DATA QUALITY MEASUREMENT TO IMPROVE SQA APPLICATION IN PT XYZ

Ratna Yulika Go[1], Rahmi Julianasari[2], Ahmad Syaifulloh Imron[3] dan Yova Ruldeviyani[4]

[1] Master of Information Technology, Faculty of Computer Science,
University of Indonesia, DKI Jakarta, Indonesia
[1]Email: ratna.yulika@ui.ac.id
[2]Email: rahmi.julianasari11@ui.ac.id
[3]Email: ahmad.syaifulloh11@ui.ac.id
[4]Email: yova@cs.ui.ac.id

ABSTRAK

Service Quality Assurance (SQA) adalah salah satu unit di PT XYZ sebuah perusahaan *e-commerce* Indonesia. SQA bertanggung jawab untuk memastikan dan menjaga layanan yang diberikan kepada pelanggan oleh agen *Customer Service* untuk memenuhi standar kualitas perusahaan. Namun aplikasi tersebut dibangun tanpa integrasi dengan sistem *Customer Relationship Management* (CRM) yang merupakan aplikasi layanan pelanggan utama di PT XYZ. Masalahnya adalah jika perusahaan membutuhkan data SQA, mereka harus menginput data secara manual dari SQA ke CRM. Hal ini dapat menyebabkan kerentanan dalam proses input data. Akibatnya berdampak pada bisnis proses perusahaan yang tidak efisien dan dapat merugikan dari sisi material maupun immaterial. Tujuan penelitian ini adalah untuk memberikan rekomendasi dalam meningkatkan kualitas data SQA agar perusahaan dapat menggunakan data tersebut untuk diintegrasikan ke dalam sistem CRM. Metode *Total Data Quality Management* (TDQM) digunakan untuk penelitian ini dengan tiga dimensi yaitu kelengkapan, validitas dan akurasi. Hasil dari masing-masing dimensi adalah 99,64% untuk dimensi kelengkapan, 84,75% untuk validitas, dan 100% untuk akurasi. Aturan bisnis pada dimensi validitas yang memiliki kualitas data terendah V1 28,57% dan V6 65,11%. Faktor permasalahan yang teridentifikasi adalah kamus data yang tidak lengkap, dokumen SOP yang tidak lengkap dan kadaluwarsa, serta tidak adanya kontrol terhadap proses bisnis dan data pada aplikasi. Rekomendasi untuk meningkatkan kualitas data aplikasi adalah PT XYZ dapat melakukan kajian mendalam terhadap peran bisnis sehingga semua proses bisnis dapat terdefinisi dengan jelas, dan regulasi dapat dituangkan dalam dokumen SOP. Penggunaan data referensi spesifik untuk setiap domain untuk meningkatkan legitimasi data dan menaikkan tingkat kesesuaian data.
Kata kunci: *total data quality management, service quality assurance, e-commerce*

ABSTRACT

Service Quality Assurance (SQA) is one of a unit in PT XYZ an Indonesian e-commerce company. SQA is responsible for ensuring and maintaining services given to customers by Customer Service agents to meet the company's quality standard. However, the application was built without integration with the Customer Relationship Management (CRM) system, the main customer service application at PT XYZ. The problem is that if companies need SQA data, they must manually input data from SQA to CRM. This can cause vulnerabilities in the data input process. As a result, it impacts the company's business processes, which are inefficient and can be detrimental both materially and immaterially. This research aims to provide recommendations to improve the quality of SQA data so that companies can use the data to be integrated into the CRM system. The Total Data Quality Management (TDQM) method was used for this study with three dimensions: completeness, validity and accuracy. The results of each dimension are 99.64% for completeness, 84.75% for validity, and 100% for accuracy. Business rules on the validity dimension that have the lowest data quality are V1 28.57% and V6 65.11%. The problem factors were identified as an incomplete data dictionary, incomplete and obsolete SOP documents, and no business processes and data control on the application. Recommendations to improve the data quality of the application are PT XYZ can conduct an in-depth study of business roles so that all business processes can be clearly defined and regulations can be set forth in SOP documents. The use of specific reference data for each domain can be used to increase data legitimacy and the level of data suitability.
Keywords: total data quality management, service quality assurance, e-commerce

1. INTRODUCTION

Data is an asset that has a high selling value today. Data can be transformed into information that can be used by companies in making strategic decisions. The quality of decisions must be supported by the presentation of accurate information so that the quality of data is crucial to be maintained and appropriately

managed [1]. Data quality has become a challenge for companies in terms of data management. Data quality is very important to achieve the goals of the company, and if the data quality in a company is poor, it will cause problems that will occur in the future [2]. Poor data quality will disrupt productivity, increase operational costs, affect customer satisfaction levels, and make decisions inaccurate [3]. It is essential to learn the factors of poor data quality and the significance of data quality.

Based on research conducted by [4], data quality problems impact business losses in the United States (US) up to 600 billion dollars per year. Other research conducted by [5] also stated that data quality problems in a retail database cost consumers as much as 2.5 billion dollars annually. That means data quality management is crucial to preventing risks, especially those that impact financial losses.

PT XYZ is an Indonesian technology company engaged in e-commerce. PT XYZ's main business is to provide online buying and selling services that connect sellers and buyers on a C2C (Customer-to-Customer) basis [6]. PT XYZ strives to be a data-driven company, processing data into information to help make decisions and find insights to improve customer service quality. One of the main focuses of PT XYZ is to provide the best service to maintain customer relationships. This responsibility is carried out by the Customer Service Management (CSM) of PT XYZ. Not only maintaining the quality of the services provided, but CSM also performs a quality control function of the processes running on the unit. This function is carried out by Service Quality Assurance (SQA). SQA is a unit that ensures and maintains the services provided to customers by Customer Service (CS). The process is carried out on a Q-Framework information system. SQA will ensure the running process is carried out properly following the applicable SOP. The SQA unit has an application that functions to automate and centralize operational processes. The application is a stand-alone system without any integration with CRM. The process of data input is done manually, and there is no automatic validation process that makes data input errors prone to occur. It becomes a challenge for SQA to ensure the completeness and accuracy of data, particularly in maintaining the quality of reporting and analysis to aid decision-making.

Currently, the system has been running for two years. As long as the system is running, there are no major obstacles in the process. But SQA and several leaders from the CSM division who use the system also experience several challenges. The manual input process provides its own challenges to ensure the correct data is entered. Another challenge is that there is no automatic validation process from the system because there is no direct integration with PT XYZ main system. These challenges can have an impact on decreasing data quality. When there is an input error, it will affect the reporting data and the analysis process up to the decision-making process. There are plans to build a new system integrated with the PT XYZ main system. The system will be integrated in terms of data and architecture so data quality and system performance will improve. Therefore, data quality analysis will be needed to provide some recommendations that can be made during the integration process.

The previous study [7] explained that this study aimed to determine the dimensions, assessment, and analysis of data quality in tertiary institutions with case studies at the Statistics Institute. This paper analyses data using the Total Data Quality Management (TDQM) framework. However, it has limitations in its research, namely, research is only carried out up to the stage of analyzing the causes of data quality problems in the company's core business and measurements are carried out only in submitting guarantee certificates and some related main data. Research on the application of the TDQM method has also been carried out [8]. The research explains that a company's success in increasing profits and managing the risk of loss is largely determined by data. Good data quality can improve the quality of decision-making at the top management level. The limitations of this research are that the research is only carried out up to the stage of analyzing the causes of data quality problems in the company's core business, and measurements are carried out only in the process of submitting guarantee certificates and some related master data.

In a different case study, research [9] explains that a company's success in increasing revenue and managing risk of loss depends on data. High-quality data results in good, quality decision-making at the top management level. The data quality assessment results show four main causes of BCP data quality problems and five recommendations for prevention and improvement. There are limitations in this study, namely this research was carried out only up to the stage of the strategy for improving data quality in a process-driven manner, so further stages were needed in the TDQM implementation process. According to [10] poor data quality can have a negative impact on organizations, such as poor decision-making and planning. Based on these problems, this paper presents and analyzes a case study developed by a government agency, BPS Kabupaten Kaur. For analysis, a data quality maturity model is used to measure the implementation of data quality management in organizations. However, this study has several limitations, namely, the research was only conducted in one office, in one district, and the data collection method used a questionnaire without direct interviews and observation.

ISSN: 2337-7631 (printed)
ISSN: 2654-4091 (Online)

Based on the problems and research that has been done before, the researchers used the TDQM method to measure the quality of the data whose output is to provide recommendations to the top management of PT XYZ SQA unit to ensure maximum data quality to increase customer loyalty and be able to use it for decision making. It is also hoped that the results of this study can be continued to conduct an assessment of data quality in other processes at PT XYZ. This research is the first case study conducted in the field of service quality assurance at PT XYZ using the TDQM method.

2. THEORY AND METHODOLOGY

SQA application

The SQA application is an information system aimed at centralizing data and automating the reporting and data analysis process carried out by the SQA unit. The processes in the system include sampling, scoring, coaching, finding, and interaction. Currently, the application does not automatically integrate with the CRM system, so data is entered manually.

Data quality

Data can be high quality if it meets the aspects its users will manage. The data must be accurate, according to the time needed, relevant, complete, accountable, and understandable [11]. In other words, the user's aim and goal significantly impact the data quality. DAMA International 2010 says that a data quality measurement process is required to set data quality expectations, measure data quality improvements, and help find the factors of data quality problems. To assess the data quality, a sequence of activities needs to be carried out that combines the definition of the data itself and the applicable business rules [12].

Data quality measurement

Measurement of data quality is inseparable from data quality. Data quality management includes planning, implementing, and controlling data activities supported by data quality methodologies [12]. Methodologies that can be used to measure data quality include TDQM, which consists of four stages. The stages of the TDQM methodology include defining the dimensions of data quality, measuring data quality based on a predetermined dimension matrix, analyzing measurement results to identify the main problems in the data, and finally, improving data quality based on the results of the analysis [13]. The TDQM methodology can be used to measure data quality in all processes within the organization [14].

Data quality dimension

Data quality dimensions need to be defined by a company to analyze and measure data quality. By definition, the data quality dimension is a characteristic of the data quality required by data users. Data users will always be interconnected with the data model and the value of the data [12]. Companies must understand the context of their business needs before determining what quality dimensions are needed. The benefit of companies using existing quality dimensions is finding out the impact of poor data quality on compliance with company regulations, reputation, costs and so on [14]. Variable dimensional data that can be used to analyze and measure data quality are accuracy, completeness, accessibility, timeliness and consistency. Accuracy can show the extent to which the data is correct, reliable and can be presented in real terms. Completeness means that the attributes in the data must meet certain specific values in the data. Accessibility means the extent to which information in data can be accessed easily and is always available. Timeliness refers to the user's expected time in accessing data and data availability. Consistency is the extent to which the value of the data presented remains consistent when accessed anytime and anywhere. Some of these points can be used as metrics to measure the dimensions of data quality [15].

Methodology comparison

Several methodologies can be used to assess data quality [16]. Based on the comparison of the methodology in Table 1, the researchers chose to use the TDQM method because it is very practical to be applied in improving data quality on an ongoing basis [17]. The phases of the TDQM include defining data quality dimensions (define), assessing data quality based on a predetermined dimensional matrix (measures), analyzing assessment results to find the significant issues in the data (analyze), and the final step is to improve data quality according to the outcome of the analysis (improved) [13]. The TDQM methodology can be used to assess data quality through the process in the organization [14].

Research instruments

This study uses qualitative methods with interview techniques to identify business processes and business rules that are running, followed by quantitative methods to measure data quality. Interviews were conducted with 3 respondents, are an SQA manager, a lead analyst, and an SQA staff. Interview questions consist of 12 questions which are divided into 3 parts, are 3 questions about business processes, 5 questions about systems, and 4 questions about data. Interview was done by google meet and recorded.

ISSN: 2337-7631 (printed)
ISSN: 2654-4091 (Online)

Table 1. TDQM, DQA, and AIMQ comparison

Parameters	TDQM	DQA	AIMQ
Acronym	Total Quality Data Management	Data Quality Assessment	A method for evaluating the quality of information
Components	The cycle is defined, measured, analyzed, and improve Focus: product information	Subjective and objective data quality evaluation, root cause analysis, comparative analysis, and action for improvement	Have three main elements: product service performance model for information quality (PSP/IQ), based on a questionnaire to measure dimensions, benchmarking and role gap analysis.
Type of data	Structured and semi-structured	Structured	Structured, semi-structured and unstructured
Assessment and development	Comprehensive from the standpoint of implementation	Disassociating between subjective and objective quality indicators	It is a subjective form of evaluation that lacks a framework for improvement.

Data collection

For data quality analysis, the data used is 3-month period data from August 1, 2021, to October 30, 2021. There are 11 tables, 85 columns, and 147,521 rows of data. The data is processed by querying directly into the internal database of the application system. To measure data quality, researchers conducted Structured Query Language (SQL) queries directly to the application database followed by conducting more interviews to analyze the outcomes of data quality assessment.

Table 2. Column and table definition

Table Name	Total of Columns	Total of Rows
users	10	1,392
users account	4	539
mapping agent	8	3,846
sampling	7	19,787
individual scoring	18	20,087
parameter score	7	301,903
case type	4	418
agent mistake	4	1,138
finding	8	1,183
conversation	7	1,874
coaching	8	167
Total	85	147,521

Based on the results of the interview data used (Table 2), the data only covered the main business processes, out of 22 tables only 11 tables are relevant to the process in our research. That includes users, users account, mapping agent, sampling individual, scoring, parameter score, case type, agent mistake, finding, conversation, coaching. This data will be used in analyzing the completeness of the dimensions.

While the data that has been collected for the dimension of validity is in the form of coded rules, namely every sampling with a CE score = 0 will enter the finding category (V1), each finding with a CE score = 0 will enter coaching (V2), the cumulative score of each sampling must be between 0-100 (V3), QA must sample the tickets handled by the CS agent according to the defined agent mapping (V4), the username in the users account table must contain the email registered in the users table (V5), tickets that are sampled must match the tickets in the CRM system (V6), sampling duration must be in the same week from Monday to Sunday for Early ticket, Middle Ticket, and End Ticket (V7). In determining the accuracy of the data, PT XYZ uses the following rules ACC_Score = (60% x CE) + (40% x NCE) (AC1), each sampling with a CE score < 60 is categorized as a defect (AC2).

Process data

The results of the interviews show that to carry out quality control every week, the QA will take sample data every Monday-week on the Zendesk and third party dashboards which contain interactions made by agents in handling customers. The sample is in the form of a transaction id which will later be entered into the Q-framework and analyzed. The data that has been analyzed by the QA will be assessed according to the SOP and categorized if the agent's score is below 100 then it will enter the training process.

ISSN: 2337-7631 (printed)
ISSN: 2654-4091 (Online)

Analysis data

Data analysis carried out in this study used SQL by querying directly to the database and analyzing each attribute using Microsoft Excel. For completeness, data is complete when all required data values are available. In addition, data must can represent a null value because in some cases the data may have no value which is related [18]. Incomplete data occurs when a null value is set for data that is supposed to have value. This indicates that the quality assessment process the data must be able to identify the cause of the null values found in the dataset before assessing the completeness of the data [18]. While validity, the data obtained is compared with the rules that have been set by the company. Rule is coded and queries directly to the database using SQL. The results of the query process are obtained if it is not 100% then there are attributes whose data is not valid. For accuracy dimension, based on the interview results obtained two rules in determining the accuracy of the data. The rules are given a code and then the query process to the database. If the results show below 100%, the data on these attributes can be categorized as inaccurate.

3. RESULTS AND DISCUSSION

Results

The research was conducted by triangulation, that is interviews for data collection and testing the data by querying directly on the company database. The stages of the study were carried out using the TDQM method. From the 6 dimensions of data quality in the DAMA Book, it was determined by respondents 3 the quality of the data. Because only 3 dimension needed by the company for the data quality assessment process, namely completeness, validity, accuracy. These three dimensions are considered to represent the quality of the data currently used.

Completeness dimension

The measurement of the completeness dimension was carried out on 11 tables taken from the application database. In this dimension, the percentage of data that is not null in each table is carried out.

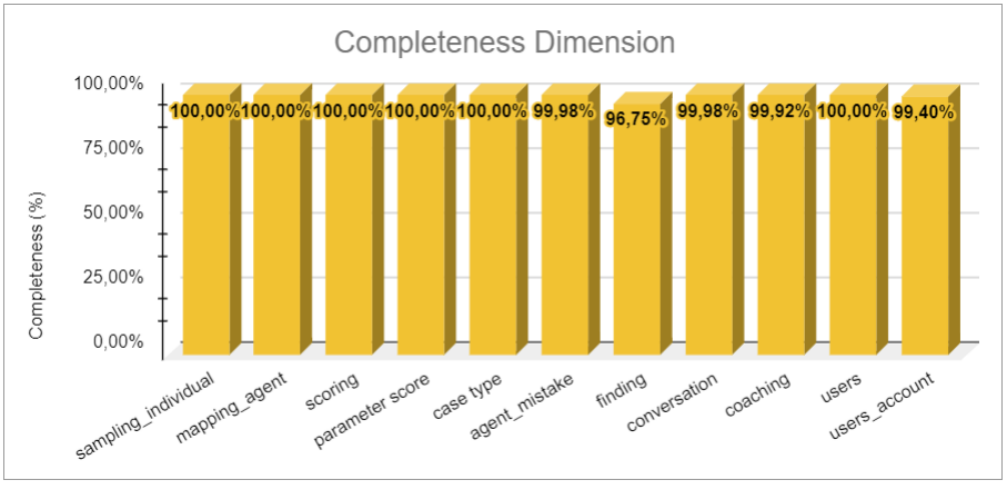

Figure 1. Completeness of data attributes percentage

Results in figure 1 show that in the completeness dimension there are 6 tables with a completeness value of 100%, which are the user table, sampling_individual, mapping_agent, parameter_score, scoring, and case_type. There are 5 tables with completeness values above 90%, that is the users_account table with a value of 99.40%, the agent_mistake table with a value of 99.98%, the finding table with a value of 96.75%, the conversation table with a value of 99.98%, and the coaching table with value 99.92%. On average, the completeness dimension has a value of 99.64%, which means that the overall data shows that all tables are filled according to the completeness of each table.

Validity dimension

The validity dimension is carried out by comparing the current business rules or SOPs with the processed data. Based on the results of interviews, obtained 7 applicable rules are every sampling with a CE score = 0 will enter the finding category (V1), each finding with a CE score = 0 will enter coaching (V2), the cumulative score of each sampling must be between 0-100 (V3), QA must sample the tickets handled by the CS agent according to the defined agent mapping (V4), the username in the users_account table must contain the email registered in the users table (V5), tickets that are sampled must match the tickets in the CRM system (V6), sampling duration must be in the same week from Monday to Sunday for Early ticket, Middle Ticket, and End Ticket (V7).

Validation results using queries directly to the database in accordance with these rules can be seen in figure 2. The average value of the validity dimension is 84.75%. It was identified that there was data that had the lowest validity value, that is the V1 rule of 28.57%, which means that not all ticket sampling data that has a score of 0 enter into finding. There is a ticket sampling that is not from the CRM system but from another system which is a third-party system that is not included in the results of interviews and documentation so it is not included in our analysis. There is also a V6 rule of 65.11% which means there is ticket sampling that is not found in the CRM system. It was found that the ticket sampling contained the transaction number or testing data, not the ticket number. With the validity value, it will certainly have an impact on invalid reporting results.

Accuracy dimension

Based on the results of interviews, obtained 2 rules that apply are the accumulated score must meet the function ACC_Score = (60% x CE) + (40% x NCE) (AC1), each sampling with a CE score < 60 is categorized as a defect (AC2).

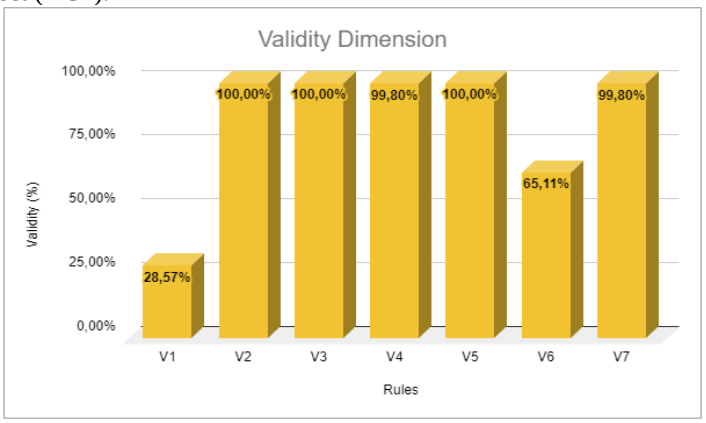

Figure 2. Validity of data attributes percentage

In contrast to the previous dimensions, the results on the accuracy dimension of both the AC1 rules and the AC2 rules both meet the accuracy criteria with an accuracy value of 100% can be seen in figure 3. This is because the calculation process has been automated by the system to reduce errors in calculations that can reduce accuracy. It can be concluded that the level of data accuracy in the two rules is quite perfect because the data input process and the data results are following the rules that have been set.

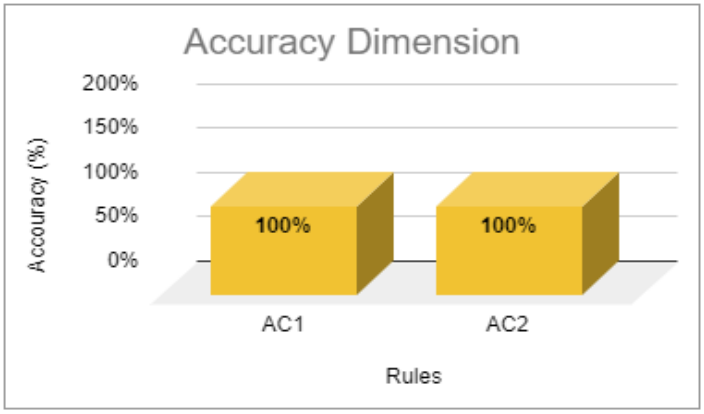

Figure 3. Accuracy of data attributes percentage

Discussion

Based on the measurement results found factors causing data quality problems. The analysis process was carried out by interviewing the SQA staff and QA leads. The following factors were identified and can be made to improve data quality. For completeness dimension, on average completeness has a value of 99.64%, which means that overall, the data shows that all tables are filled according to the completeness of each table. Incompleteness does not reach 100% because there are some QA staff who do not complete the table provided. As research conducted by [19] states that the accuracy of the data and the weak participation of experts is a problem in this dimension. Research [20] provides recommendations to carry out regular reminders that are required for QA staff to complete data input.

Besides that, based on the results of the completeness dimension in Figure I, it shows that the completeness of the data in conducting quality analysis on agents is very good and the QA team follows

ISSN: 2337-7631 (printed)
ISSN: 2654-4091 (Online)

the instructions for filling in the data carefully. Even though there are tables that are not filled in completely because the requested data is optional or not required to be filled in or the fields already include other tables. Research [21] explains that the completeness of the data can also be based on identifying all existing data, documentation, and codes. One approach to data completeness is to add information or instructions for using data such as SOPs and describing data in the repository system.

Repository is a data storage place that contains various data. The diversity of knowledge and expertise possessed by repository management staff becomes very important, the data that must be managed is very diverse Knowledge of metadata also needs special attention, not everyone have the ability to understand various standard metadata. Research [22] mentions completeness of metadata is needed to provide information related to data saved. Standard metadata that is generally used for data exchange it contains title, creator, description, subject, date, publisher, and contributor.

Assessment of data quality depends on the understanding of a deep data curator defines a data that includes accuracy, relevance, representation, and accessibility. This is an obstacle in ensuring the quality of published data through the repository system. Therefore, increasing the competence of human resources regarding the assessment of data quality should be of particular concern. Data curation activities not just checking metadata, complete supporting data, instructions for use data, as well as legal aspects. However, research [22] state it is necessary to understand in substance, so that the data is managed and published can be reused and generate new data (reproduce).

Based on the result of validity, V1 rule of 28.57% is ticket sampling data that has a score of 0 enter into finding. V6 rule of 65.11% is ticket sampling that is not found in the CRM system. Ticket sampling that is not from the CRM system but from another system which is a third-party system. Therefore the value of the two rules is low because there is data that is pulled separately from the system being analyzed. Futhermore, the problem is data dictionaries and system documentation are incomplete and not up to date. Such as nullity information and documentation of business rules running on the system. This makes the analysis must be done manually as well as confirmation and validation of data to subject matter experts. Some business rules that are running have not all been documented and regulated in the SOP so errors can occur when inputting data such as incorrectly inputting the ticket number into a transaction number. There is no automatic data control process because the integration is done manually, such as the ticket data validity control that is sampled, which must be contained in the CRM system. In addition, there is no data identifier to separate test data.

Validity can conduct in-depth studies on the business process that is running in the application and update SOPs and system documentation regularly, add a data identifier for data testing, carry out data control where the database in the CRM system becomes a reference to ensure legitimacy and improve data conformity. Create tools to measure data quality on a regular basis and determine the minimum threshold that meets business expectations as a control for monitoring data quality.

The results on the accuracy dimensions of both rule 1 and rule 2 both meet the accuracy criteria with an accuracy value of 100%. This is because the calculation process has been automated by the system thereby reducing errors in calculations which can reduce accuracy. It can be concluded that the level of data accuracy in the two rules is perfect, because in the data entry process and the results of the data are in accordance with predetermined rules. Maintaining the current level of accuracy which has reached 100%. Existing automation processes can simplify and improve data accuracy. However, it is still necessary to back up such as manual processes in accordance with SOPs so that the maximum level of accuracy can be achieved.

The research is expected to have an impact on related parties. For research side is adding new insights regarding data quality in e-commerce in Indonesia so that it is hoped that it can be applied and become a reference for other e-commerce. For company side is provide evaluation results and recommendations for improving data quality. So that before the integration process can do a good data quality screening before automating data integration between Q-Framework, Zendesk and the call center. In addition, PT XYZ can conduct data quality assessments in other processes.

4. CONCLUSION

The assessment of data quality at PT XYZ is carried out using the TDQM methodology with data quality measures on the aspects of completeness, validity, and accuracy. On the overall dimensions of completeness, the average has reached a perfect or near-perfect value. Each dimension which is 99.64% for the completeness dimension, 84.75% for validity, and 100% for accuracy. However, there are two rules on the validity dimension which have a low value, that is the V1 rule of 28.57% because there is a sampling ticket data that has a CE score = 0 but does not enter the finding process. Then the V6 rule is 65.11% because not all sampling tickets come from the CRM system, but from another system which is a third-party system that

ISSN: 2337-7631 (printed)
ISSN: 2654-4091 (Online)

is not included in the results of interviews and documentation, so it is not included in our analysis. In addition, testing data was found where there is currently no identifier to separate testing data. There are also ticket data that contains transaction numbers that are not regulated in the SOP.

From the results of observations and data quality problems factor analysis, several factors were obtained which are data dictionary and incomplete documentation, incomplete and up-to-date SOPs, and no control processes on business processes in the system. Therefore, our recommendations are given as a possible strategy for improving the data quality of the application. It is expected to be a suggestion for PT XYZ, especially the SQA unit to add more value to data as a company asset.

REFERENCES

[1] C. Arthur, "Tech Giants May Be Huge, But Nothing Matches Big Data," *Technology*, 2019. [Online]. Available: https://en.wikipedia.org/wiki/Clive_Humby#cite_note-10. [Accessed: 02-Mar-2023].

[2] P. Woodall, A. Borek, and A. K. Parlikad, "Data quality assessment: The Hybrid Approach," *Information & Management*, vol. 50, no. 7, 2013, doi: https://doi.org/10.1016/j.im.2013.05.009.

[3] D. Loshin, "Business Impacts of Poor Data Quality," *Pract. Guid. to Data Qual. Improv.*, pp. 1–16, 2011, doi: https://doi.org/10.1016/b978-0-12-373717-5.00001-4.

[4] A. Ohlsson, T. Leeke, J. Watson-MacDonell, and S. Forbes, "Data Quality Assessment for Perinatal Events in an Online Obstetrical Database (O-DB) and a Health Records Database (HR-DB)," *Pediatr. Res.*, vol. 41, pp. 207–207, 1997, doi: https://doi.org/10.1203/00006450-199704001-01246.

[5] L. English, "Information quality management: The next frontier," *DMReview Mag.*, p. 2, 2000, [Online]. Available: https://www.dmreview.com/article_sub_articleId_2073.html. [Accessed: 25-Mar-2023].

[6] PT Bukalapak, "https://www.bukalapak.com/terms#introduction," 2021. [Online]. Available: https://www.bukalapak.com/terms#introduction. [Accessed: 25-Mar-2023].

[7] W. Wijayanti, A. N. Hidayanto, N. Wilantika, I. R. Adawati, and S. B. Yudhoatmojo, "Data quality assessment on higher education: A case study of institute of statistics," *ISRITI*, 2018, doi: https://doi.org/10.1109/ISRITI.2018.8864476.

[8] W. A. Bowo, A. Suhanto, M. Naisuty, S. Ma'mun, A. N. Hidayanto, and I. C. Habsari, "Data Quality Assessment: A Case Study of PT JAS Using TDQM Framework," *ICIC*, 2019, doi: https://doi.org/10.1109/ICIC47613.2019.8985896.

[9] I. D. Andini, Y. Ruldeviyani, A. H. Maulana, and A. Hidayat, "Penilaian Kualitas Data Broadband Customer Profiling (BCP) Pelanggan Fixed Broadband PT Telekomunikasi Indonesia Tbk.," *Iptek-Kom*, vol. 22, no. 1, 2020, [Online]. Available: https://202.89.117.136/index.php/iptekkom/article/viewFile/2916/1361.

[10] D. Plotkin, "Data Stewardship: An Actionable Guide to Effective Data Management and Data Governance", 2014, doi: https://doi.org/10.1016/b978-0-12-410389-4.00007-6.

[11] L. Jiang and J. Zhao, "An Empirical Study On Risk Data Quality Management," *ICIII 2012*, vol. 1, pp. 511–514, 2012, doi: https://doi.org/10.1109/ICIII.2012.6339714.

[12] M. Mosley, *The DAMA Guide to The Data Management Body of Knowledge*, First Edition, 2009.

[13] M. Bertoni, "A case study on the analysis of the data quality of a large medical database," *DEXA*, pp. 308–312, 2009, doi: https://doi.org/10.1109/DEXA.2009.82.

[14] A. Esposito, *Data Quality (Concepts, Methodologies and Techniques)*. 2006.

[15] A. Lucas, "Corporate data quality management: From theory to practice," *5th Iber. Conf. Inf. Syst. Technol.*, pp. 1–7, 2010, [Online]. Available: https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=5556606&isnumber=5556594. [Accessed: 25-Mar-2023].

[16] C. Batini, C. Cappiello, C. Francalanci, and A. Maurino, "Methodologies for data quality assessment and improvement," *ACM Comput. Surv.*, vol. 41, no. 3, 2009, doi: https://doi.org/10.1145/1541880.1541883.

[17] R. Vaziri, M. Mohsenzadeh, and J. Habibi, "TBDQ: A pragmatic task-based method to data quality assessment and improvement," *PLoS One*, vol. 11, no. 5, 2016, doi: 10.1371/journal.pone.0154508.

[18] M. Bovee, R. P. Srivastava, and B. Mak, "A conceptual framework and belief-function approach to assessing overall information quality," *Int. J. Intell. Syst.*, vol. 18, no. 1, 2003, doi: 10.1002/int.10074.

[19] M. M. C. Francisco, S. N. Alves-Souza, E. G. L. Campos, and L. S. De Souza, "Total data quality management and total information quality management applied to costumer relationship management," *ACM Int. Conf. Proceeding Ser.*, 2017, doi: 10.1145/3149572.3149575.

[20] I. Nurrohmah, M. Ayu, A. Dewi, and N. Sahadi, "Measuring the e-Government Maturity in Indonesia using the Ranking of e-Government of Indonesia (PeGI)," *ASRJETS*, vol. 32, no. 1, pp. 49–63, 2017. [Online]. Available: https://core.ac.uk/reader/235050280. [Accessed: 25-Mar-2023].

ISSN: 2337-7631 (printed)
ISSN: 2654-4091 (Online)

[21] L. Peer, A. Green, and E. Stephenson, "Committing to Data Quality Review," *Int. J. Digit. Curation*, Vol. 9, No. 1, pp. 263–291, 2014, doi: 10.2218/ijdc.v9i1.317.
[22] S. Riyanto, E. Marlina, H. Subagyo, H. Triasih, and A. Yaman, "Metode Penilaian Kualitas Data Sebagai Rekomendasi Sistem Repositori Ilmiah Nasional," *BACA*, Vol. 41, No. 1, 2020, doi: 10.14203/j.baca.v41i1.544.

ISSN: 2337-7631 (printed)
ISSN: 2654-4091 (Online)