The Development of Data Quality Metrics using Thematic Analysis

Puteri Nor Ilya Nadia Zulkiffli, Emelia Akashah P. Akhir, NorShakirah Aziz, Karl Cox

Abstract: ***Data quality management remains a challenge in every organization in which high quality data needed to help in decision making. Poor data quality management has a negative impact that can result in financial loss, loss of privacy, business process failure and inefficiencies, creates legal and security risks and loss of reputation. Much research has been conducted on data quality metrics and related information such as selecting data quality dimensions but most of the studies on data quality metrics are less than fit to help in decision making[1]–[8]. There is also lack of standardization regarding usage of each dimension. These reasons provide the rationale for the objectives of this study which are (1) To investigate the appropriateness of dimensions used in existing data quality metrics that used in assessing data quality and (2) To propose a data quality metrics to assess quality of data. The study conducted an extensive literature review to address the objectives. We used thematic analysis to determine the theme studied by each paper. Then, a data quality metrics is suggested with dimensions that are discussed. The study found that key to importance is accuracy, completeness, timeliness and scope of data.***

Keywords: data quality, data quality dimension, accuracy, completeness, timeliness

I. INTRODUCTION

Research on data quality is not something new in the world as it is needed to help in decision making in any organizations or industry that has data that is processed. However, some organizations still do not have proper data quality guidelines to qualify the data because the cost of doing so is high. The negative impacts of poor data quality management are felt not only in term of monetary inefficiencies or loss but also could be significantly worse as described above. Although the issue of data quality has existed for a long time, maintaining high quality data faces new challenges in this data-driven world, where data about the usage of everything is of potential importance to organizations. Data now comes in from a myriad of source, not just from shopping habits, or process efficiencies statistics or from accounting and sales trends within a company. Data is now everywhere, flowing continually from multiple sources and platforms.

Revised Manuscript Received on May 22, 2019.

Puteri Nor Ilya Nadia Zulkiffli, Computer and Information Science Department, UniversitiTeknologi PETRONAS, Perak, Malaysia.

Emelia Akashah P. Akhir, Positive Computing, UniversitiTeknologi PETRONAS, Perak, Malaysia

NorShakirah Aziz, Center for Research in Data Science, UniversitiTeknologi PETRONAS, Perak, Malaysia

Karl Cox, School of Computing, Engineering and Mathematics, University of Brighton, East Sussex, United Kingdom

The era of really Big Data is here as companies chase the goose that laid the golden egg of massive market share potential simply from all the data of our everyday lives. Companies who engage in the deal end of the Big Data Era have to face these incredible challenges of managing data effectively and ensuring those nuggets of data they do need to extract from the masses are high quality and fit for purpose. By implementing data quality metrics, every data is received by the decision makers, will be sorted and filter in order to ensure the data that complies with standard of quality required in order to aid in making decisions. Otherwise, the data will be eliminated. This paper started by discussing about definition of data quality metrics, the most discussed dimensions in previous literature then based on the extensive literature review, completeness, accuracy and timeliness are choosing to be focus on. The discussion will go deeper in these dimensions' definition and measurement which expected to achieve the objective of this paper that are (1) to investigate the appropriateness of dimensions used in existing data quality metrics that used in assessing data quality and (2) to propose a data quality metrics to assess quality of data.

II. LITERATURE REVIEW

Definition of Data Quality and Overview of Data Quality Metric as Dataset Checker

Data could be in many form or format such as it could be in combination of figures, facts and statistics that telling about the overview of the organizations' past activities that will lead to profit or loss[9]. Even though many scholars state that there is still no a standard definition of data quality [1]–[3], [5], [6], [10] or that an exact data quality measurement is yet to be proven successful[11], most of them come to almost the same conclusion that the current definition of data quality is "fitness for use" [3], [7], [8], [10], [12]. The differences in opinions on data quality definitions by the scholars will always happened but the fundamental concept of the data quality itself is not only to discuss about the data quality features only but must relate with the nature of the business that the data quality will be applied because data quality have its own requirements that need to be achieve in order to consider the data is quality [3]. As a simple explanation, high quality data will need to be used accordingly based on its purpose and field that it meant for [1], [10], [12].

According to [13], data quality metric is a set of measurement that will be used to assess what quality means to measure the data and the metric could be a formula or set of rule that could calculate the 'quality' of the data. Data quality metric is one of the various techniques in quality management which will consist of different dimension[13]. Data quality metric will need to be defined clearly as it will be the guidance to ensure the initiatives done by organizations to improve their data quality management is effective[14] while data quality dimension is "a set of data quality attributes that represent a single aspect or construct of data quality"[3]. The selected dimensions will be different according to the scope of the study and business environment involved[3].

Dimensions Mostly Discussed in Previous Literature

There are many existing researches that highlighted on the dimensions chose with different method on deciding which dimensions that suit most. The dimensions will need to be capable enough to measure the quality of data for the targeted organization as the dimension need to suit the actual circumstance of targeted organization [3]and the right choices of dimensions and they are correlation, it will help in creating high quality of data [6], [7]. There are 60 literatures that been referred, to look for the most mentioned dimensions by the previous scholars and also to know how the previous scholars defined each dimension.

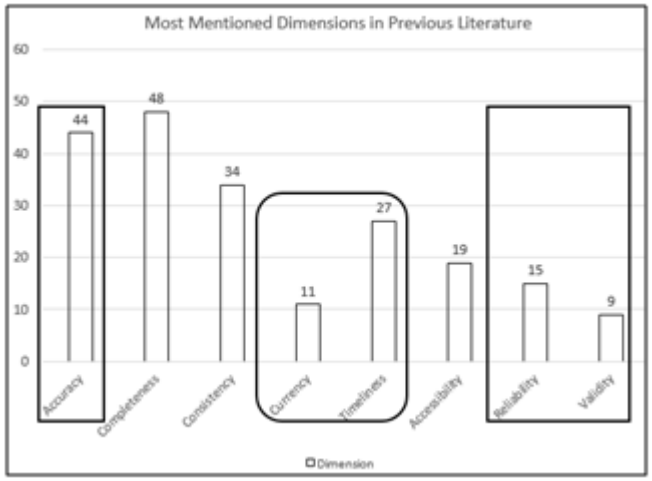

Fig. 1 Votes for Most Mentioned Dimensions in Previous Literature

There are many researchers that come out with various set of dimensions that they had selected for their targeted organization or for their general research purposes according to the fit for purpose definition of data quality [1], [15], [16]. Based on Figure 1, this study will be highlighted only on completeness, accuracy and timeliness because these three dimensions are most mentioned by most researchers for the past few years back. The finding shows that 48 authors consistently mentioned completeness as one of the vital dimensions in any data quality metrics followed by accuracy (in rectangle shape) with 44 authors and last but not least, timeliness (in semi-rectangle shape) and currency are two same dimensions that been mentioned interchangeably by few scholars[17]–[19]which means that if sum both dimensions' votes together it will be the third highest dimension been mentioned by previous scholars.

The role of these three dimensions seems to be crucial especially when the data nowadays entering the big data era that having massive amount of information that have the potential to help the organization to improve their business [3]. Table 1 shows the summary of scholars' opinions which taken according to different domain which became their main focus in their study become an evidence to be one of the reasons to select only these three dimensions for further study as the importance of these three dimensions been supported by previous scholars.

Table. 1 Scholars that Supported the Importance of Completeness, Accuracy and Timeliness

No.	Authors Agreed the Crucial of the Dimensions	Dimension Selected with Summary of Scholars' Statement
1	[20]–[24]	Accuracy is a fundamental issue and most important dimension
2	[2], [20], [25], [26]	Completeness is among the most commonly studied and expected to make decision
3	[3], [16], [20], [27], [28]	Timeliness is important in big data era which can effect decision making

The importance to have these dimensions in organizations' data quality metrics and the real scenario happened when completeness, accuracy and timeliness been ignored by organizations are discussed in section C for completeness, section D for accuracy and section E for timeliness. Besides, each dimensions' section will discuss on the summary of the definition of each dimension and sub-dimensions after extensive literature review been done by the author.

Completeness Dimension

Data is usually used for more than one purpose and the completeness dimension will ensure that the data need to acquire all values that been needed [12]. As stated by [29]a system database will be a challenge to an organization if completeness been ignored. Based on previous literature, some definitions used by most scholars already stated the measurements that involve to measure completeness. The sub-dimensions for data completeness are complete, sufficient breadth and depth, data is not missing and null is defined are among the command measurement stated in definition of completeness.

There are five (5) authors that having the same thought that sub-dimensions for completeness should be complete or contain all necessary information [12], [30]–[32]while[28]used word of comprehensive to define completeness which according to Oxford Dictionary means, "Including or dealing with all or nearly all elements or aspects of something". Due to the same nature of words, these 5 definitions bring in the same measurement for completeness. The other six (6) authors agreed on sufficient breadth and depth as completeness measurement[8], [26], [33]–[36] while ten (10) authors used "data is not missing" as sub-dimension for data completeness[1], [2], [18], [26], [33], [35], [37]–[40]. [41], [42] are having the same thought that null value should be defined as measurement for completeness.

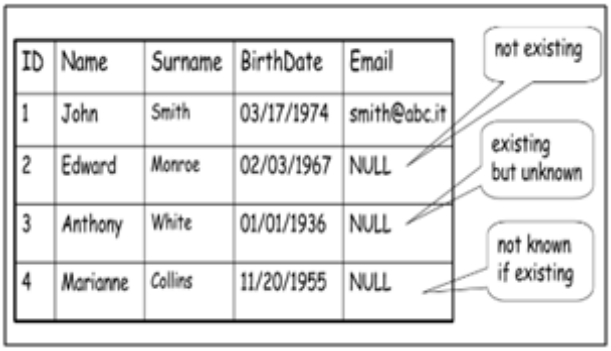

Fig. 2 Measurement Parameter for Completeness (Source: Batini et al., 2009)

Figure 2 shows the example of relational database that how the measurement for completeness take place which is the data is complete when there are sufficient of breadth and depth, and data is not missing by the null spaces are defined. For completeness, it is vital to know the reason why the value is missing either it really not exists, or it exist but it is unknown or it is not known whether it is exist[43].Based on the measurement used by scholars for completeness, there are few authors are having clashing of opinions about the sub-dimensions for completeness such as[18], [43], [44]. [18]has relates reliability with completeness even though other scholars such[16], [23], [26], [33], [36] have relate the reliability to measure accuracy while[43] has defined completeness almost the same as definition of accuracy by[44]which contradict from other scholars that defined accuracy and completeness as something different.

Accuracy Dimension

Accuracy is needed to assess quality of data as the bigger the volume of data collected, more errors such as duplication will be found [15]. Without quality of accuracy in data will increase the difficulty of data assessment such as in short term, it will affect the statistics that users want to do while for a long term effect, an efficient decision making cannot be achieved[45].

Based on the dimensions that most mentioned by past literatures, some of the dimensions are having almost the same definition and referred to almost the same indicators. Based on the definitions of accuracy given by previous scholars, there are few sub-dimensions are having the same meaning which also refer to accuracy such as correctness, reliability, precision and validity. [28], [46] defined reliability the same as definition of accuracy while [34] defined correct and precise with the same key word for accuracy which are reliable and exact. The same thing done by [28] has defined validity the same as accuracy. Besides having almost the same definition, there are scholar like[18] that used correctness to measure accuracy, while[47]used precision to help to assess accuracy. As for [48] has group correctness, validity and precision together as sub-dimensions for accuracy. Some of the scholars are using different definition on the dimension that actually lead to measure the same dimensions which later it been group together to be use as measurement for accuracy. Correctness, reliability, precision and validity as they having almost the same definition with accuracy that could become the measurement to assess accuracy.

There are also clashing of opinions between scholars as many scholars agreed on the measurements mentioned above are having relation with accuracy but some scholars like[10], [13], [18], [35], [49]are having different thought as their definitions of reliability and validity are referring to dimension of consistency and not accuracy. There are 17 authors stated that the measurement for accuracy is correct or accurate as according to Oxford Dictionary, accurate means "especially of information, measurements, or predictions) correct in all details; exact." which have similarities with meaning of correct.

Timeliness Dimension

The rapid movement of data in the world nowadays, make the assessment of quality of data needs towards timeliness dimension [4]. In current Big Data era, the needs for the data receive by the decision makers must have their own timestamp otherwise users of the data will not know which one comes first and which one later [17] due to the speedily movement of the data. One of the challenges to the retailing industry in the latest research done shows that receiving the lack of timeliness data for the business is not good for a competitive environment such retailing industry[50].

Based on the definitions of timeliness given by previous scholars, having the same meaning with other sub-dimension such as currency and volatility. There are three kind of different opinions portrayed by previous research about timeliness dimension and the sub-dimensions. Scholars like[5], [11] defined timeliness, currency and volatility as one which related to each other, while [17]–[19]used timeliness and currency interchangeably in their research that both have almost the same definition. Last but not least, [40] mentioned that according to his research, several scholars even gave different definitions for currency and timeliness.

These different opinions from previous literature shows that a standard data quality metrics or uniform definition for these dimensions is not yet available. Further discussion on the measurement for timeliness stated that there are four sub-dimensions seem to be used by many previous literatures which are the data under dimension of timeliness needs to be timely, the data is available whenever needed, frequency of the updated data and last but not least current or up-to-date measurement are considered as one because according to online Oxford dictionary with latest update in 2017, they are having the same definition.

III. FINDINGS AND RESULT

Previously in literature review section, we have discussed the importance of each selected dimension and also the definitions used by past literature for the selected dimensions and sub-dimensions which considered the agreement between the scholars and also the difference in term of their opinions in expressing the definition of each dimensions and sub-dimensions. This section will summarize the discussion in previous section for each selected dimensions starting from completeness, followed by accuracy and timeliness. Last but not least this section will discuss on the overall result for preliminary data quality model.

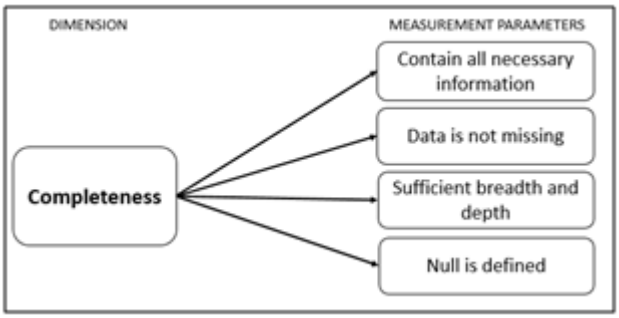

Fig. 3 Dimension of Completeness and the Sub-dimensions

The summary for completeness could be seen based on Figure 3 above which shows the dimension discussed previously. Completeness and the sub-dimensions are group together after done theme analysis. These sub-dimensions are mostly mentioned by the past literature on the definitions they used on completeness. The summary for each sub-dimensions' definition that been discussed in literature review section is shown in the Table 2 below.

Table. 2 Summary of Sub-Dimensions Definitions

No.	Sub-dimension	Summary of definition
1	Contain all necessary information	All required parts of an entity's description are included
2	Sufficient breadth and depth	Sufficient breadth and depth to describe the corresponding set of the real world
3	Data is not missing	The cells that are not empty and have a meaningful value assigned
4	Null is defined	The term null must have the definition either it is not existing, or existing but unknown or not known if it is existing

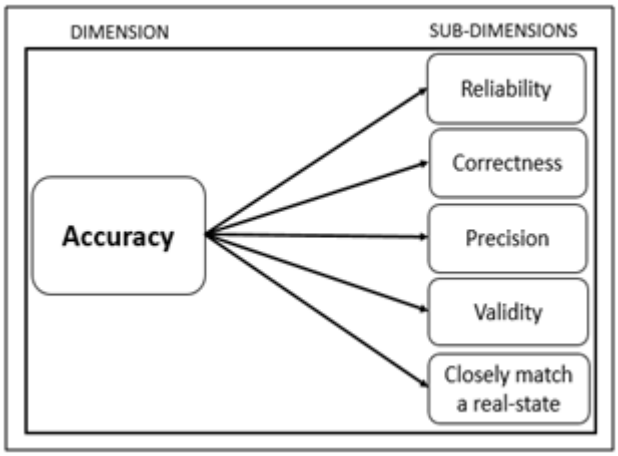

Fig. 4 Dimensions of Accuracy and the Sub-dimensions

The summary for this section could be seen based on Figure 4that shows the dimension discussed above, which is accuracy and the sub-dimensions which are reliability, correctness, precision, validity and closely match a real-state that been group together after done theme analysis that shows these sub-dimensions are mostly mentioned by the past literature in the definitions or measurement they used to discuss about accuracy. The summary for each definition that been discussed above is shown in the Table 3 below.

Table. 3 Summary of Sub-dimensions Definition

No.	Sub-dimension	Summary of definition
1	Reliability	Specification can be depended on
2	Correctness	Being free of errors
3	Precision	Attributes that are exact or that provide discrimination (understanding) in a specific context of use
4	Validity	The contents are within the pre-specified value domain
5	Closely match the real-state	The real-world entities reflected by the values receives

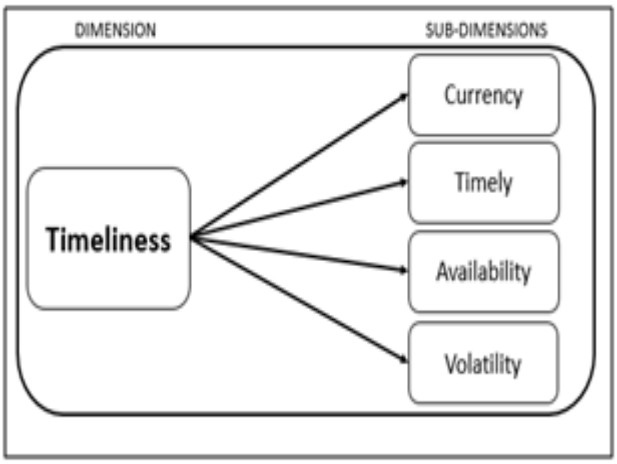

Fig. 5 Dimension of Timeliness and the Sub-Dimensions

The summary for this section could be seen based on Figure 5 that shows the dimension discussed above, which is timeliness and the sub-dimensions which are currency, timely, availability and volatility that been group together after done theme analysis that shows these sub-dimensions are mostly mentioned by the past literature in the definitions or measurement they used to discuss about timeliness. The summary for each definition that been discussed above is shown in the Table 4 below.

Table. 4 Summary of Sub-dimensions Definition

No.	Sub-dimension	Summary of definition
1	Currency	How promptly data are updated
2	Timely	Near real-time as possible
3	Availability	Available to organization within a specified time period
4	Volatility	Frequency data vary in time

Based on discussion on the definitions in previous literature review section has resulted in the development of the preliminary data quality model as shows in Figure 6 below. The sub-dimensions are taken from the definitions that discussed in previous literatures. Based on the definitions that discussed in the tables as stated in this findings and result section, the author come out with the overall data quality model with the sub-dimensions such as accuracy (in rectangle shape) measure the data according to the degree of reliable, precise, correct, valid and the data closely match a real-state. For completeness, the measurement are the data is complete when there are sufficient of breadth and depth, and data is not missing by the null spaces are defined while for timeliness dimension (in semi rectangle), the measurement parameters proposed for preliminary study is the data need to be up-to-date, timely, available whenever needed and frequency of updated data is considering.

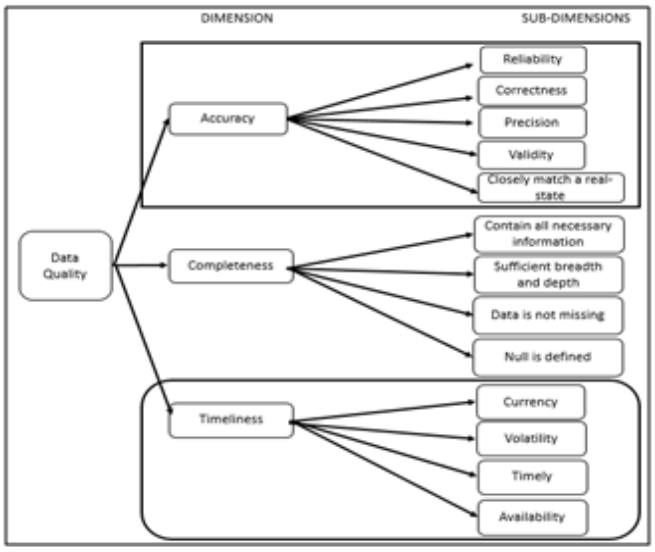

Fig. 6 Preliminary Data Quality Model

IV. CONCLUSION AND FUTURE WORK

Accuracy, completeness and timeliness were selected as focus of this study as these dimensions are most crucial in big data era as well as most mentioned by the literature as important. This paper examined the measurements used to assess the dimensions discussed based on definitions and measurement used in the literature. Previous work has used different definitions for different dimensions. This became the main findings of this study; that many sub-dimensions are actually referring to the same dimensions as shown in Figure 6. As the measurements have been grouped together, our findings have addressed the objectives of this study which are (1) To investigate the appropriateness of dimensions used in existing data quality metrics that used in assessing data quality in retail dataset; (2) To proposed a data quality metrics to assess quality of data in retail dataset. Data quality metrics produced need to be able to cater to the Big Data era. Future work will be focus on retailing dataset so that the data quality metrics developed must be tailored to the needs of that industry.

V. ACKNOWLEDGEMENTS

We would like to express our gratitude to our reviewers for giving good review and comments to improve this paper. We also would like to show our appreciation to those support financially for this research

REFERENCES

1. J. Agre, M. S. Vassiliou, and C. Kramer, "Science and Technology Issues Relating to Data Quality in C2 Systems," no. February 2011, pp. 1–31, 2011.
2. P. Angeles and F. García-ugalde, "A Data Quality Practical Approach," vol. 2, no. 2, pp. 259–274, 2009.
3. L. Cai and Y. Zhu, "The Challenges of Data Quality and Data Quality Assessment in the Big Data Era," Data Sci. J., vol. 14, no. 0, p. 2, 2015.
4. Taleb, H. T. El Kassabi, M. A. Serhani, R. Dssouli, and C. Bouhaddioui, "Big Data Quality: A Quality Dimensions Evaluation," 2016 Int Ieee Conf. Ubiquitous Intell. Comput. Adv. Trust. Comput. Scalable Comput. Commun. Cloud Big Data Comput. Internet People, Smart World Congr., pp. 759–765, 2016.
5. M. A. Serhani, H. T. El Kassabi, I. Taleb, and A. Nujum, "An Hybrid Approach to Quality Evaluation across Big Data Value Chain," 2016 IEEE Int. Congr. Big Data (BigData Congr., no. August, pp. 418–425, 2016.
6. F. Sidi, P. H. S. Panahy, L. S. Affendey, M. A. Jabar, H. Ibrahim, and A. Mustapha, "Data quality: A survey of data quality dimensions," 2012 Int. Conf. Inf. Retr. Knowl. Manag., pp. 300–304, 2012.
7. R. Vaziri, M. Mohsenzadeh, and J. Habibi, "Measuring data quality with weighted metrics," Total Qual. Manag. Bus. Excell., vol. 3363, pp. 1–13, 2017.
8. Wahyudi, G. Kuk, and M. Janssen, "A Process Pattern Model for Tackling and Improving Big Data Quality," Inf. Syst. Front., 2018.
9. K. Chongbang, "A Study on Information Management System in Tesco Plc," 2012. [Online]. Available: https://www.scribd.com/doc/%0A79844828/PDF-Information-Management-System-in-Tesco%0A. [Accessed: 10-Oct-2018].
10. L. Bertossi and M. Milani, "Ontological Multidimensional Data Models and Contextual Data Qality," no. April, 2017.
11. P. Michael, "Continuous Data Quality Assessment in Information systems," J. Hydroinformatics, vol. 17, no. 4, pp. 640–661, 2015.
12. S. M. Ali, N. Anjum, M. N. Kamel Boulos, M. Ishaq, J. Aamir, and G. R. Haider, "Measuring management's perspective of data quality in Pakistan's Tuberculosis control programme: a test-based approach to identify data quality dimensions," BMC Res. Notes, vol. 11, no. 1, p. 40, 2018.
13. D. Becker, "Flexible & Generic Data Quality Metrics , Measurements & Assessments." The MITRE Corporation, 2012.
14. M. Lebied, "The Ultimate Guide to Modern Data Quality Management (DQM) For An Effective Data Quality Control Driven by The Right Metrics," Datapine, 2018. [Online]. Available: https://www.datapine.com/blog/data-quality-management-and-metrics/. [Accessed: 21-Oct-2018].
15. Aggarwal, "Identification of Quality Parameters associated with 3V ' s of Big Data," pp. 1135–1140, 2016.
16. R. Silvola, J. Harkonen, O. Vilppola, H. K. Vehkapera, and H. Haapasalo, "Data quality assessment and improvement," Int. J. Bus. Inf. Syst., vol. 22, no. 1, p. 62, 2016.
17. W. Fan and F. Geerts, Foundations of of Data Quality Management. 2012.
18. N. Larburu, R. Bults, M. Van Sinderen, and H. Hermens, "Quality-of-data management for telemedicine systems," Procedia Comput. Sci., vol. 63, no. Icth, pp. 451–458, 2015.
19. H. M. Sneed, B. Demuth, and B. Freitag, "A process for assessing data quality," Proc. - IEEE 6th Int. Conf. Softw. Testing, Verif. Valid. Work. ICSTW 2013, pp. 114–119, 2013.
20. Bronselaer, J. Nielandt, T. Boeckling, and G. De Tre, "Operational Measurement of Data Quality," Commun. Comput. Inf. Sci., vol. 81 PART 2, pp. 529–540, 2018.
21. R. Aliguliyev and Y. Imamverdiyev, "Conceptual Big Data Architecture for the Oil and Gas Industry," Probl. Inf. Technol., vol. 08, no. 1, pp. 3–13, 2017.
22. H. T. Moges, V. Van Vlasselaer, W. Lemahieu, and B. Baesens, "Determining the use of data quality metadata (DQM) for decision making purposes and its impact on decision outcomes - An exploratory study," Decis. Support Syst., vol. 83, pp. 32–46, 2016.
23. H. T. Moges, "A Contextual Data Quality Analysis for Credit Risk Management in Financial Institutions," no. 466, 2014.
24. K. Orr, "Data quality and systems theory," Commun. ACM ACM, vol. 41, no. 2, pp. 66–71, 1998.
25. M. I. Günes, "Data Quality Assessment in Credit Risk Management by a Customized Total Data Quality Management Approach," no. February, 2016.
26. H. Jingyu and C. Kejia, "Ranking Wikipedia article' s data quality by learning dimension distributions," vol. 3, no. 3, pp. 207–227, 2014.
27. J. Eweje, R. Turner, and R. Müller, "Maximizing strategic value from megaprojects : The influence of information-feed on decision-making by the project manager," JPMA, vol. 30, no. 6, pp. 639–651, 2012.
28. M. Gustavsson and C. Wänström, "Assessing information quality in manufacturing planning and control processes," Int. J. Qual. Reliab. Manag., vol. 26, no. 4, pp. 325–340, 2009.
29. S. D. Thomson and N. Beagrie, "Preserving Transactional Data," DPC Technol. Watch Rep., vol. 16-02, no. May, p. 44, 2016.
30. K. J. Reiche and E. Hofig, "Implementation of metadata quality metrics and application on public government data," Proc. - Int. Comput. Softw. Appl. Conf., pp. 236–241, 2013.
31. M. Joly, "Refinery production planning and scheduling: The refining core business," Brazilian J. Chem. Eng., vol. 29, no. 2, pp. 371–384, 2012.
32. N. Zellal and A. Zaouia, "A measurement model for factors influencing data quality in data warehouse," 2016 4th IEEE Int. Colloq. Inf. Sci. Technol., pp. 46–51, 2016.
33. P. J. Mirski and R. Bernsteiner, "Data Quality Assessment in Digital Labor Markets," in Research Gate, 2017, no. July.
34. Rafique, P. Lew, M. Q. Abbasi, and Z. Li, "Information Quality Evaluation Framework: Extending ISO 25012 Data Quality Model," Int. J. Comput. Electr. Autom. Control Inf. Eng., vol. 6, no. 5, pp. 568–573, 2012.
35. M. Salati et al., "The European thoracic data quality project: An Aggregate Data Quality score to measure the quality of international multi-institutional databases," Eur. J. Cardio-thoracic Surg., vol. 49, no. 5, pp. 1470–1475, 2016.
36. D. M. Strong, Y. W. Lee, and R. Y. Wang, "Data Quality in Context," Commun. ACM, vol. 40, no. 5, pp. 103–110, 1997.
37. Bronselaer, R. De Mol, and G. De Tre, "A Measure-Theoretic Foundation for Data Quality," IEEE Trans. Fuzzy Syst., vol. 26, no. 2, pp. 627–639, 2016.
38. V. Charles, J. Stiller, P. Kiraly, W. Bailer, and N. Freire, "Data Quality Assessment in Europeana : Metrics for Multilinguality Data Quality Assessment in Europeana : Metrics for Multilinguality," 2017, no. September.
39. P. N. Mendes, H. Mühleisen, and C. Bizer, "Sieve: Linked Data quality assessment and fusion," ACM Int. Conf. Proceeding Ser., pp. 116–123, 2012.
40. Vetrò, L. Canova, M. Torchiano, C. O. Minotas, R. Iemma, and F. Morando, "Open data quality measurement framework : Definition and application to Open Government Data," Gov. Inf. Q., 2016.
41. L. David, "Data Quality Fundamental," 2010, no. 301, pp. 1–47.
42. M. Torchiano, A. Vetro, and F. Iuliano, "Preserving the Benefits of Open Government Data by Measuring and Improving Their Quality: An Empirical Study," 2017 IEEE 41st Annu. Comput. Softw. Appl. Conf., pp. 144–153, 2017.
43. C. Batini, C. Cappiello, C. Francalanci, and A. Maurino, "Methodologies for data quality assessment and improvement," ACM Comput. Surv., vol. 41, no. 3, pp. 1–52, 2009.
44. G. Quirin and K. Marcus, "An Indicator Function for Insufficient Data Quality - A Contribution to Data Accuracy," Lect. Notes Bus. Inf. Process., vol. 129, no. September 2012, pp. 169–184, 2012.
45. Y. Sun, T. Lu, and N. Gu, "A method of electronic health data quality assessment: Enabling data provenance," Proc. 2017 IEEE 21st Int. Conf. Comput. Support. Coop. Work Des. CSCWD 2017, pp. 233–238, 2017.
46. B. Heinrich, D. Hristova, M. Klier, A. Schiller, and M. Szubartowicz, "Requirements for Data Quality Metrics," J. Data Inf.

Qual., vol. 9, no. 2, pp. 1–32, 2018.
47. Levitin and T. Redman, “Quality Dimensions of a Conceptual,” Science (80-.)., vol. 31, no. 1, pp. 81–88, 1995.
48. Batini, A. Rula, M. Scannapieco, and G. Viscusi, “From Data Quality to Big Data Quality,” J. Database Manag., vol. 26, no. 1, pp. 60–82, 2015.
49. R. Tate, D. Kalra, R. Boggon, N. Beloff, S. Puri, and T. Williams, “Data quality in European primary care research databases,” 2014 IEEE-EMBS Int. Conf. Biomed. Heal. Informatics, BHI 2014, no. September 2013, pp. 85–88, 2014.
50. Quatrro, “TOP-5 CHALLENGES THAT WILL AFFECT GROCERY OPERATORS IN 2017,” 2017. [Online]. Available: https://www.quatrrobss.com/challenges-grocery-operators.php. [Accessed: 09-Oct-2018].