge is ensuring an appropriate level of data quality to make sure that all translating processes are successful [4].

The translators in modern CAD systems show a similar level of performance and robustness as known from previous benchmarks and long-term experience [4] and, therefore, all models can be transferred lossless without exception to Solid-Works. However, it appears that in some cases automatic healing algorithms have slightly adjusted the geometry to satisfy the continuous condition [39]. It could not be predicted definitely to what extent of additional problems in further processing this will lead [4].

Further comparisons in the model properties like center of gravity, moments of inertia, as well as cloud of points, were also executed systematically (Fig. 7.7). All values remained within the allowed tolerances and showed no abnormalities what indicated a mature and stable and reliable process [4].

7.6 Discussion and Future Perspectives

Despite of its huge achievements described in the use cases above, data quality is still subject to intensive research and practical improvements. The valuation of data quality can be facilitated in two directions: operational control of data quality as autonomous procedure (supported by an application) and data quality as inherent component of the business process as a whole.

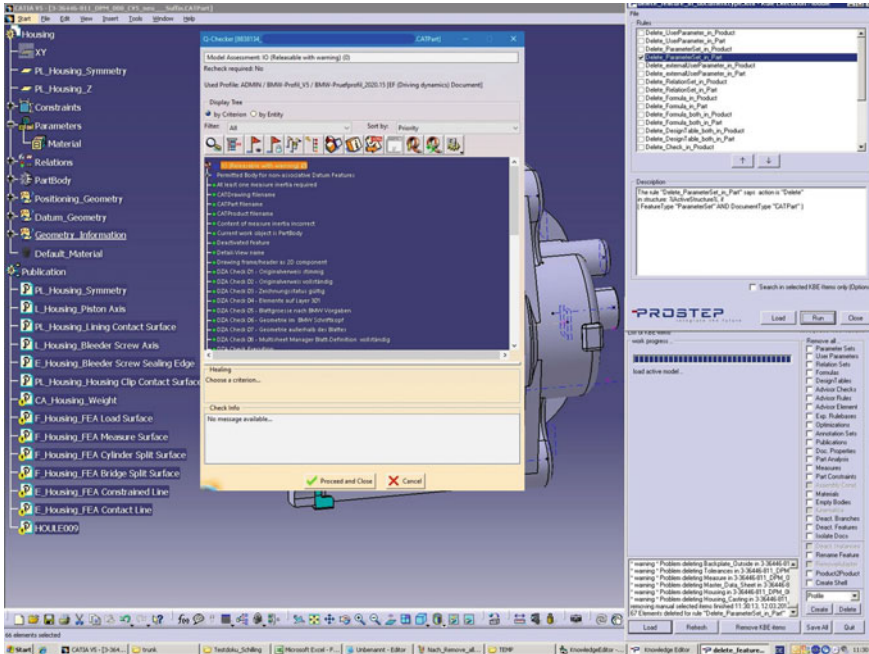


Fig. 7.7 Results of data quality check of migrated data

When considering the operational control of data quality, a number of developments have put data quality on a stronger footing over the last decade. In particular, the development and uptake of data quality dedicated design methods, the uptake of data quality assessment functionality in most major CAD and PDM systems, and the development of supporting methodologies as well as check tools have strongly contributed to the higher sensitivity for the data quality. However, a number of challenges remain to be solved.

On the operational level, better quantitative metrics are required to measure higher semantic level quality aspects. CAD model reuse is particularly sensitive to hidden errors and anomalies. Design intent is still poorly addressed by MQT tools, but there are other bottlenecks for different stakeholders (Original Equipment Manufacturers (OEM), lower tier suppliers, and SME's) [40]. Profiles must become robust (changes do not produce unexpected failures) and flexible, e.g. modular (allow as many changes as necessary) [41]. If we can experimentally validate the hypothesis that the flexibility of the profiles depends not on the amount but on the semantic level of the constraints, valid metrics for flexibility may follow. For instance, detecting excessive use of poor 'fix' relations that lock point coordinates is an example of the type of high semantic quality tests that are not supported by current CAD check tools [32].

Dominant OEMs force top-down interoperability with their suppliers, which results in 'defensive' or 'conservative' designs, which are robust but hardly creative. Interoperability is a main concern for OEMs, as reusability is guaranteed by the

best practices they impose, whereas simplification tasks are transferred to suppliers [42]. A hidden problem that hinders interoperability is the lack of proven modeling guidelines. Best practices are checked by MQT tools but also imposed and tuned by the OEM, whose current goal involves improving interoperability by abandoning explicit representations and adopting STEP AP 242. Validation rules require setup, and although quantitative metrics for shape errors already exist, they are context dependent and are governed by computational threshold values that are different for each MQT [41].

We see that the demands on data quality are getting higher and higher [42]. Therefore, the way practiced here appears to be very advantageous, preferably to use the models from a higher-quality library of CAD models. The alternative approach of performing object recognition via feature recognition in several stages requires a data quality check and subsequent repair procedure [43].

Considering a holistic approach, development of a data quality assessment tool—a dashboard—in addition to policies and protocols to manage data quality could be the solution for data quality issues on the enterprise level. Moreover, it should include systematic guidelines for planning the data quality assessment activity, extracting requirements for the data quality management, setting priorities to expedite the adaptation, identifying dimensions and metrics to ease the understanding, and visualizing these dimensions and metrics to assess the overall data quality [44].

Finally, a universal standard for describing, modeling, analyzing, measuring, testing, simulating, and building the real-world objects, products, and services remains a vision of Digital Twin. It promises to provide a platform whose standardized digital representation of real-world objects enables consistent, seamless exchange of technical information and interoperability across domains, industry silos, vertical markets, tools, and applications.

7.7 Conclusions and Outlook

As the preceding sections and use cases indicate, data quality plays an important role for the generation of Digital Twin. The generation of high-quality models is an important and integral part of digital processes today. The earlier approach of the ‘digital master’ has now evolved into a comprehensive approach of a Digital Twin. The developed approach with object recognition replaces the human hand in generation of Digital Twin providing the same level of data quality.

3D CAD models not only serve to depict the product shape geometrically, but also provide a basis for a large number of subsequent tasks and processes that are managed by the PDM system. The continuous further use and reuse of the CAD models by different users, such as CAE and CAM, can increase the effectiveness of virtual product creation and shorten the product creation time considerably [45]. As can be observed from the use cases, knowledge from downstream disciplines such as manufacturing must be implemented into data quality methods. At present, many of the original CAD models are either time-consuming post-processing by downstream

users or, in some cases, even created from scratch. Many processes in PDM are still document-oriented [46]. Therefore, the sophisticated methods and techniques of quality management must be used for the virtual products of a digital planner as well as for physical products. Likewise, every model error leads to loss of time and additional costs. Ultimately, this knowledge has led to the use of fully-fledged parametric models to build the Digital Twin.

As a future development, a more comprehensive theoretical basis is required to define quantitative metrics for complex quality requirements (e.g. higher level semantics, data quality templates, relationship to Digital Thread). A new challenge for future developments represents the increasing product complexity caused by upcoming technologies, especially in area of electric, electronics, and software [47]. As a typical supporting topic, data quality will keep the speed with the development of the subordinated systems.

References

1. Otto B, Österle H (2016) Corporate data quality. Springer, Heidelberg. <https://doi.org/10.1007/978-3-662-46806-7>
2. Haug A, Stentoft Arlbjørn J (2011) Barriers to master data quality. *J Enterp Inf Manag* 24:288–303. <https://doi.org/10.1108/17410391111122862>
3. Sha K, Zeadally S (2015) Data quality challenges in cyber-physical systems. *J Data Inf Qual* 6(2–3):8. <https://doi.org/10.1145/2740965>
4. Bondar S, Ruppert C, Stjepandić J (2014) Ensuring data quality beyond change management in virtual enterprise. *Int J Agile Syst Manage* 7(3/4):304–323. <https://doi.org/10.1504/IJASM.2014.065346>
5. Sebastian-Coleman L (2012) Measuring data quality for ongoing improvement: a data quality assessment framework. Elsevier, Amsterdam. <https://doi.org/10.1016/C2011-0-07321-0>
6. Sadiq S (2013) Handbook of data quality. Research and practice. Springer, Heidelberg. <https://doi.org/10.1007/978-3-642-36257-6>
7. Kraus M (2000) Entwicklung eines methodischen Konzeptes zur Sicherung der Datenqualität beim CAD-Datenaustausch, Diploma thesis, TU Darmstadt
8. Ofner MH (2013) Datenqualitätsmanagement aus Prozessperspektive: methoden und Modelle, PhD thesis, Universität St. Gallen
9. Yang J, Han S, Kang H, Kim J (2006) Product data quality assurance for e-manufacturing in the automotive industry. *Int J Comput Integr Manuf* 19(2):136–147. <https://doi.org/10.1080/09511920500171261>
10. Zhang R, Jayawardene V, Indulska M, Sadiq SW, Zhou X (2014) A data driven approach for discovering data quality requirements. In: Proceedings of the thirty fifth international conference on information systems: ICIS 2014
11. Pfouga A, Stjepandić J (2018) Leveraging 3D geometric knowledge in the product lifecycle based on industrial standards. *J Comput Des Eng* 5(1):54–67. <https://doi.org/10.1016/j.jcde.2017.11.002>
12. Schuh G, Rebentisch E, Riesener M, Ipers T, Tönnies C, Jank MH (2019) Data quality program management for digital shadows of products. *Procedia CIRP* 86:43–48. <https://doi.org/10.1016/j.procir.2020.01.027>
13. Boyd M (2009) Product data quality—different problem, different solutions. In: MIT information quality industry symposium, 15–17 July 2009. http://mitiq.mit.edu/IQIS/Documents/CDOIQS_200977/Papers/02_08_2D-2.pdf.

14. El-khoury J, Berezovskyi A, Nyberg A (2019) An industrial evaluation of data access techniques for the interoperability of engineering software tools. *J Ind Inf Integr* 15(2019):58–68. <https://doi.org/10.1016/j.jii.2019.04.004>
15. da Silva Serapião Leal G, Guédria W, Panetto H (2019) An ontology for interoperability assessment: a systemic approach. *J Ind Inf Integr* 16:100100. <https://doi.org/10.1016/j.jii.2019.07.001>
16. Wagner L (2020) Kein Digital Twin ohne digitale Durchgängigkeit. *Zeitschrift für den wirtschaftlichen Fabrikbetrieb*, Band 115, April 2020, pp 43–45. <https://doi.org/10.3139/104.112315>
17. Stark J (2015) *Product lifecycle management*, vol 1. Springer International Publishing, Switzerland. <https://doi.org/10.1007/978-3-319-17440-2>
18. Stark J (2019) *Product lifecycle management (vol 4): the case studies*. Springer Nature Switzerland AG. <https://doi.org/10.1007/978-3-030-16134-7>
19. Bondar S, Potjewijd L, Stjepandić J (2013) Globalized OEM and tier-1 processes at SKF. In: Stjepandić et al (eds) *Concurrent engineering approaches for sustainable product development in a multi-disciplinary environment*. Springer, London, pp 789–800. https://doi.org/10.1007/978-1-4471-4426-7_67
20. Biahmou A, Emmer C, Pfouga A, Stjepandić J (2016) Digital master as an enabler for industry 4.0. *Adv Transdiscipl Eng* 4:672–681. <https://doi.org/10.3233/978-1-61499-703-0-672>
21. Shen H, Bednarz T, Nguyen H, Feng F, Wyl T, Hoek PJ, Lo EHS (2019) Information visualisation methods and techniques: state-of-the-art and future directions. *J Ind Inf Integr* 16:100102. <https://doi.org/10.1016/j.jii.2019.07.003>
22. Emmer C, Fröhlich A, Stjepandić J (2013) Advanced engineering visualization with standardized 3D formats. In: *IFIP advances in information and communication technology*, vol 409. Springer, Berlin Heidelberg, pp 584–595. https://doi.org/10.1007/978-3-642-41501-2_58
23. Zilinski L, Sapp Nelson MR (2014) Thinking critically about data consumption: creating the data credibility checklist. In: *ASIS&T proceedings of the American society for information science and technology*, Seattle, WA, vol 51. <https://doi.org/10.1002/meet.2014.14505101107>
24. Young S, Palmer J (2007) Pedigree and confidence: issues in data credibility and reliability. In: *2007 10th international conference on information fusion*. <https://doi.org/10.1109/ICIF.2007.4408053>
25. Belkoura S, Zanin M, LaTorre A (2019) Fostering interpretability of data mining models through data perturbation. *Expert Syst Appl* 137:191–201. <https://doi.org/10.1016/j.eswa.2019.07.001>
26. ISO 8000-60 (2017) *Data quality—Part 60: data quality management: overview*
27. Lee C, Kim S, Lee J, Lee D, Caballero I (2019) Perez del Castillo R, Yan H (2019) Data quality assessment for manufacturing processes. *ICIC Express Lett Part B: Appl* 10(12):1107–1113. <https://doi.org/10.24507/icicelb.10.12.1107>
28. ISO 8000-61 (2016) *Data quality—Part 61: data quality management: process reference model*.
29. ISO 8000-62 (2018) *Data quality—Part 62: process data quality management: organizational process maturity assessment*. Appl ISO/IEC 330xx Family Stand
30. IEC 62264-1 (2013) *Enterprise-control system integration—Part 1: models and terminology*
31. Fischer T, Martin HP, Endres M, Stjepandić J, Trinkhaus O (2000) *Anwendungsorientierte Optimierung des neutralen CAD-Datenaustausches mit Schwerpunkt Genauigkeit und Toleranz*, VDA, Frankfurt
32. N.N., 4955 (2006) *Umfang und Qualität von CAD/CAM-Daten*, VDA, Frankfurt
33. N.N., ISO/PAS 26183:2006 (2005) *SASIG-product data quality guidelines for the global automotive industry*. International Organization for Standardization, Switzerland
34. Kuhn O, Liese H, Stjepandić J (2011) Methodology for knowledge-based engineering template update. In: Cavallucci D, Guio R, Cascini G (eds) *Building innovation pipelines through computer-aided innovation*, Springer, Berlin Heidelberg, pp 178–191. https://doi.org/10.1007/978-3-642-22182-8_14
35. Paukkunen L (2018) *Development of metrics and automation for product model verification*, Master thesis, Aalto University

36. González Lluch MC, Company P, Contero González MR, Camba JD, Colom J (2017) A case study on the use of model quality testing tools for the assessment of MCAD models and drawings. *Int J Eng Educ* 33(5):1643–1653
37. Denkena B, Dittrich MA, Stobrawa S, Stjepandić J (2020) Efficient generation of a digital twin using object detection for data acquisition and XML-interface for model creation. *Procedia CIRP* 93:274–279. <https://doi.org/10.1016/j.procir.2020.03.016>
38. Yang W, Fu C, Yan X, Chen Z (2020) A knowledge-based system for quality analysis in model-based design. *J Intell Manuf* 24. <https://doi.org/10.1007/s10845-020-01535-8>
39. Hirz M, Rossbacher P, Gulánová J (2017) Future trends in CAD—from the perspective of automotive industry. *Comput-Aided Des Appl* 1–8. <https://doi.org/10.1080/16864360.2017.1287675>
40. González-Lluch C, Company P, Contero M, Camba JD, Plumed R (2017) A survey on 3D CAD model quality assurance and testing tools. *Comput Aided Des* 83:64–79. <https://doi.org/10.1016/j.cad.2016.10.003>
41. Ostrosi E, Stjepandić J, Fukuda S, Kurth M (2014) Modularity: new trends for product platform strategy support in concurrent engineering. *Adv Transdiscipl Eng* 1:414–423. <https://doi.org/10.3233/978-1-61499-440-4-414>
42. Sun J, Hiekata K, Yamato H, Nakagaki N, Sugawara A (2014) Virtualization and automation of curved shell plates manufacturing plan design process for knowledge elicitation. *Int J Agile Syst Manage* 7(3/4):282–303. <https://doi.org/10.1504/IJASM.2014.065348>
43. Bondar S, Shammaa A, Stjepandić J, Tashiro K (2015) Advances in parameterized CAD feature translation. *Adv Transdiscipl Eng* 2:615–624. <https://doi.org/10.3233/978-1-61499-544-9-615>
44. Gürdür D, El-khoury J, Nyberg M (2019) Methodology for linked enterprise data quality assessment through information visualizations. *J Ind Inf Integr* 15:191–200. <https://doi.org/10.1016/j.jii.2018.11.002>
45. Stekolschik A (2007) Ein Beitrag zum ganzheitlichen Qualitätsmanagement von CAD-Modellen in der Produktentstehung, Dissertation, Ruhr-Universität Bochum
46. Lämmer L, Theiss M (2015) Product lifecycle management. In: Stjepandic J, Wognum N, Verhagen WJC (eds) *Concurrent engineering in the 21st century*. Springer International Switzerland, pp 455–490. https://doi.org/10.1007/978-3-319-13776-6_16
47. Van Der Velden C, Bil C, Xu X (2012) Adaptable methodology for automation application development. *Adv Eng Inform* 26(2):231–250. <https://doi.org/10.1016/j.aei.2012.02.007>