Available online at www.sciencedirect.com

ScienceDirect

Procedia Manufacturing 29 (2019) 583–591

Procedia MANUFACTURING

www.elsevier.com/locate/procedia

18th International Conference on Sheet Metal, SHEMET 2019

Data quality assessment for improved decision-making: a methodology for small and medium-sized enterprises

Lisa C. Günther[a,*], Eduardo Colangelo[a], Hans-Hermann Wiendahl[a], Christian Bauer[b]

[a]*Fraunhofer Institute for Manufacturing Engineering and Automation IPA, Nobelstrasse 12, 70569 Stuttgart, Germany*
[b]*Trumpf GmbH + Co. KG, Johann-Maus-Str. 2, 71254 Ditzingen, Germany*

Abstract

Industrial enterprises rely on prediction of market behavior, monitoring of performance measures, evaluation of production processes and other data analyses to support strategic and operational decisions. However, although an adequate data quality (DQ) is essential for any data analysis and several methodologies for DQ assessment exist, not all organizations consider DQ in decision-making processes. E.g., inaccurate and delayed data acquisition leads to imprecise master data and poor knowledge of machine utilization. While these aspects should influence production planning and control, current approaches to data evaluation are too complex to use them on a-day-to-day basis. In this paper, we propose a methodology that simplifies the execution of DQ evaluations and improves the understandability of its results. One of its main concerns is to make DQ assessment usable to small and medium-sized enterprises (SME). The approach takes selected, context related structured or semi-structured data as input and uses a set of generic test criteria applicable to different tasks and domains. It combines data and domain driven aspects and can be partly executed automated and without context specific domain knowledge. The results of the assessment can be summarized into quality dimensions and used for benchmarking. The methodology is validated using data from the enterprise resource planning (ERP) and manufacturing execution system (MES) of a sheet metal manufacturer covering a year of time. The particular application aims at calculating logistic key performance indicators. Based on these conditions, data requirements are defined and the available data is evaluated considering domain specific characteristics.

Keywords: Data quality assessment; Data quality control; Information quality; Benchmarking; Production planning and control

* Corresponding author. Tel.: +49-711-970-1984; fax: +49-711-970-1009.
E-mail address: lisa.charlotte.guenther@ipa.fraunhofer.de

2351-9789
10.1016/j.promfg.2019.02.114

1. Introduction

Data and knowledge about DQ are crucial for any industrial enterprise to take informed decisions. However, despite its importance, three decades of DQ research and various methodologies for DQ assessment, DQ is still not necessarily considered in everyday decisions. Two reasons can be identified: *(1)* Up to now, research has led to "fragmented and sparse results in the literature" [1] with techniques and tools still missing. *(2)* There is a gap between the techniques and measurements developed by the research side and their actual use in practice. In this article, we propose a practical approach to DQ assessment applicable to SME. It takes selected, context related data and evaluates DQ with a set of generic metrics that can be adapted to different tasks and domains. The methodology is demonstrated in context of production planning and control (PPC) in sheet metal manufacturing (SMM).

A SMM environment is typically arranged as a flow shop or job shop and consists of some main work processes: material is separated (e.g. by punching or laser cutting), formed (e.g. by bending), parts are welded if required, followed by a surface finish and an optional assembly process. However, despite the simple setup, PPC in SMM is especially challenging. Typical tasks are nesting of parts [2], scheduling [3] and tool selection [4]. At the same time, interdependencies and restrictions need to be considered. In addition to that, an increasing individualization of products leads to decreasing lot sizes which comes with new challenges for production and order planning and scheduling respectively. Examples are the influence of nesting on scheduling [5] and maximum allowed time between abrasive blasting and coating. Furthermore, processing times are often hard to predict. An example are bending operations that depend on individual springback of material [6]. As a result all PPC decisions require not only experience but need to be supported by high quality data. And as the influence of approaches like AI applications (e.g. [7]) and new technologies such as cloud manufacturing [8] grows, the need for meaningful data increases even more. But data acquisition in SMM proves to be difficult. The production environment is characterized by a high variety of products which makes master data management challenging. As operations are influenced by numerous parameters such as material type, sheet thickness and cutting edge requirements, the estimation of processing times is difficult. Batch processing with varying compilation of parts, high setup times compared to processing times and labor-intensive operations makes tracking of actual processing times even harder. With these aspects in mind – the requirement of data on the one hand and the challenges in data acquisition on the other hand – knowledge about DQ is essential for taking informed business decisions and efficiently improve the DQ.

Despite the difference between the terms information and data, no clear distinction is made between information quality (IQ) and DQ in literature (see e.g. in [9, 1]). In the remainder of the paper, the term DQ is used unless a citied publication specifically uses the term IQ. The article is organized as follows. Section 2 provides an overview of DQ research. The methodology is introduced in section 3. Section 4 details the approach and presents a use-case. The paper finishes with a summary and suggestions for future work in section 5.

2. Related work

2.1. DQ dimensions and interdependencies

DQ as presented in literature is a multi-dimensional concept [10] with the work of Wang and Strong [11] being considered as one of the first systematic approaches to identify and describe DQ dimensions [12]. Wang and Strong determine three different approaches to study DQ: the empirical, the theoretical and the intuitive approach. These result in three perspectives to define DQ dimensions: the user, the real world and the data perspective [13]. The user perspective defines DQ dimensions from users' expectations and intended use. The real-world perspective assumes that an information system (IS) represents an application domain; it derives from the theoretical approach that examines the origin of data deficiencies and thus allows definition of comprehensive set of DQ dimensions [11]. The data perspective selects quality dimensions according to the goals of the specific application and enables an objective and automatic assessment of DQ. Wang and Strong choose the empirical approach. They conduct a two stage survey and a two-phase sorting study to identify more than a hundred DQ attributes important to data consumers. These are then grouped into 20 DQ dimensions each representing a single aspect of DQ. Further notable classification schemes are those of Wand and Wang [14] who categorize 26 quality dimensions using a theoretical approach and of Naumann [15] who selects 22 quality criteria in four categories with an intuitive approach supported by literature research. Price and Shanks state that quality criteria should not be based on a single approach but be "both theoretically grounded and practical" [16] and propose a framework with 16 DQ dimensions. Batini et al. compare several DQ classifications and conclude that "no general agreement exists either on which set of dimensions defines the quality of data, or on the exact meaning of each dimension" [17]. Still some common quality dimensions can be identified. Among those are:

- *Accuracy* evaluates "the extend to which data is correct, reliable and certified free of error" [11] and can be calculated as the "quotient of the number of correct values in a source and the overall number of values" [15].
- *Completeness* takes into account if a data set includes all data necessary to "represent every meaningful state of the represented real world system" [14] and should consider why a value is missing [17].
- *Consistency* refers to "the violation of semantic rules defined over a set of data items" [17] and "the extent to which data are always presented in the same format and are compatible with previous data" [11].
- *Timeliness* is influenced by system volatility (rate of change), currency (time of data update) and the time the data is used [14] and described e.g. as "the extent to which the age of the data is appropriate for the task at hand" [11] or "the average age of data in a source" [15].
- *Relevancy* evaluates whether available data types are pertinent to the intended use of the data [16] and if "the provided information satisfies the users need" [15].

Even though quality measures are still not consistently defined, a number of publications analyze interdependencies between quality dimensions. Sadeghi and Clayton [18] and Han and Venkatasubramanian [19] address the trade-off between timeliness and accuracy. Ballou and Pazer [20] model the relationship between completeness and consistency. Panahy et al. [21] qualitatively describe the dependencies between four measurable dimensions and their relation to the DQ improvement process: Accuracy, completeness and consistency are considered as independent variables influencing timeliness. The improvement process is a control variable. The framework is validated with an empirical study among users of IS in different organizational positions [22]. Gackowski [9] analyses the logical interdependence between DQ attributes and ranks them hierarchically. Therefore he distinguishes whether attributes influence business decisions directly or indirectly and qualitatively or quantitative and between mandatory and desirable attributes. While the proposed framework aims at providing an economical and practical approach to support evaluation of DQ in the context of business decision making without being oversimplifying, the paper focuses on a qualitative explanations and lacks an analytical example. Amics, Barone and Batini [23] present a purely data-driven approach. The proposed analytical framework describes how often quality dimensions are affected by erroneous observations. The dependencies between quality measures are then calculated via entropic values and equations. No a priori-knowledge about the data is needed. [24] extends the framework to explore multi-variate dependencies between DQ dimensions through a Bayesian Network. Both frameworks are successfully tested on real datasets that are evaluated in terms of syntactic accuracy, completeness, timeliness and consistency.

2.2. Methodologies to assess and merge DQ

Besides definition and characterization of quality measures numerous publications address the assessment and improvement of DQ. The most general approach involves the following phases: *(1)* state reconstruction to collect contextual information, e.g. on data structure and organizational processes, *(2)* DQ assessment and *(3)* improvements, either data-driven or process-driven [17]. Below some major DQ assessment strategies are introduced. Methodologies related to the cooperative context are chosen whenever possible. A more detailed introduction can be found in [1].

Stvilla et al. [25] develop a general IQ assessment framework. They use sources of IQ problems, such as mapping and context changes, to develop a taxonomy IQ dimensions. The framework is supplemented by 41 general metric functions, that can also be used to develop context-specific IQ metrics. Even and Shankaranarayanan [26] highlight the importance of considering the utility of the data when evaluating its quality. They argue that on the one hand different datasets may have different significance in the same context while on the other hand the utility of same dataset may vary in a different context. Thus, they suggest to evaluate the amount of defects as well as to what extend quality defects reduce the value of the data within a defined context and task.

Lee et al. introduce a methodology that aims at providing "a practical IQ tool to organizations" [27] that helps monitor IQ improvements over time. A questionnaire is used to assess several IQ dimensions relevant for information users. The results are interpreted with two quantitative techniques to detect problem areas and prioritize IQ improvement tasks. While the approach ensures integrating contextual and subjective aspects into the assessment Gackowski criticizes that the method focuses on preferences of information consumers not on business purposes [9]. Gustavsson and Wänström [28] propose another approach as they note "a lack of studies identifying the dimensions needed to describe IQ deficiencies in the manufacturing planning and control specific setting". They select ten DQ dimensions and illustrate them by conducting an IQ assessment in three different enterprises examining IQ at different planning levels. In each case, the assessment heavily relies on interviews with top and middle managers and PPC staff supplemented by further data sources such as ERP data. Pipino et al. [10] propose to combine subjective DQ assessment via questionnaire with an objective assessment. The latter one is conducted by calculating task-dependent and task-independent metrics that reflect the DQ with and without considering contextual knowledge such as business rules and database constraints. Discrepancies between the results of the subjective and objective DQ indicate how to improve DQ. [29, 30] present an architecture designed to help managing DQ in cooperative information systems and to avoid distribution of low quality data through the system. Instead of sporadic DQ monitoring and long term improvement the proposed framework aims at ad-hoc evaluation of data and monitoring of changes in DQ.

Despite the variety of approaches, few publications detail the development of scores that combine multiple DQ measures. Kennedy et al. [31] summarize the quality of a data element in a vector. Each element of the vector represents a specific DQ dimension expressed with a number between 1 and 5. Several methods are presented to aggregate a DQ vector into a single DQ indicator. Naumann [15] focusses on evaluating the IQ of query answers retrieved from integrated IS, mainly the world wide web. Within this application he considers the problems of combining quality criteria with different ranges, units, scales and weighting into an aggregated score and the challenge of merging the scores of multiple data resources.

2.3. Conclusion

During the last decades of DQ research DQ dimensions have been described from user and data perspective and summarized in various categories, their interdependencies have been investigated qualitatively and quantitatively and assessment methodologies have been developed that are generic or context specific and practical or theoretical [13]. Today DQ research is still ongoing. A growing field adapts and extends existing ideas to the context of big data and web data. This puts focus on other forms of data representations [32], requirements of new analysis methods [33] and different DQ dimensions [34]. Further current research addresses challenges of specific environments (see e.g. [35]). Still there are aspects that hinder the consideration of DQ assessment in day-to-day business decisions:

- Much DQ research focusses on development and classification of DQ dimensions. This does not only apply to DQ research in itself, but also to papers introducing methodologies. Many of which define their own framework of DQ dimensions (e.g. [25], [27], [30]). This makes e.g. selection of an approach for a specific application more difficult.
- The execution of DQ analysis and the interpretation of the results often requires background knowledge of definitions and methods (e.g. [21], [23], [25]). It cannot necessarily be done by a data consumer.
- DQ assessment approaches that address a cooperative context and consider experiences and expectations of data users usually include a questionnaire (e.g. [10], [27], [28]). Due to that, the procedure is time consuming.
- In general, many methodologies aim at occasional DQ assessment and long-term improvement of DQ. An exception – an architecture designed for ad-hoc evaluation of data – is presented in [29] and [30].

3. Introduction to methodology

With regard to the categories presented in [16] the proposed methodology can be outlined as follows: It involves the first and second phases of the most general approach, i.e. it starts with a state reconstruction phase, followed by the DQ assessment regarding completeness, relevance, accuracy, timeliness, consistency and plausibility. A data improvement phase is not included. Costs associated with DQ deficiencies are not considered. The exemplary DQ assessment below is applied to structured data but the approach can be extended to semi-structured data. The methodology is not restricted to a specific IS, though its application is limited in some cases, e.g. a web-based IS.

Table 1: Test criteria for characterization of datasets and related quality dimensions.

Test category	Perspective	Evaluation object	Examples of test criteria	DQ dimension
Individual tests	Objective	Character sequences	Empty cells, 0-values, null values	Completeness
Form related tests	Both objective and subjective	Characteristics of dataset	'Round' values (numerical, time delta, date time)	Accuracy
		Distribution of dataset	Character patterns, outliers, near median values	Relevance, Timeliness
Content related test	Subjective	Content	Default values, values identical to those of another dataset	Consistency, Plausibility

The basic idea is to start DQ assessment by describing data anomalies. In contrast, other methodologies begin by definition and classification of DQ dimensions. I.e., we suggest to ask "how can data be characterized?" instead of "what causes DQ deficiencies?" or "how should data look like to meet the requirements?" We propose 20 preliminary test criteria to describe data features. These are grouped into the following three categories (see also **Error! Reference source not found.**):

- *Individual tests* evaluate the data of a given attribute by testing each individual value against a specific condition unrelated to further attributes or values. These are e.g. tests to check if cell values are null or empty.
- *Form related tests* are comparative tests that describe the characteristics and distribution of values. The data of an attribute is evaluated with regard to a condition that *(1)* depends on other values of the same attribute and that *(2)* needs some specification. Examples are:
  (1) Outlier detection depending on median and the absolute distance to the median; identification of values that are noticeably 'round' depending on – in case of numeric values – the total range of values; finding patterns where patterns are sequences frequently occurring elements within the dataset
  (2) An indicator when a value is considered as an outlier; the minimal length of a sequence of elements that is considered as a pattern; a threshold when a value is considered as "occurring frequently"
- *Content related tests* are another type of comparative tests. These tests assess the data of a given attribute with regard to additional data. Examples for additional data are default values or the values of another data attribute.

The categories reflect the evolution from objective to subjective assessment criteria. Individual test criteria are purely data-driven and independent of context. Form related tests are adapted to a specific context and affected by experience and expectations of data users. Definition and assessment of content related test relies on the context knowledge of data users and domain experts. Each tests returns a value in the range of *[1, 0]* that indicates the ratio of attribute values that match the given condition.

The methodology consists of five main steps (Fig. 1). First a specific usage context or data dependent task is defined. Related data sources are identified and a data model is developed. General context knowledge and the predefined test criteria add further information. Based on the specific use-case, data attributes and test criteria are selected. They form the basis for the definition of a set of data quality metrics. After this one-off preparation phase, that combines domain- and data-driven input as well as use-case specific and general information, the actual DQ assessment takes place. The characteristics of given input data are automatically calculated using the metrics and summarized into DQ dimensions. The results can be used for the data analysis task at hand, for tracking DQ over time, for benchmarking or as a starting point for a data improvement process.

4. Assessment of logistics data in sheet metal processing

To illustrate the method we examine a sheet metal manufacturer who calculates key figures and logistic operating curves to support production control. As we know from experience, a common DQ problem in SMM is caused by imprecise or missing process completion data, which results in insufficient knowledge about actual operation times. I.e. as a use-case, we focus on assessment of operation time data. The context knowledge consists of information on

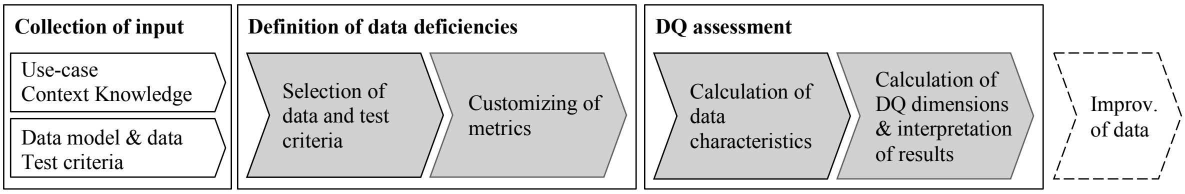

Fig. 1. Main steps of the DQ assessment methodology.

the production system and knowledge in the domain of PPC. Since SMM is a single item production, we choose the funnel model (Fig. 2) as an appropriate basis to visualize and control logistic processes within the production [36]. The funnel model describes the throughput of a load center, e.g. one or more workstations, in terms of input, work in process and output. Multiple planning and feedback measures derive from this approach. These are quantity measures like the number of rejected parts, time specifications like actual processing or planned setup time and events like the finish time of a job. The data used for PPC is retrieved from an ERP system. For performing the DQ assessment, the theoretical and the ERP data model are matched. The test criteria, developed for describing data characteristics, are not modified in this phase.

After compiling the input, we select the data and test criteria relevant for the intended analysis. As the funnel model divides operation times into setup and processing times, we use both parameters for the assessment. As additional information, the time units corresponding to each value are extracted from the database. Since a common source for errors results from transferring set values to actual values, we also select set values from the database to compare both parameters. Due to the data types of the selected parameters and the properties of the database, eight of the 20 test-criteria are chosen for DQ assessment. E.g., no tests addressing date or time values are conducted and also testing for null or default values is not necessary for the selected dataset. Furthermore, processing and setup times are derived from timestamps of single events and are stored with different time units, therefore checking for patterns and negative values is not required either. Data and test criteria are used to define a set of metrics. We choose to perform all selected tests on both setup and processing times and specify the test criteria if necessary. Among the tests that are conducted are: checks if values exist, checks if values are noticeably round (10 minutes, 20 minutes etc.), checks for outliers, checks for correctness of the given time unit and checks to compare set and actual values.

The calculation is performed on a dataset covering a year of time. To investigate how DQ changes over time, we split the data into three-month periods, each corresponding to 17,000 to 25,000 operations. The results are transformed into DQ dimensions: Test results related to completeness are summed, results related to relevance are combined via weighted averaging. Accuracy, consistency and plausibility in this case derive directly from single test criteria. Timeliness is not evaluated since the selected data does not contain any date or time values. All quality dimension values are displayed in Fig. 3. The DQ dimensions arranged from left to right demand an increasing context and domain knowledge for specification of test criteria and interpretation of results. E.g., completeness is easily understood as the ratio of available data compared to the total amount of required data. In contrast, plausibility is in this case evaluated by comparing predefined set times and actual times. In general, DQ of processing times is higher than that of setup times. The difference is especially noticeable regarding plausibility and should be analyzed in more detail.

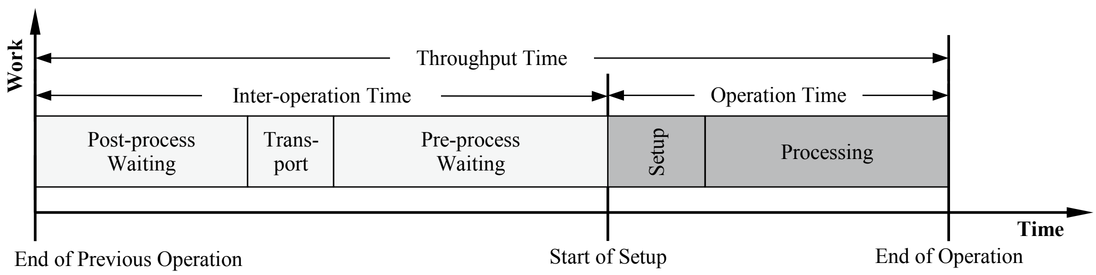

Fig. 2. Throughput element (operation or production order) as described in the funnel model (according to [36]).

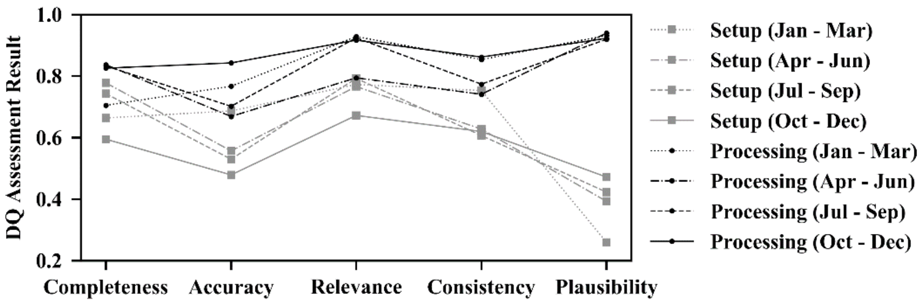

Fig. 3. Results of assessing the DQ of operation times grouped into setup and processing times and quarters of the year.

With regard to a concrete PPC task the planner should keep in mind that knowledge about actual setup times is limited. For about 30 % of all operations the actual setup time is not known; 60 % of the available setup times might not be plausible. Processing times on the other hand have a high relevance, consistency and plausibility. Completeness and accuracy can still be improved. The development of processing time DQ measures shows no positive or negative trend during the year, while DQ measures of setup times tend to decrease.

5. Conclusion and future work

In this paper, we propose a general-purpose DQ assessment method that is adapted to a specific context and use-cases. Therefore, it is valid for different areas of application. At the same time data assessment takes a minimum effort, as only data relevant for the use-case is considered. Since the calculation of data characteristics is fully automated DQ assessment can be integrated into regular data analysis tasks and decision-making. Furthermore, it is possible to extend the provided set of test-criteria or to restrict them to individual and/or form related tests if limited context knowledge is available. The method was successfully applied to assess DQ of ERP and MES data of several sheet metal manufacturers. Section 4 presents some of the results.

The usage of data-driven approaches for areas such as quality and process optimization in SMM is being widely researched. For them, the possession of correct data plays undoubtedly an important role. The focus of the presented DQ assessment method is to be able to perform logistic analysis, which allows improvements in the overall equipment efficiency, for example by means of reduction of downtimes or increase of the utilization of workstations. These analyses present the problem that most of the data required, like timestamps and times, is created manually (especially within SME). The assessment of the data quality of several manufacturers (SME) revealed that the data quality for logistic analysis is low in all of them. Many reasons can be identified, such as the flexibility of the manufacturing systems involved, the possession of different technologies and the lack of correct feedback systems. Even the results presented section 4, which correspond to a relatively big manufacturer of machine tools with congruent technologies (high-end machines) and manufacturing methods, show that the problems in DQ for logistic purposes is still an issue. The DQ assessment presents therefore a way to focus on improving the data acquisition of high relevance and impact (given the limited resource of most enterprises). E.g., the addition of a few control points, like simple sensors, can already enable of performance of important analysis (e.g. of those based on timestamps).

However, the method is not without limitations. In context of the illustrative example, it is noticeable that the logistics data analysis tasks are only allowed a limited complexity. E.g., up to now it might be necessary to break a query into several steps to differentiate between different types of work stations, or if it involves multi-stage jobs that are processed on welding or assembly stations. Furthermore, the aggregation of results into quality dimensions leads to a loss of information. Methods to improve the traceability of DQ dimensions, e.g. by visualizations of results, should be developed. Future work might also address a guideline for the preparation and improvement phase and the application to different contexts and domains.

References

[1] C. Batini, M. Scannapieco, Data and information quality, Springer International Publishing, (2016).
[2] S. Q. Xie, G. G. Wang, Y. Liu, Nesting of two-dimensional irregular parts: an integrated approach, International Journal of Computer Integrated Manufacturing, 20.8 (2007) 741-756.
[3] P. Helo, D. Phuong, Y. Hao, Cloud manufacturing–Scheduling as a service for sheet metal manufacturing, Computers & Operations Research, (2018).
[4] T. Giannakakis, G. C. Vosniakos, Sheet metal cutting and piercing operations planning and tools configuration by an expert system, The International Journal of Advanced Manufacturing Technology, 36.7-8 (2008) 658-670.
[5] T. Sakaguchi, H. Ohtani, Y. Shimizu, Genetic algorithm based nesting method with considering schedule for sheet metal processing, Transactions of the Institute of Systems, Control and Information Engineers, 28.3 (2015) 99-106.
[6] S. K. Panthi, N. Ramakrishnan, K. K. Pathak, J. S. Chouhan, An analysis of springback in sheet metal bending using finite element method (FEM), Journal of Materials Processing Technology, 186.1-3 (2007) 120-124.
[7] S. Kumar, H. M. Hussein, A. Hussein, eds., AI Applications in Sheet Metal Forming, Springer Singapore (2017).
[8] P. Helo, Y. Hao, Cloud manufacturing system for sheet metal processing, Production Planning & Control, 28.6-8 (2017) 524-537.
[9] Z. J. Gackowski, Q. U. D. Imensions, A. P. U. Ocused, V. I. E. W. On, Logical Interdependence of Some Attributes of Data/Information Quality, IQ, (2004) 126-140.
[10] L. L. Pipino, Y. W. Lee, R. Y. Wang, Data quality assessment, Communications of the ACM, 45.4 (2002) 211-218.
[11] R. Y. Wang, D. M. Strong, Beyond accuracy: What data quality means to data consumers, Journal of management information systems, 12.4 (1996) 5-33.
[12] S. Madnick, R. Y. Wang, Y. W. Lee, H. Zhu, Overview and framework for data and information quality research Journal of Data and Information Quality (JDIQ), 1.1 (2009) 2.
[13] M. Ge, M. Helfert, A review of information quality research—develop a research agenda, Paper presented at the International Conference on Information Quality 2007, (2007).
[14] Y. Wand, R. Y. Wang, Anchoring data quality dimensions in ontological foundations, Communications of the ACM, 39.11 (1996) 86-95.
[15] F. Naumann, Quality-driven query answering for integrated information systems, Vol. 2261, Springer, (2003).
[16] R. Price, G. Shanks, A semiotic information quality framework: development and comparative analysis, Enacting Research Methods in Information Systems, Palgrave Macmillan, Cham, (2016) 219-250.
[17] C. Batini, C. Cappiello, C. Francalanci, A. Maurino, Methodologies for data quality assessment and improvement, ACM computing surveys (CSUR), 41.3 (2009) 16.
[18] A. Sadeghi, R. Clayton, The quality vs. timeliness tradeoffs in the BLS ES-202 administrative statistics, Federal Committee on Statistical Methodology, (2002) 1-7.
[19] Q. Han, N. Venkatasubramanian, Addressing timeliness/accuracy/cost tradeoffs in information collection for dynamic environments, Real-Time Systems Symposium, 2003, RTSS 2003, 24th IEEE, IEEE, (2003).
[20] D. P. Ballou, H. L. Pazer, Modeling completeness versus consistency tradeoffs in information decision contexts, IEEE Transactions on Knowledge and Data Engineering, 15.1 (2003) 240-243.
[21] P. H. S. Panahy, F. Sidi, L. S. Affendey, M. A. Jabar, H. Ibrahim, A. Mustapha, A framework to construct data quality dimensions relationships, Indian Journal of Science and Technology, 6.5 (2013) 4422-4431.
[22] P. H. S. Panahy, F. Sidi, L. S. Affendey, M. A. Jabar, H. Ibrahim, A. Mustapha, Discovering dependencies among data quality dimensions: A validation of instrument, Journal of Applied Sciences, 13.1 (2013) 95-102.
[23] F. De Amicis, D. Barone, C. Batini, An Analytical Framework to Analyze Dependencies Among Data Quality Dimensions, ICIQ, (2006).
[24] D. Barone, F. Stella, C. Batini, Dependency discovery in data quality, International Conference on Advanced Information Systems Engineering, Springer, Berlin, Heidelberg, (2010).
[25] B. Stvilia, L. Gasser, M. B. Twidale, L. C. Smith, A framework for information quality assessment, Journal of the American society for information science and technology, 58.12 (2007) 1720-1733.
[26] A. Even, G. Shankaranarayanan, Understanding Impartial Versus Utility-Driven Quality Assessment In Large Datasets, ICIQ, (2007) 265-279.
[27] Y. W. Lee, D. M. Strong, B. K. Kahn, R. Y. Wang, AIMQ: a methodology for information quality assessment, Information & management, 40.2 (2002) 133-146.
[28] M. Gustavsson, C. Wänström, Assessing information quality in manufacturing planning and control processes, International Journal of Quality & Reliability Management, 26.4 (2009) 325-340.
[29] M. Mecella, M. Scannapieco, A. Virgillito, R. Baldoni, T. Catarci, C. Batini, Managing data quality in cooperative information systems, OTM Confederated International Conferences "On the Move to Meaningful Internet Systems", Springer, Berlin, Heidelberg, (2002) 486-502.
[30] C. Marchetti, M. Mecella, M. Scannapieco, A. Virgillito, R. Baldoni, Data quality notification in cooperative information systems, Proceedings of the First International Workshop on Data Quality in Cooperative Information Systems, Siena, Italy. 2003.
[31] D. J. Kennedy, D. C. Montgomery, B. H. Quay, Data quality, The International Journal of Life Cycle Assessment, 1.4 (1996)199-207.
[32] Färber, Michael, et al. "Linked data quality of dbpedia, freebase, opencyc, wikidata, and yago." Semantic Web Preprint (2016): 1-53.
[33] D. Ardagna, C. Cappiello, W. Samá, M. Vitali, Context-aware data quality assessment for big data, Future Generation Computer Systems 89 (2018) 548-562.

[34] A. Zaveri, A. Rula, A. Maurino, R. Pietrobon, J. Lehmann, S. Auer, Quality assessment for linked data: A survey. Semantic Web, 7.1 (2016) 63-93.
[35] L. Houston, Y. Probst, A. Martin, Assessing data quality and the variability of source data verification auditing methods in clinical research settings, Journal of biomedical informatics, 83 (2018) 25-32.
[36] H. H. Wiendahl, Auftragsmanagement der industriellen Produktion: Grundlagen, Konfiguration, Einführung, Springer-Verlag, (2011).