Journal of Theoretical and Applied Information Technology
31st August 2017. Vol.95. No.16

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

THE STATE OF DATA QUALITY ARTS IN (TECHNICAL) SERVICE REPORTING

[1]A. T. KHALILIAN, [2]OTHMAN IBRAHIM
[1] Faculty of Computing, Universiti Teknologi Malaysia,UTM Skudai, 81310 Johor, Malaysia
[2] Faculty of Computing, Universiti Teknologi Malaysia,UTM Skudai, 81310 Johor, Malaysia
E-mail: [1] A.T.Khalilian@hotmail.com, [2] othmanibrahim@utm.my

ABSTRACT

Service Level Management (SLM) in IT Service Management (ITSM) contains Technical Service Reports (TSRs) to report Service Quality (SQ) based on the Service Level Agreement (SLA). However, producing TSRs for a large enterprise has Data Quality (DQ) challenges. The source of technical metrics in TSRs comes from large, unverified and non-normalized system-generated events and logs in a large enterprise environment. Moreover, configuration items and service information meta-data that are essential for producing these SLM reports are facing DQ problems. These challenges lead to low reports' Data Quality (DQ) that destroy customer's trust and management visibility, which leads to financial penalties and SQ issues. In order to improve the TSRs' DQ and consequently improving the SQ and reducing the risks of financial penalties, researchers need to know the limitations and definitions of DQ for TSRs, and this is not feasible, except for having a comprehensive overview of DQ dimensions and its processes. This paper provides a statement on the situation of the DQ in existing literature by having eyes on technical service reporting issues.

Keywords: *Data Quality, Technical Service Report, Service Level Management, DQ, TSR, SLM*

1. INTRODUCTION

IT Service Management (ITSM) is a concept under IT Operation Management (ITOM). ITSM is for implementing and managing qualified IT services to meet the needs of a business or organization by an IT Service provider through an appropriate mix of information technologies, process and people [1]. Based on Mora et al. (2014), there are seven international IT Service Management (ITSM) frameworks which are Information Technology Infrastructure Library (ITIL) (include v2 and v3), ISO 20000, Control Objectives for Information and Related Technology (COBIT) 4.0, Capability Maturity Model Integration for Services (CMMI-SVC), Microsoft Operations Framework (MOF) 4.0, and IBM Tivoli Unified Process (ITUP) [2].

Service reporting that is context of this literature review and is a part of service delivery processes beside Service Level Management (SLM). Service reporting is a process to define, document, produce and use Service Reports (SRs) [3, 4]. In fact, SRs must be documented and agreed by both service provider and interested parties to be used by service provider to take action based on SRs findings and Service Level Agreement (SLA). SRs are for both management and customer that must include performance against target, relevant information about significant events and all measurable aspect of service for both current and historical analysis. These measureable aspects could process success rate or service end result. In fact, Technical Service Reports (TSRs) defined and appeared when customer requires a measureable performance of technical service result based on system- generated result such as successful backup rate or storage service up-time.

Service Level Agreements (SLAs) [5] include the contracted key elements to describe the full success of delivered services and specify the desired metrics to examine the effectiveness of service activities, functions and processes [6]. IT Service Management (ITSM) and Information Technology Infrastructure Library (ITIL) as a famous major implemented framework in ITSM [7] define the service reporting as the main process of SLA monitoring and the one which is responsible to provide Key Performance Indicator (KPI) reports for Service Level Management (SLM). Many IT service organizations consider the measurement of IT service management processes, especially service support processes, as a considerable challenge [8].

Some companies outsource their IT services to a third party organization called IT Service Provider (ITSP) and they have to provide service reporting based on SLA. However, SLA monitoring is a critical issue in ITSM. In ITSM, SLA contains financial penalties for some defined target metric in service reports with the name of Service Level Target (SLT). The SLT is a part of service reporting, and its quality is the concern of both service provider and customer. Quality of these reports are becoming crucial when metrics have not defined based on a simple process or system output and also when it requires lots of data gathering, verification and calculation.

Service Reports that are based on technical metrics definition require system-generated logs to be produced, collected and cleansed. Technical metrics are a type of performance metrics in service reporting that are based on the end result of service which requires system-generated event logs from IT infrastructure for its calculation like "percent of backup success" and are not based on IT process like "percent of successful change". Reports that contain these explained technical metrics define as Technical Service Reports (TSRs). Generating TSRs are more complex and costly because system-generated logs need experts' verification, data cleansing and complex processing to fit in all business rules and condition defined in SLA. In fact, an expert must exclude testing issues or any kind of issues which is not genuine based on SLA definitions or exclude log noises that are not real service results. These verifications are costly and time consuming [9] which make the report delivery and report monitoring out of expected time. However, Operations and customers need reports in a right-time that be believable (a data quality dimension). An acceptable report in service reporting is a report which shows all customer services with correct status and matched to real implementations and SLA which represent the final verified figure of metrics [10]. These difficulties mostly lead to manual data cleansing logs and reports data which are time consuming and lead to late delivery report. The process of data cleansing and data quality improvement has lots of open problems and highly is domain related and needs to be explored for each specific context [11]. The manual reports mostly could be based on simple system-generated events and uptimes which mixed by testing and valid events [9]. These all challenges lead to manual or semi manual generation of technical metrics for SLM purpose and the low quality brings doubts and disputes of customers.

In addition to system-generated logs, there are other Data Sources (DSs) that must be used in report generations such as Asset Management System (AMS) and Configuration Management Database (CMDB). AMS is a system on top of the Asset Management (AM) process that manages activities or processes of tracking and reporting the properties, value and ownership of assets throughout their lifecycle [12]. Besides, CMDB is a database that is used to store configuration records and their attributes throughout their lifecycle [12]. AMS and CMDB contain some meta-data such as Configuration Items (CIs), Service Catalogues Information and some service definitions, rules and categories. CI is any component or service asset that is required to be managed and stored in order to deliver an IT service [12]. Besides, service catalogue information or in abbreviation Service Information (SI) is structured information about live IT services [12]. However, completeness and accuracy of these meta-data and DSs are depend on maturity of their process and their adaption in the organization. In fact, process design deficiencies like "Incomplete representation", "Ambiguous representation" and Operation deficiencies would lead actual data in reality that are required for generation of reports to be missed [13]. Thus, because of direct dependency between Service Reports (SRs) and these data sources, Data Sources' Data Quality (DSDQ) issues have direct impact on Reports' Data Quality (RDQ) issues and become part of problem.

Although, some researches have begun to examine data quality (DQ) and Information Quality (IQ), only few researches has been paid attention to specific area of data quality in reporting of IT services. Many researches discuss data quality for Information systems, EIP, data warehouse and decision systems when data sources are other information systems or human-generated sources. However, no attention has been paid for system-generated data in complex IT services. Thus, it remains to be seen how SLM reports' data quality in this context can be improved and so, Technical Service Reports (TSRs) Data Quality (DQ) is the gap. To be clearer on the differences between terms of Data Quality (DQ), Information Quality (IQ) and Report Quality (RQ), researcher followed Madnick (2009) and consider DQ and IQ in a same scope, opposite others who define DQ to be more on technical issues and IQ to be non-technical issues. Report quality in this research considered as overall quality of Reports and not only data of reports that is more focuses on representation quality and considered out of this research scope.

2. DATA QUALITY

The understanding and availability of rules to interrelate and validate the data elements is a definition for data quality and it represents a substantial project risk. Any problem in data quality may completely or largely unfit data for use [14]. If data are of poor quality, decisions are likely to be unsound [15]. Although there are many researchers that they have examined data quality (DQ) and Information Quality (IQ) [13, 16-20], there are not enough attentions paid to specific areas of data quality in reporting of data services. Madnick et al. (2009) claims that many researchers discuss about data quality for Information systems, EIP, data warehouse and decision systems when data sources are other information systems or human-generated sources. However, no enough attention has been paid for system-generated data in complex IT services. Thus, it is important to know how SLM reports' data quality in this context can be improved and consequently, Data Quality (DQ) in Technical Service Reports (TSRs) remains as the considerable gap.

Researchers followed Madnick to clarify the differences between terms of Data Quality (DQ), Information Quality (IQ) and Report Quality (RQ). He considered DQ and IQ in a same scope, opposite others who define DQ to be more on technical issues and IQ to be non-technical issues [18].

Data Quality is a complex concept and its definition is not straightforward [21]. Based on Orr (1998) Feedback Control System (FCS) model view, data quality is "the measure of the agreement between the data views presented by an information system and that same data in the real world" [22]. As suggested by Redman (1998), poor data quality can jeopardize the effectiveness of an organization's tactics and strategies [23]. Poor data quality can be a factor leading to serious problems [24]. The impact of data quality and information about data quality on decision making has been investigated in several studies [17, 25-29]. From the point of view to assess the "fitness for use" of data, data quality initiatives are critical for an organization's use of IT to support its operations and competitiveness. Organizations have begun to move from reactive to proactive ways of managing the quality of their data [18].

The importance of data quality in data mining claimed in many publications [30-32] and they have presented the influences of data quality on the validity of the results. All conclude the interpretations of processes needs the ensured data quality and accuracy.

There are four major categories of DQ/IQ introduced in the literature; Data Quality Impact, Database-Related Technical Solutions for Data Quality, Data Quality in the Context of Computer Science and Information Technology and Data Quality in Curation [18]. In this paper DQ and IQ considered as a whole similar context and discussed more from their quality dimensions when presented as a report. Although some framework and tools cover IQ concept from angels [33] which are more useful for assessment of IQ in organization level and not from reports perspective, this paper focused on used-based DQ/IQ improvement method for purpose of service reporting. The aim of this study was to review the accessible literature.

2.1 Quality Dimensions

If the data quality in a specific context is poor, questionable or unknown, it will be less valuable to be a tool for improving the quality of that specific context. Researches need to be confident of the quality of the source data. To investigate quality improvement, it is required to find all aspects of quality itself that named Quality Dimensions (QDs). In order to define dimensions of the data quality, it is important to know the specific viewpoint and the philosophy of how to see quality. In many publications, data quality defined as a term to show how well data satisfies data consumers. According to this definition a broader conceptualization of data quality is achievable by concentrating on perception of data consumers about quality. It relies on data consumers' perception more than perceptions of information systems professionals which is limited to intrinsic levels and accuracy dimensions. In this context, data becomes like a product and users judge how fit it is. So, it is hard to say that data quality has the same meaning for different users. In fact, each data consumer requires the used data to fulfil a certain criterion which he presumes essential for his own tasks at hand. These criteria or aspects or attributes of DQ are known as DQ Dimensions (e.g. Accuracy, Timeliness, Precision, Completeness, Reliability and Error recovery [34-36].

Although there are many publications presented in the literature that they are dealing with the measurement of data quality dimensions, they do not specify the methods used and most of them deal with the improvement of data quality with no specific attention to the targeted dimension [37].

Generally, among datasets, the classification of data quality considerably relies on the type and intended use of the dataset [37]. According to some publications, data quality is based on the "fitness for purpose" and has six dimensions; relevance,

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

accuracy and reliability, timeliness, accessibility, interpretability and coherence [38] while the other publication claimed seven dimensions for the quality of data which are valid, complete, consistent, unique, timely, accurate and precise [39]. Another publication by David Loshin (2011) stated eight dimensions to monitor the performance of data quality. According to the publication, data quality dimension includes uniqueness, accuracy, consistency, completeness, timeliness, currency, conformance and referential integrity [40].

There are 120 different names presented in [37] for dimensions of data quality extracted from 69 publications. They are categorized in four groups; accuracy, completeness, capture and others. An iterative approach by Wang and Strong to develop a framework of data quality, gathered 15 dimensions of quality in four categories; intrinsic, contextual, representational and accessibility [41].

According to Madnick et al. (2009), based on consumers' perception conceptualization and achieving from 159 dimensions, there are four dimensions that have been emphasized most frequently in publications; accuracy, completeness, consistency, and timeliness (

Table 1).

According to Madnick (2011) quality dimension comparison, all required customer attribute and aspect of data quality required based on initial case investigation exist in Wang Strong research (1996) and not in other models. There are four main categories have been identified by them as a conceptual framework of data quality describing in Fig.1. The framework and focus of Wang and Strong (1996) are on intrinsic DQ, attribute believability and accuracy as well as contextual DQ more on Completeness and Timeliness.

Quality dimensions have correlations together. In some cases, putting focus on one dimension causes effect on others however sometimes trade-offs may be done between timeliness and a dimension among accuracy, completeness, and consistency. Because, having accurate (or complete or consistent) data may require time, thus timeliness is in oppose [21]. This trade-off could be acceptable if in context, quality in one dimension still remain in acceptable amount to consumer.

Similar to case problems which are looking for reports' accuracy, Wang and Strong (1996) did a comprehensive questionnaire around data quality dimensions which data consumer require. The results of that research led to develop a model that captures scopes of data quality which are important for data consumers other than only accuracy (Fig.1). If the area of measurement and quality improvement adapts from Wang and Strong framework (1996), the quality of reports will divide to four main categories and 15 sub-categories.

Intrinsic data quality dimensions refer to "the extent to which data values are in conformance with the actual or true values" [41]. These four dimensions are;

- Believability: The extent to which data are accepted or regarded as true, real, and credible.
- Accuracy: The extent to which data are correct, reliable, and certified free of error.
- Objectivity: The extent to which data are unbiased (unprejudiced) and impartial.
- Reputation: The extent to which data are trusted or highly regarded in terms of their source or content.

Conceptual dimensions are the next category that refer to "the extent to which data are applicable to or pertain to the task of the data user" [41]. These five dimensions are:

- Value-added: The extent to which data are beneficial and provide advantages from their use.
- Relevancy: The extent to which data are applicable and helpful for the task at hand.
- Timeliness: The extent to which the age of the data is appropriate for the task at hand.
- Completeness: The extent to which data are of sufficient breadth, depth, and scope for the task at hand.
- Appropriate Amount of Data: The extent to which the quantity or volume of available data is appropriate.

Representational dimensions are third category that refer to "the extent to which data are presented in an intelligible and clear manner" [41]. These four dimensions are:

- Interpretability: The extent to which data are in appropriate language and units and the data definitions are clear.
- Ease of Understanding: The extent to which data are clear without ambuity and easily comprehended.
- Representational Consistency: The extent to which data are always presented in the same format and are compatible with previous data.

- Concise Representation: The extent to which data are compactly represented without being overwhelming (i.e., brief in presentation, yet complete and to the point).

Accessibility dimensions are the last category that refer to "the extent to which data are available or obtainable" [41]. These two dimensions are:

- Accessibility: The extent to which data are available or easily and quickly retrievable.
- Access Security: The extent to which access to data can be restricted and hence kept secure.

Based on objectives of each research, must focus on some of data quality dimensions. These Quality Dimensions (QDs) have been used as part of evaluation and measure of data quality improvement. Main solid usage of these QDs could be in evolutionary questionnaires that collects practitioners' evaluation of reports data quality improvement in the end of each research's cycle. Another usage could be for coding the impact of actions on different aspects of data quality improvement which helps interpretation and conclusion in reflection and learning stages. In fact, by using QDs, interpretations can be coded in different categories that would improve understanding and learning by looking from different defined angles to data and results.

The matter of data quality has its own constrains in Big Data context. Big Data is not only about data, but also about a complete conceptual and technological stack including raw and processed data, storage, ways of managing data, processing and analytics [42]. Three Data Quality characteristics for assessing the levels of Data Quality-in-Use in Big Data projects are Contextual Adequacy, Operational Adequacy and Temporal Adequacy. A challenge that becomes even trickier is the management of the quality of the data in Big Data environments. More than ever before the need for assessing the Quality-in-Use gains importance since the real contribution – business value – of data can be only estimated in its context of use. Although there exists different Data Quality models for assessing the quality of regular data, none of them has been adapted to Big Data [42].

2.2 Data Quality Processes

According to the definition claimed by Batini and Scannapieco (2006), the data quality activity refers to any process performed directly on data to improve the data quality [43]. Typically, the data quality process is a part of improving data quality which involves implementing integrity constraints in databases and setting up data quality processes as well as dedicated organizational structures [44]. The published classic approaches cannot successfully prevent data users from dealing with inadequate quality there must be a model for increasing the data quality based on the relevant processes.There are many data quality activities which are new data acquisition, standardization (or normalization), object identification (or record linkage, record matching, entity resolution), data integration, source trustworthiness, quality composition, error localization, error correction and cost optimization.

"New data acquisition" refers to a process of data collection to achieve new quality data and complete data. Based on the matter of time and also costs to prepare an adequate data acquisition process it seems very vital to perform innovative new data acquisition systems that can satisfy the desired constraints without losing data quality even improve the data quality [45].

"Standardization" is the data normalization in artefact core and refers to modification and transforming data to new data format or representation based on a defined standard. Data Warehousing (DWH) is one of the best solutions and practices for reporting and analysis [46, 47]. However based on data quality "standardization" methodology suggestions [19, 43, 48] and context characteristics, normalization concept [48] injected to data warehouse model to provide ability of data quality improvement. Thus new innovative data warehouse architectures is an interesting area of study to provide ability of data control for data warehouse concepts and control system models and apply Feedback Control System (FCS) theory to the reporting system based on data warehouse concept.

Feedback Control System (FCS) theory explains about importance of feedback in a system and how it should control the system input to improve quality of output [53]. FCS define a controller element to control plant (the object to be controlled) and generate input that can be extend to Database and DWH. This extend has been used in information systems and has suggested to use for improving data quality in information systems besides increasing the usage of data [22]. Orr suggested a model that embedded Information System (IS) in a large framework of Feedback-Control System model.

"Object identification" or better to say entity resolution is the process to extract entity from service logs to unique entity, configuration item or service information that exist in real world. This process happens in normalization stage that also reduces size of data and link entity with foreign key to main service logs. There are some concepts for object identification and functional dependencies is one of the proposed concepts for specifying matching rules. Matching dependencies is another concept that can be applied to different data quality applications such as detecting the violations of integrity constraints [49].

The created data of the new sciences and technologies pushes researchers for integration of data achieving from heterogeneous and diverse purposed sources, business rules, underlying models [50]. **"Data integration"** is happening to unify different data sources to one table. It also could happen by creating a unified view on different sources.

"Error localization" or error detection could use to find error in data based on some rules. As an example, it can be used in research artefact to capture data sources errors and prevent errors be transferred from Layer 1 to Layer 2. In some cases, errors are accepted as a part of environment situation that will catch by other mechanism. "The problems relevant to Error Localization concerns finding the minimum number of fields in a record such that by modifying the values in these fields the new record satisfies a given set of rules" [51]. Moreover, **"Error correction"** refers to a set of rules that fixes data errors.

In addition, it is relevant to review and consider data collection process part of Data Warehousing and DWH quality as well. Researches show in order to improve quality in DWH, it is required to enrich metadata facilities for the exploitation of the knowledge collected in a DWH [52]. One of the data quality models introduced for data warehouses is Information Production MAP (IP-MAP). Based on IP-MAP principles, data can be seen as particular product of a manufacturing activity [43]. However, for TSRs, data produce by machine and activities on IT environment beside service provider operation activity. In fact, IP-MAP does not define operation process specifically and it makes issue for the context. These two viewpoints and context are not fit together and in this case we cannot consider all information are product of organization which it would result in rejection of IP-MAP for TSR context.

3. DISCUSSION

This study has gone through the peer-reviewed literature to present approaches to the definition, classification, standardization, measurement, improvement and reporting of data quality. Quality dimensions have correlations together. In some cases, putting focus on one dimension causes effect on others however sometimes trade-offs may be done between timeliness and a dimension among accuracy, completeness, and consistency. Because, having accurate (or complete or consistent) data may require time, thus timeliness is in oppose [21]. This trade-off could be acceptable if in context, quality in one dimension still remain in acceptable amount to consumer.

Because the research community is still arguing about the exact meaning of each quality dimension, it is suggested to define more context related to each of Quality dimensions [21]. Thus, using a data quality framework with similar viewpoint and purpose helps researcher to choose quality dimensions' definition and evaluation questionnaire that fit to their research's context.

Similar to case problems which are looking for reports' accuracy, Wang and Strong (1996) did a comprehensive questionnaire around data quality dimensions which data consumer require. The results of that research led to develop a model that captures scopes of data quality which are important to Technical Service Reports in SLA context. In addition, this model has been introduced as it is the only model that supports believability and reputation dimensions which are important to the case organization. In fact, by referring to a wide industrial research (Chandler, 2012), believability behind SLA Dashboard is a broad issue among out sources IT services.

Based on objectives of data quality improvement in TSR, must focus on some of data quality dimensions. These Quality Dimensions (QDs) can be used as part of evaluation and measure of data quality improvement. Main solid usage of these QDs could be in evolutionary questionnaires that collects practitioners' evaluation of reports data quality improvement in the beginning and end of each improvement cycle. Another usage could be for coding the impact of actions on different aspects of data quality improvement which helps interpretation and conclusion in reflection and learning stages when an Action Design Research (ADR) methodology use for DQ improvement research. In fact, by using QDs, interpretations can be coded in different categories that would improve understanding and learning by looking from different defined angles to data, actions and results.

There are many data quality activities which are new data acquisition, standardization (or normalization), object identification (or record linkage, record matching, entity resolution), data integration, source trustworthiness, quality composition, error localization, error correction and cost optimization. Any conducted research must determine which ones are the used processes in its artefact or improvement model. In addition, the researcher must pay attention to the fact that although a special process may be used, there are some other methods which are not applicable to the TSR context or when DQ consider as "fitness for use" of data.

Feedback Control Systems (FCS) model are a unique model of the data quality process. FCS plays a major role as an effective cyclic improvement model in TSR context and covers all reviewed dimensions of data quality. FCS is defined based on "fitness for use" view point and compensates the lack of meta-data or CIs information quality which is known as a part of TSR quality challenges.

4. CONCLUSION

The review has been started with the terms of data quality. Because there are many points of view in this area, it is narrowed to the data quality which is effective on service reports and in service reports whatever is related to technical metrics definitions are chosen. Therefore, data quality in technical service reporting came up as the final narrowed criteria of the review.

The main purpose of this literature review is to clear the state of arts in data quality in technical service reporting and help researchers to find a profound understanding on this research area to organize and present their research study. Generally, this review set up the background for the study and reviewed knowledge area and technologies.

IT Service Management is a concept under IT Operation Management for implementing and managing qualified IT services to meet the needs of a business or organization by an IT Service provider through an appropriate mix of information technologies, process and people and there are seven international IT Service Management (ITSM) frameworks which are ITIL, ISO 20000, COBIT, CMMI-SVC, MOF, and ITUP [2]. Service reporting that is the main context of this literature review is a part of service delivery processes beside Service Level Management (SLM). Service reporting is a process to define, document, produce and use Service Reports (SRs) [3, 4]. SRs are for both management and customer that must include performance against target, relevant information about significant events and all measurable aspect of service for both current and historical analysis. All of these measureable aspects could process success rate or service end result and Technical Service Reports (TSRs) defined and appeared when customer requires a measureable performance of technical service result based on system- generated result such as successful backup rate or storage service up-time. Service Level Agreements include the contracted key elements to describe the full success of delivered services and specify the desired metrics to examine the effectiveness of service activities, functions and processes [6]. IT Service Management (ITSM) and Information Technology Infrastructure Library (ITIL) as a famous major implemented framework in ITSM [7] define the service reporting as the main process of SLA monitoring and the one which is responsible to provide Key Performance Indicator (KPI) reports for Service Level Management (SLM). Many IT service organizations consider the measurement of IT service management processes, especially service support processes, as a considerable challenge [8].

There are many data quality review papers but they only focused on the general quality dimensions of process. If there is a review on both dimension and process, it only define the general criteria with no specific purpose. This review focused on data quality which is used in service reporting and specifically on technical service reporting. So the scope became very narrow to reach the maximum efficiency of the review.

In order to improve TSRs' DQ and consequently improving SQ and reducing the risks of financial penalties, researchers need to know the constraints of DQ and this is not feasible except having a comprehensive overview on the dimensions of DQ and its processes. Generating TSRs are more complex and costly because system-generated logs need experts' verification, data cleansing and complex processing to fit in all business rules and condition defined in SLA. In fact, an expert must exclude testing issues or any kind of issues which is not genuine based on SLA definitions or exclude log noises that are not real service results. Wang and Strong (1996) Data Quality Dimension Model and Orr (1998) FCS model are two valuable models to be used for data quality improvement in TSR context. Thus, it is important to know how SLM reports' data quality in this context can be improved and consequently, Data Quality (DQ) in Technical Service Reports (TSRs) remains as the considerable gap.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

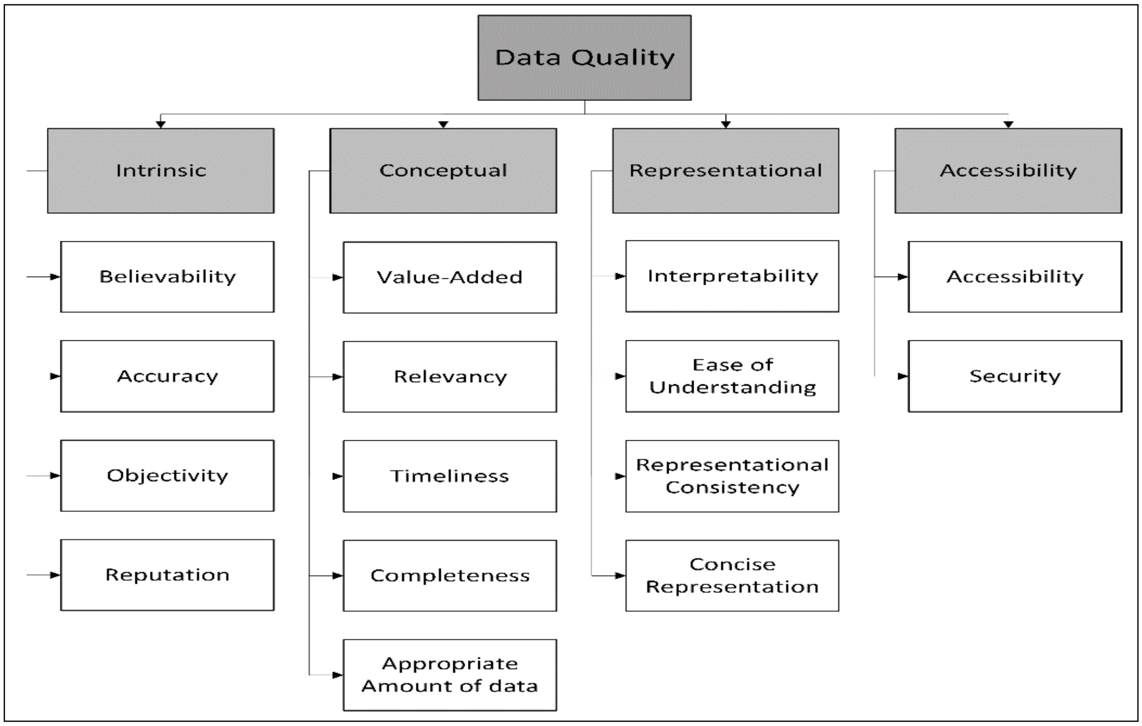

Fig 1: A conceptual framework of data quality (Wang and Strong, 1996)

Table 1: Data quality dimensions in different proposals (Madnick et al., 2009)

Dimensions	WandWang 1996	WangStrong 1996	Redman 1996	Jarke 1999	Bovee 2001
Accuracy	X	X	X	X	X
Completeness	X	X	X	X	X
Consistency / Representational Consistency	X	X	X	X	X
Time-related Dimensions	X	X	X	X	X
Interpretability		X	X	X	X
Ease of Understanding / understandability		X			
Reliability	X			X	
Creditability				X	X
Believability		X			
Reputation		X			
Objectivity		X			
Relevancy/ Relevance		X	X		X
Accessibility		X		X	X
Security / Access Security		X		X	
Value-added		X			
Concise representation		X			
Appropriate amount of data/ amount of data		X	X		
Availability				X	
Portability			X	X	
Responsiveness/ Response Time				X	

REFRENCES:

[1] S. Taylor, *The Official Introduction to the ITIL Service Lifecycle*. London: TSO, 2007.

[2] M. Mora, M. Raisinghani, R. O'Connor, J. Gomez, and O. Gelman, "An extensive review of IT service design in seven international ITSM processes frameworks: Part I," *International Journal of Information Technologies and Systems Approach,* vol. 7, pp. 83–107, 2014 2014.

[3] "IEEE Standard - Adoption of ISO/IEC 20000-1:2011, Information technology – Service management – Part 1: Service management system requirements," ed, 2013.

[4] C. Hsuan-Yu, C. Y. D. Sim, and L. Ching-Her, "Dual-Band Meander Monopole Antenna for WLAN Operation in Laptop Computer," *Antennas and Wireless Propagation Letters, IEEE,* vol. 12, pp. 694-697, 2013.

[5] W. Linlin and B. Rajkumar, "Service Level Agreement (SLA) in Utility Computing Systems," in *Grid and Cloud Computing: Concepts, Methodologies, Tools and Applications*, ed Hershey, PA, USA: IGI Global, 2012, pp. 286-310.

[6] N. Ranaldo and E. Zimeo, "Capacity-driven utility model for service level agreement negotiation of cloud services," *Future Generation Computer Systems,* vol. 55, pp. 186-199, 2// 2016.

[7] S. D. Galup, R. Dattero, J. J. Quan, and S. Conger, "An overview of IT service management," *Commun. ACM,* vol. 52, pp. 124-127, 2009.

[8] A. Lahtela, J. M, x0E, ntti, and J. Kaukola, "Implementing an ITIL-Based IT Service Management Measurement System," in *Digital Society, 2010. ICDS '10. Fourth International Conference on*, 2010, pp. 249-254.

[9] C. Anacleto, "Defining and Observing the Compliance of Service Level Agreements: A Model Driven Approach," 2010, pp. 165-170.

[10] A. Paschke and E. Schnappinger-Gerull, "A Categorization Scheme for SLA Metrics," in *MKWI*, 2006.

[11] H. Müller and J. C. Freytag, *Problems, Methods, and Challenges in Comprehensive Data Cleansing*: Humboldt-Univ. zu Berlin, 2005.

[12] A. Hanna, *ITIL glossary and abbreviations*. London: TCO, 2011.

[13] Y. Wand and R. Y. Wang, "Anchoring Data Quality Dimensions in Ontological Foundations," *Commun. ACM,* vol. 39, pp. 86–95, November 1996 1996.

[14] K. Bagchi, P. Kirs, G. Udo, and R. Cerveny, "Characteristics and determinants of insourced and offshored projects: A comparative analysis," *Journal of World Business,* vol. 50, pp. 108-121, 1// 2015.

[15] A. Karkouch, H. Mousannif, H. Al Moatassime, and T. Noel, "Data quality in internet of things: A state-of-the-art survey," *Journal of Network and Computer Applications,* vol. 73, pp. 57-81, 9// 2016.

[16] K. Ahn, H. Rakha, D. Hill, and C. Systematics, "Data quality white paper," US Department of Transportation, Federal Highway Administration2008 2008.

[17] R. Blake and G. Shankaranarayanan, "Data and Information Quality: Research Themes and Evolving Patterns," 2015.

[18] S. E. Madnick, R. Y. Wang, Y. W. Lee, and H. Zhu, "Overview and framework for data and information quality research," *Journal of Data and Information Quality (JDIQ),* vol. 1, p. 2, 2009 2009.

[19] L. L. Pipino, Y. W. Lee, and R. Y. Wang, "Data quality assessment," *Communications of the ACM,* vol. 45, pp. 211–218, 2002 2002.

[20] R. Y. Wang and D. M. Strong, "Beyond accuracy: What data quality means to data consumers," *Journal of management information systems,* pp. 5–33, 1996 1996.

[21] M. Scannapieco, P. Missier, and C. Batini, "Data Quality at a Glance.," *Datenbank-Spektrum,* vol. 14, pp. 6–14, 2005 2005.

[22] K. Orr, "Data quality and systems theory," *Communications of the ACM,* vol. 41, pp. 66–71, February 1998 1998.

[23] T. C. Redman, "The impact of poor data quality on the typical enterprise," *Communications of the ACM,* vol. 41, pp. 79–82, 1998 1998.

[24] C. W. Fisher and B. R. Kingma, "Criticality of data quality as exemplified in two disasters," *Information & Management,* vol. 39, pp. 109-116, December 2001 2001.

[25] S. Raghunathan, "Impact of information quality and decision-maker quality on decision quality: a theoretical model and simulation analysis,"

Decision Support Systems, vol. 26, pp. 275-286, October 1999 1999.

[26] I. N. Chengalur-Smith, D. P. Ballou, and H. L. Pazer, "The impact of data quality information on decision making: an exploratory analysis," *IEEE Transactions on Knowledge and Data Engineering,* vol. 11, pp. 853-864, November 1999 1999.

[27] C. W. Fisher, I. Chengalur-Smith, and D. P. Ballou, "The Impact of Experience and Time on the Use of Data Quality Information in Decision Making," *Info. Sys. Research,* vol. 14, pp. 170–188, June 2003 2003.

[28] W. Jung, L. Olfman, T. Ryan, and Y.-T. Park, "An experimental study of the effects of contextual data quality and task complexity on decision performance," in *Information Reuse and Integration, Conf, 2005. IRI -2005 IEEE International Conference on.*, 2005, pp. 149-154.

[29] G. Shankaranarayanan and Y. Cai, "Supporting data quality management in decision-making," *Decision Support Systems,* vol. 42, pp. 302-317, October 2006 2006.

[30] L. Berti-Équille, "Measuring and Modelling Data Quality for Quality-Awareness in Data Mining," in *Quality Measures in Data Mining*, F. J. Guillet and H. J. Hamilton, Eds., ed Berlin, Heidelberg: Springer Berlin Heidelberg, 2007, pp. 101-126.

[31] D. J. Hand, "Principles of Data Mining," *Drug Safety,* vol. 30, pp. 621-622, 2007// 2007.

[32] J. Hipp, U. Güntzer, and U. Grimmer, "Data Quality Mining -- Making a Virtue of Necessity," in *6TH ACM SIGMOD WORKSHOP ON RESEARCH ISSUES IN DATA MINING AND KNOWLEDGE DISCOVERY (DMKD 2001)*, 2001, pp. 52-57.

[33] Y. W. Lee, D. M. Strong, B. K. Kahn, and R. Y. Wang, "AIMQ: a methodology for information quality assessment," *Information & management,* vol. 40, pp. 133–146, 2002 2002.

[34] C. Batini and M. Scannapieca, "Introduction to Data Quality," in *Data Quality: Concepts, Methodologies and Techniques*, ed, 2006, pp. 1–18.

[35] A. Klein and W. Lehner, "Representing data quality in sensor data streaming environments," *Journal of Data and Information Quality (JDIQ),* vol. 1, p. 10, 2009 2009.

[36] D. M. Strong, Y. W. Lee, and R. Y. Wang, "Data Quality in Context," *Commun. ACM,* vol. 40, pp. 103–110, May 1997 1997.

[37] G. M. O'Reilly, B. Gabbe, L. Moore, and P. A. Cameron, "Classifying, measuring and improving the quality of data in trauma registries: A review of the literature," *Injury,* vol. 47, pp. 559-567, 3// 2016.

[38] A. B. o. S. (ABS), "Information Paper: Quality Dimensions of the Australian National Accounts," ed, 2007.

[39] P. Nousak and R. Phelps, "A Scorecard approach to improving Data Quality," in *SUGI 27*, Orlando, Florida, 2002.

[40] D. Loshin. (2011) Monitoring Data Quality Performance Using Data Quality Metrics. *Informatica*.

[41] R. Y. Wang and D. M. Strong, "Beyond Accuracy: What Data Quality Means to Data Consumers," *Journal of Management Information Systems,* vol. 12, pp. 5-33, 1996/03/01 1996.

[42] J. Merino, I. Caballero, B. Rivas, M. Serrano, and M. Piattini, "A Data Quality in Use model for Big Data," *Future Generation Computer Systems,* vol. 63, pp. 123-130, 10// 2016.

[43] C. Batini and M. Scannapieco, *Data Quality - Concepts, Methodologies and Techniques*. New York: Springer, 2006.

[44] C. Reuter, F. Brambring, J. Weirich, and A. Kleines, "Improving Data Consistency in Production Control by Adaptation of Data Mining Algorithms," *Procedia CIRP,* vol. 56, pp. 545-550, // 2016.

[45] E. Marcucci and V. Gatta, "How Good are Retailers in Predicting Transport Providers' Preferences for Urban Freight Policies?... And Vice Versa?," *Transportation Research Procedia,* vol. 12, pp. 193-202, 2016/01/01 2016.

[46] W. H. Inmon, "The data warehouse and data mining," *Communications of the ACM,* vol. 39, pp. 49–50, 1996 1996.

[47] W. H. Inmon, *Building the data warehouse*: John wiley & sons, 2005.

[48] P. P.-S. Chen, "The entity-relationship model—toward a unified view of data," *ACM Transactions on Database Systems (TODS),* vol. 1, pp. 9–36, 1976 1976.

[49] S. Song and L. Chen, "Efficient discovery of similarity constraints for matching dependencies," *Data & Knowledge Engineering,* vol. 87, pp. 146-166, 9// 2013.

[50] O. Brazhnik and J. F. Jones, "Anatomy of data integration," *Journal of Biomedical Informatics,* vol. 40, pp. 252-269, 6// 2007.

[51] J. Riera-Ledesma and J.-J. Salazar-González, "A heuristic approach for the continuous error localization problem in data cleaning," *Computers & Operations Research,* vol. 34, pp. 2370-2383, 8// 2007.

[52] P. Vassiliadis, M. Bouzeghoub, and C. Quix, "Towards quality-oriented data warehouse usage and evolution," *Information Systems,* vol. 25, pp. 89-115, April 2000 2000.

[53] Doyle, J. C., B. A. Francis, and A. R. Tannenbaum. "Feedback Control Theory,(1992)." Maxwell Communication Group, New York (2001).