Strategies to Improve Data Quality Management Using Total Data Quality Management (TDQM) and Data Management Body of Knowledge (DMBOK): A Case Study of M-Passport Application

Rina Rahmawati[1*], Yova Ruldeviyani[2], Puja Putri Abdullah[3], and Fathurahman Ma'ruf Hudoarma[4]
[1−4]Faculty of Computer Science, Universitas Indonesia
Jawa Barat 16424, Indonesia
Email: [1]rina.rahmawati11@ui.ac.id, [2]yova@cs.ui.ac.id, [3]puja.putri@ui.ac.id,
[4]fathurahman.maruf@ui.ac.id

***Abstract*—M-Passport is a mobile application developed for Indonesians to request for passport online. The applicants independently input all required data using this application, so the quality of data entered must be considered to ensure the passport's validity as an official state document. However, input errors increase the time needed for the interview process and make the data verification procedure inefficient. The research aims to assess the data quality of M-Passport for organizations to take deliberate actions to enhance the data quality. The research applies the Total Data Quality Management (TDQM) method and the Data Management Body of Knowledge (DMBOK). Six data quality dimensions are used. It consists of completeness, validity, accuracy, timeliness, uniqueness, and consistency. The measure phase is carried out on 17 entities in the M-Passport database through a query process in the production environment. Then, the analysis phase observes the problems based on the pre-determined dimensional classification groups. The result indicates that the average values of completeness, validity, accuracy, consistency, timeliness, and uniqueness are 99.20%, 99.41%, 100%, 90.68%, 78.52%, and 99.98%, respectively. According to the findings, timeliness and consistency are the lowest dimensions in fulfilling business rules. It indicates that organizations need to focus more on improving data quality in these dimensions. Then, based on the DMBOK, the research also generates recommendations for resolving technical and operational issues.**

***Index Terms*—Data Quality Management, Total Data Quality Management (TDQM), Data Management Body of Knowledge (DMBOK), M-Passport Application**

Received: March 13, 2022; received in revised form: July 13, 2022; accepted: July 15, 2022; available online: March 17, 2023.
*Corresponding Author

I. Introduction

As a member of the International Civil Aviation Organization (ICAO), Indonesia must adhere to the stipulated rules regarding travel documents, such as passports. It is issued to the citizens, permitting them to travel to other nations while remaining valid for a certain period [1]. A passport is officially used to identify an individual when carrying out activities both in and outside the country.

An electronic-based government is one of the efforts to improve public service quality [2]. The Directorate General of Immigration (DGI) introduced the M-Passport applications on September 6, 2021, at three immigration offices in South Jakarta, Central Jakarta, and Tangerang. However, it was intended to be expanded throughout the country on Immigration Service Day, held on January 26, 2022. This process aids applicants in applying for a passport by inputting data and uploading scanned files. In the M-Passport application, some features that accommodate the diverse stages, such as data entry, national identity validation, and billing code collection have already been carried out face-to-face at the immigration office. Applicants can also make scheduled visits to continue the biometric collection and interview process. Afterward, the immigration officer prints the passport document based on the entered data.

Data quality has been a topic of interest for many years, but it is becoming more important as science and practice move forward. High-quality data are valuable asset and a competitive strategy [3]. Irrespective of the fact that problems are inescapable, it is far simpler to

gauge and avoidable by understanding its definition [4]. Knowing data quality issues enables organizations to consider whether the problems identified affect their businesses [5]. It also results in inaccurate information and poor performance. Weaknesses in data quality can also result in significant losses [6]. The organization's management is essential, especially for government agencies [7]. Data quality measurement can also be performed on a large scale, not only in Database Management System (DBMS) [8, 9]. In the case of M-Passport, the quality of the entered data must be considered to ensure the validity of the passport as an official state document. Errors tend to increase the duration of the interview process and make the e-government seem inefficient.

Several obstacles are also encountered during the implementation of the M-Passport. Based on the report sent to the DGI, constant queues at the immigration office are caused by several registered applicants who have made payments and still have to wait for their data to appear in the passport back-office system to move to the biometrics stage. Another obstacle is that during the interview process, it turns out that the uploaded documents are not transferred to the passport back-office system. Therefore, they must be re-scanned as it prolongs the service process. These challenges are also directly reported by the public through email and DGI's official social media, especially the inability to download a cover letter from the immigration office after making payment. The root cause of these constraints is the improperly managed data.

Effectively managed data enable these organizations to achieve their strategic goals. Reversely, they are bound to encounter several obstacles. As stated in The Leader's Data Manifesto, fundamental and lasting change requires the leadership's involvement at every level [10]. It is one of the most significant changes in determining the strategies used by organizations to manage and improve their data quality. The Data Management Body of Knowledge (DMBOK) ensures that the acquired information fits consumers' needs. Managerial activities include defining high-quality data, strategy, and scope of the initial assessment. After identifying these elements, an initial assessment is performed, which aids in identifying and prioritizing the improvements needed. Finally, data quality is developed and becomes operational.

The Total Data Quality Management (TDQM) method is commonly used to measure data quality. This approach is popularly used in various fields despite several new methods for measuring data quality. For example, it has been adopted in the financial sector [11], healthcare sector [12], educational sector [13], and industrial sector [14].

Based on the problems mentioned, measurement or assessment of data quality needs to be carried out at the beginning of the implementation to evaluate the readiness of Indonesians to use e-government services and the system to serve mobile users throughout the country. Data quality measurement is conducted by accessing the M-Passport. Hence, the organization can immediately improve factors with a negative impact. The research is carried out based on the TDQM framework designed by previous research [15]. It consists of four phases: define, measure, analysis, and improve. Compared to other models, TDQM is selected based on the keywords and the type of measurement to answer research questions. The research also combines TDQM and DMBOK for the improvement phase to ensure that the data fit the users' needs. The research questions are how the data quality is on the M-Passport application and what strategy the organization should adopt to improve the data quality.

II. Literature Review

A. Data Quality

In every organization, managing data quality is necessary. The right decisions are made with accurate results available to the users at the appropriate time [16]. Data quality refers to its characteristics and the capacity to satisfy stated requirements. It is defined as information that is "fit for use" by its users [17]. The first step in managing data quality is to extract it from sources that need to be checked [18]. There are several reasons for encouraging active participation in efforts to maintain high-data quality, as follows [3].

1) It is a valuable asset.
2) It tends to increase customer satisfaction.
3) It increases revenue and profits.
4) It can be perceived as a competitive strategy or advantage.

B. Data Quality Dimension

According to DAMA [10], the data quality dimension is a measurable feature. Furthermore, data quality is defined, measured, improved, monitored, and managed using dimensions [19]. These dimensions must be directly related to the potential risks in the critical process. Organizations need to define characteristics that are relevant to business processes, and these are known as business rules to measure data quality. These principles define business processes and rules of action [19], and data are one of the outputs. Its violation implies that the business process does not function properly. These rules are also used as input to load data quality [19].

Three approaches (theoretical, empirical, and intuitive) are adopted to propose a comprehensive definition of dimension [17]. The theoretical approach employs a format model in its definition. The empirical approach defines dimensions based on data that are important to consumers through experiments, interviews, and questionnaires. The intuitive approach defines it according to practical experiences and an understanding of essential attributes. This method allows each researcher to select the most relevant features to the objectives.

The Data Management Association UK claims that organizations usually measure their data quality using six key dimensions [4]. It can be seen as follows.

- Completeness: The percentage of data stored versus the maximum possible extent [20].
- Validity: whether the data value is consistent with the specified domain [10].
- Accuracy: The degree to which the data is precise, reliable, and error-free [17].
- Timeliness: Whether current data are promptly made available to users [16].
- Uniqueness: No entity appearing more than once in the data set [10].
- Consistency: The degree to which data are always presented in the same format and must be compatible with previous ones [17].

C. Data Quality Measurement

Data quality measurement is a procedure that checks whether information concerning business rules and standards is acceptable for the intended use [20]. Its advantage is tangible proof of the issues that underlie the business needs and detect data quality issues [19]. Data quality measurement is carried out to analyze the specific causes of errors generated. It will determine the relevant actions that need to be taken to eliminate its sources, erase incorrect information, and prioritize efforts included in the project plan [21].

Here are some methodologies that can be applied in various business environments:

1) Task-based Data Quality Method (TBDQ)
   TBDQ is a process-driven data quality method that specifically helps organizations where employees play an essential role, directly or indirectly, in creating or manipulating information. TBDQ focuses on structured data, and it has two main phases, namely assessment (planning and evaluation) and improvement (evolution and execution) [7]. TBDQ uses a questionnaire survey and a simple ratio to measure data quality [22].
2) Data Quality Assessment (DQA)
   DQA is the evaluation of the acquired data in comparison to the established goals [23]. It comprises three steps: carrying out data quality assessments, comparing the results, and taking necessary actions for improvement. DQA uses stakeholders' expectations and quantitative metric definitions for data quality measurement [22].
3) Total Information Quality Management (TIQM)
   TIQM is a method used for establishing quality information with continuous improvement by applying effective management principles. It divides aspects of data quality into two categories, inherent and pragmatic, in contrast to several other frameworks. TIQM also employs user expectations and data quality metric definitions during measurement [22].
4) Total Data Quality Management (TDQM)
   TDQM aims to provide users with high-quality information to facilitate the implementation of a policy that covers its entirety [15]. Its cycle involves continuous definition, measurement, analysis, and improvement. It also applies business rules considerations and the definition of data quality metrics in measurement [22]. The research applies TDQM because it can be used to answer the research questions, including the suitability of the measurement type. Figure 1 shows that its methodology consists of four stages.

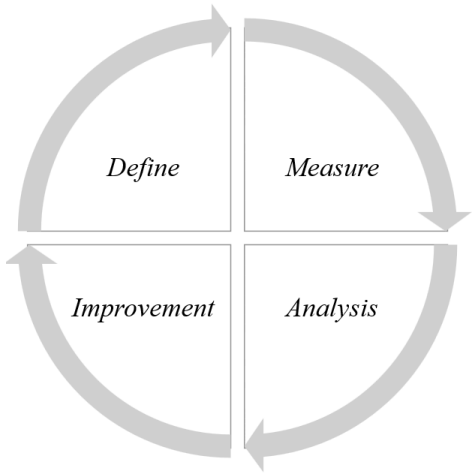

Fig. 1. Total Data Quality Management (TDQM) phase.

D. Data Management Body of Knowledge (DMBOK)

DMBOK is a data management guideline designed by DAMA International [10]. It is related to activities covering several knowledge areas, such as data quality management. The planning, implementation, and control of operations by applying effective managerial approaches to guarantee that the information is fit for

consumption and satisfies the demands of users are known as data quality management. It comprises six main dimensions: completeness, uniqueness, timeliness, validity, accuracy, and consistency [4].

III. Research Method

The research uses both primary and secondary data. Primary data are collected through direct observations and interviews with three respondents. The interview is conducted with the sub-coordinator of the Application Development of DGI. The essence is to select tables needed to measure M-Passport data quality. Business rules are determined for each of them concerning the dimensions of completeness, validity, accuracy, timeliness, uniqueness, and consistency. Furthermore, a representative of the M-Passport development team is also interviewed to ascertain the validity of the findings. The person in charge of detecting technical and operational problems is also interviewed. Then, direct observation is carried out by accessing the database from the M-Passport application. Data quality is analyzed using the query through HeidiSQL in the production environment. Specifically, for the consistency and timeliness dimensions, there is a need to compare the passport data system with the immigration office database. The three largest immigration offices, namely South Jakarta, Central Jakarta, and Tangerang, which become the M-Passport pilot project, are taken as data sampling.

Meanwhile, secondary data is obtained through literature studies from books, journals, proceedings, and relevant sources to measure data quality. Data processing is carried out quantitatively in the measurement process and qualitatively during the result analysis. The research applies the TDQM framework by adopting the data quality problem classification technique, as shown in Fig. 2. Then, from DMBOK, the variables used for measuring data quality are completeness, validity, accuracy, timeliness, uniqueness, and consistency.

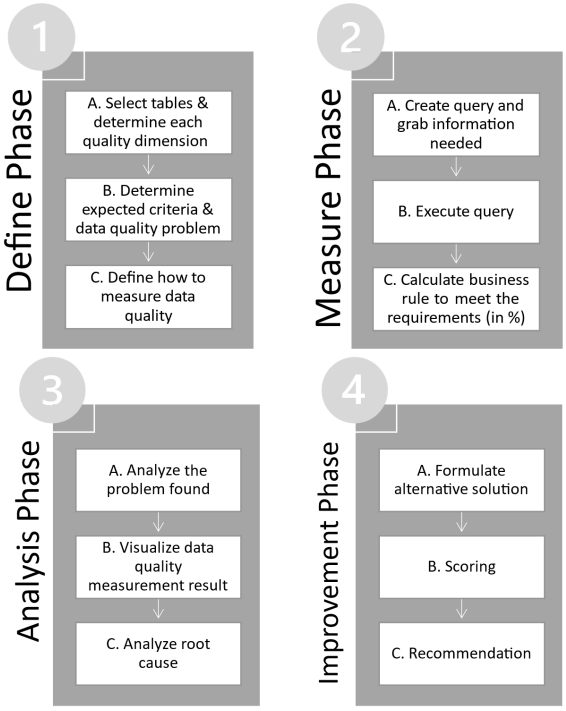

Fig. 2. Research method used.

A. Define Phase

As shown in Fig. 2 in point 1, the research commences with the selection of tables, dimensions, and criteria measured through interviews with the organization's Sub-coordinator of Application Development. Meanwhile, the selection of the table is determined based on the types (master and transaction tables) used in the functional application process. Afterward, the expected quality dimensions of each measured table are determined. It is conducted using an empirical approach, such as interviews, to determine the organization's needs. Completeness, validity, accuracy, timeliness, uniqueness, and consistency are the data quality dimensions employed in the research. Finally, it is necessary to ascertain the expected criteria and data quality problems that emerge from each dimension.

B. Measure Phase

The data are measured in a way that has been defined in the previous stage. In Fig. 2 in point 2, the results are obtained using a query in the application database. Specifically for the timeliness and completeness dimensions, measurements are also made by accessing each office's server to retrieve information on data availability. In this phase, the results are used to obtain the percentage of business rule criteria that can or cannot be met.

C. Analysis Phase

Data analysis is carried out by describing the specific problems of each measured attribute, as shown in Fig. 2 in point 3. The results are used to construct a bar chart for visualization, and the average information that meets the business rules in each dimension is used to determine the data quality. The business rules are represented in each dimension by a bar on the chart. Meanwhile, those dimensions that are 100% fulfilled are visualized in orange, while those depicted in blue are not fulfilled. Then, the problems detected are cross-checked with the application developer to ensure the

TABLE I
TABLES FOR MEASUREMENT.

Table Code	Table Description
TB01	Submitted service request
TB02	Applicant's list
TB03	Identity of each applicant
TB04	Applicant's relative
TB05	Required document in the application
TB06	Documents uploaded by the applicant
TB07	Master of work unit details
TB08	Immigration office quota master per day
TB09	Master code for each work unit
TB10	Immigration office quota master per session
TB11	User's master data
TB12	User's master data
TB13	Application to be sent to back office
TB14	Billing transactions to be sent to the back office
TB15	Payment component
TB16	Applicant's relatives to be sent to the back office
TB17	Applicant's identity to be sent to the back office

validity of the findings. It leads to the root cause analysis and further groups into the problem category.

D. Improvement Phase

At the end of the TDQM cycle, as shown in Fig. 2 in point 4, alternative solutions are used to resolve the problems detected. These solutions are formulated with a process-driven strategy that focuses on improving certain techniques in data creation to ensure better quality. The scoring method determines which solution is most likely to be adopted by the organization in the short term.

IV. RESULTS AND DISCUSSION

The process of measuring data quality is carried out in accordance with the four phases in the TDQM framework, which are described as follows.

A. Define Phase

The significant entities used to assess the data quality are identified in this phase. It includes 17 of 54 entities in the M-Passport system database. The selection of tables and columns (attributes) is based on the analysis process of each function and usability in the application. The selected table, as shown in Fig. 2 in point 1A, is codified and described in Table I.

The research employs six dimensions to measure data quality: completeness, validity, accuracy, consistency, timeliness, and uniqueness from DMBOK. The expected business rules or provisions, the classification of problems in each dimension, and the method of measuring each are shown in Fig. 2 in points 1B and 1C. It can also be seen in Table II.

The null value is one of the common data quality checks [4]. For completeness, expected criteria refer to a similar method used previously [11, 12], namely some important attribute that has value. Then, for validity, expected criteria refer to problem types in previous research [24], such as inexactness of timestamps, typing mistakes, values outside the domain range, and violation of mutual dependency. The inexactness of timestamps implies that it is recorded imprecisely. However, the research is slightly modified as if it does not match the sequence of activities.

Moreover, for accuracy, the expected criteria refer to rules used previously [12] that the data must be realistic. For instance, the national identity number needs to have 16 numeric characters and gender only written as male or female. The research also includes some strict criteria, such as the national identity number that must be validated with the web service. Next, for consistency, the expected criteria are similar to the rules used in previous research [7], stating that data need to have the same format and should be compatible with previous ones. The checking process for inconsistency is carried out by comparing the data field format with the determined one [5]. The business rules in the research mandate that the immigration back-office database is compatible with previous ones in M-Passport.

Furthermore, for the timeliness, the expected criteria defined [8] stipulate that the integration of data records needs to be within the permissible time range. Timeliness concerns the data's relevance to being sufficiently up-to-date [7]. In reference to the definition mentioned earlier, the research defines business rules using the timeliness dimension that text and file data must be sent to the back office before the applicant's arrival. Lastly, the expected criteria are defined in previous research [12] for uniqueness, indicating that the data should not be duplicated. It means that information can be uniquely identified. Uniqueness can be ensured by enforcing a primary key constraint [5].

B. Measure Phase

In this phase, 17 entities in the M-Passport database are measured according to the classified problems concerning 6 dimensions selected based on the suitability of attributes in the case study. This process is carried out by querying access to the system's production environment. It seeks to determine whether the data on the system have fulfilled or failed to meet the business rules of each defined dimension. The research adopts Eq. (1) [9, 25]. It shows N_i as a number of incomplete/invalid/inaccurate/ unavailable/duplicate/inconsistent values, N_t as a total number of values, and Δ as a metric associated with the dimensions (%). Based on the results, some business

TABLE II
RULES OF DATA QUALITY DIMENSIONS AND METHODS OF MEASUREMENT OF DATA PROBLEMS.

Expected Criteria	Data Quality Problem	Measurement Method
Completeness Columns are complete if they are filled	Missing value	Query data containing NULL or blank strings in all columns that cannot be empty.
Validity (A) The activity timestamp is valid if it is in accordance with the sequence of activities.	Invalid timestamps of sequential activities	Query for timestamps columns that need to have sequential values according to each other.
(B) The value of the ID reference column is valid if it matches the reference table data.	Outside domain range	Query for foreign key columns with table references. The calculated data does not meet validity assuming the data are obtained outside the reference table.
(C) The column value is valid if there is no typo	Typing mistakes	Query to determine a value that has no meaning according to the value that should fill the column.
(D) Column values that depend on other columns are valid if there are no contradictory values.	Mutually dependent attributes having contradicting values	Query data between contradicting columns
Accuracy Column values are accurate if they correspond to real entities	Not representing real-life entities	Identity number, name, and date of birth that have been validated with the national identity web service.
Timeliness Data are the timeliness of the available information following the expected time.	Data not available in the expected timeframe	Text and file data that are sent to the back office before the scheduled arrival.
Uniqueness Data are unique if there is no duplication of data	Data duplication	No entity that is saved more than once. Identity number, name, date of birth, arrival schedule, and service code are identified as duplicate entities.
Consistency Column values are consistent across systems	Data inconsistency	Comparison of data values from M-Passport and back-office database in an immigration office

rule criteria can be met. The details are explained in the analysis phase.

$$\Delta = 1 - \frac{N_i}{N_t} \tag{1}$$

C. Analysis Phase

As shown in Fig. 2 in point 3, the measurement is continued by conducting the analysis. It involves describing and evaluating the problems detected based on the pre-determined dimensional classification groups. Then, each business rule is visualized as a bar on the chart. Interviews are also held with the development team to ensure the validity of the problems and to find the root cause, which is codified with the technical problem (TP) and operational problem (OP) code.

The measurement of the completeness dimension aims to determine the entirety of the data from the attributes expected to be in the M-Passport. The criteria are presumed to be complete if the provisions of the business rules fill the column. Based on the analysis, as many as 81 mandatory attributes are selected based on the urgency in explaining the application type, the need for a passport, personal, and Sistem Informasi PNBP Online (SIMPONI) payment data. This billing system facilitates the payment of non-tax state income in Indonesia. If an attribute includes null or simple spaces, it is grouped as one that does not meet the criteria of a business rule.

The 81 mandatory attributes used are from 12 entities consisting of TB01 to TB12, as shown in Table III. Examples of mandatory attributes in TB01 are the submission category (C01A1), office (C01A3), and submission date (C01A4). This mandatory attribute is defined as a business rule. For the completeness dimension, with a total of 81 business rules, the average percentage of business rules that fulfill the requirements is 99.20%. Figure 3 shows that some attributes have not met the 100% expected value. It includes C01A6/information on cover letter documents generated by the system (67.15%) due to a scheduler bug (TP1) and C02A8/billing code (99.99%) and C02A9/ billing rate (99.99%) due to a scheduler bug in processing payment billing (TP2). Then, it also has C02A11/application code (99.99%), C06A2/link of the uploaded document (99.93%), C07A11/main office setting (69.23%), and C11A4/user password (99.89%) due to system bug in storing application data (TP3).

The validity dimension is measured if there are anomalies resulting from attributes not consistent with the stated domain value. Based on the analysis, 66 attributes meet the validity dimension and are also defined as business rules. For the validity dimension with 66 business rules, the average percentage that fulfills

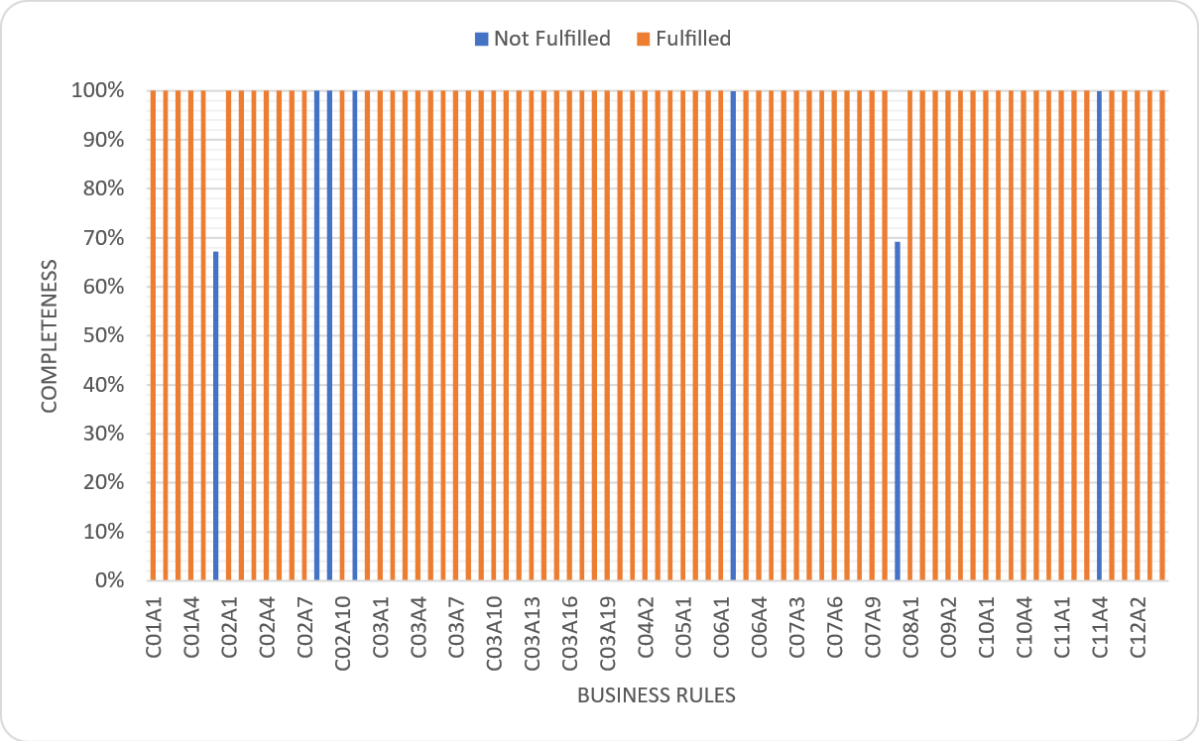

Fig. 3. Completeness dimension chart.

TABLE III
RESULT OF COMPLETENESS MEASUREMENT.

Table	Column	Sum of Business Rules	
		Fulfilled	Not Fulfilled
TB01	C01A1-C01A6	5	1
TB02	C02A1-C02A12	9	3
TB03	C03A1-C03A20	20	
TB04	C04A1-C04A4	4	
TB05	C05A1-C05A3	3	
TB06	C06A1-C06A4	3	1
TB07	C07A1-C07A11	10	1
TB08	C08A1-C08A2	2	
TB09	C09A1-C09A4	4	
TB10	C10A1-C10A6	6	
TB11	C11A1-C11A5	4	1
TB12	C12A1-C12A4	4	
Total		74	7
Average		99.20%	

TABLE IV
RESULT OF VALIDITY MEASUREMENT.

Table	Column	Sum of Business Rules	
		Fulfilled	Not Fulfilled
TB01	C01B1-C01B4	4	
TB02	C02B1-C02B11	6	5
TB03	C03B1-C03B15	12	3
TB04	C04B1-C04B5	3	2
TB05	C05B1-C05B3	3	
TB06	C06B1-C06B5	5	
TB07	C07B1-C07B2		2
TB10	C10B1-C10B4	3	1
TB11	C11B1-C11B7	5	2
TB13	C13B1-C13B5	5	
TB14	C14B1-C14B4	3	1
TB15	C15B1	1	
Total		50	16
Average		99.41%	

the requirements is 99.41%, as shown in Table IV. The provisions of the existing business rules in the validity dimension are as follows.

- Timestamps of Sequential Activities (A): The timestamps on attributes must align with the activities' sequence.
- Outside Domain Range (B): The id value for certain attributes must match the one in the reference table.
- Typing Mistakes (C): there are no typographical errors on certain attributes.
- Mutually Dependent Attributes Having Contradicting Values (D): the value of certain attributes is similar to the others.

The following are the analysis results of each attribute that does not fulfill the validity criteria, as depicted in Fig. 4. In criteria A, C02B1/billing code expiration date (99.98%) fails to meet the ex-

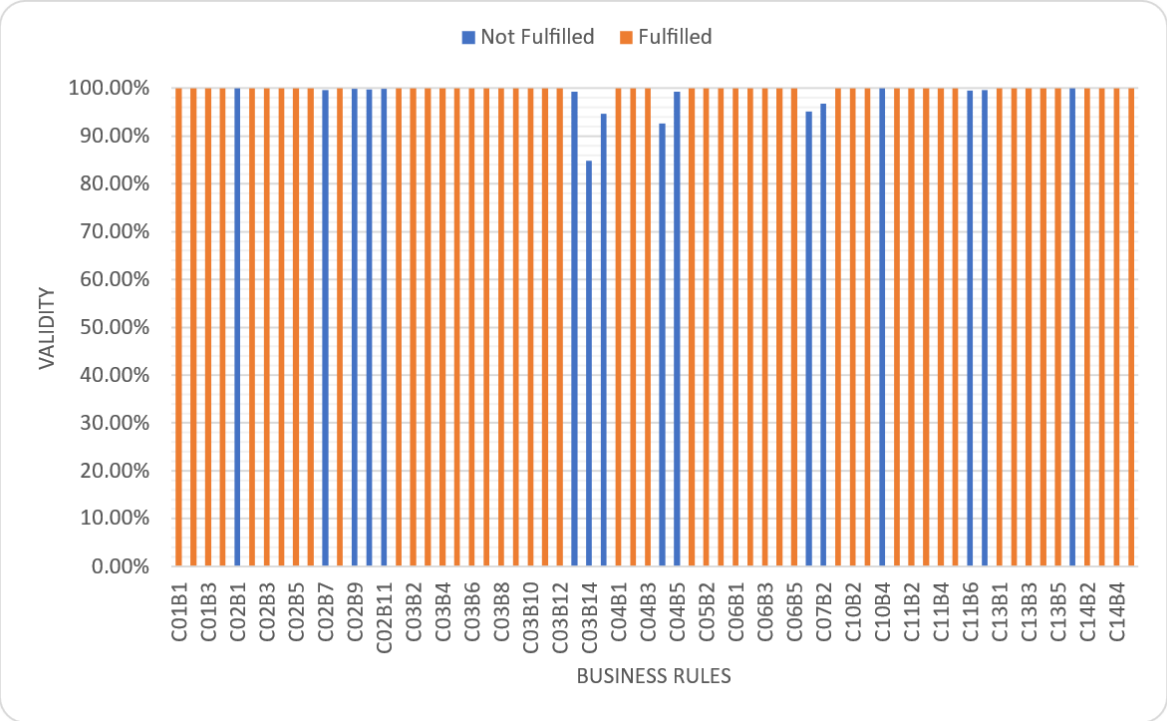

Fig. 4. Validity dimension chart.

pected value of 100% because the timestamp is not in order of occurrence. Meanwhile, in criteria B, C07B1/office id (95.12%), C07B2/main office id (96.75%), and C14B1/bank id (99.97%) do not meet the expected value of 100% because some values do not match the reference table. Next, in criteria C, attributes of C03B13/applicant's phone number (99.31%), C03B14/Ministry of Religion recommendation number (84.80%), C03B15/labor identity number (94.70%), C04B4/emergency contact number (92.66%), C04B5/emergency contact name (99.32%), C11B6/user email (99.50%), and C11B7/user phone number (99.67%) fail to meet the expected value of 100% because the data entry method is in the form of free text. The application is unable to control the input. Last, in criteria D, C02B7/booking status (99.64%), C02B9/sync status (99.90%), C02B10/document link (99.70%), C02B11/document key (99.83%), and C10B4/remaining quota (99.99%) do not meet the expected value of 100%, due to the contradictory data. These problems occur due to a system bug validating the inputted data (TP4).

The accuracy dimension is measured to guarantee that the data stored in the system are accurate, consistent, and unambiguous. The criterion is the degree of truth of the applicant's data validated with national identity as reference information. Measurements in this dimension depend on the applicant's main data, namely name (C02C1), date of birth (C02C2), and national identity number (Nomor Induk Kependudukan (NIK)) (C02C3). NIK is a national identification number identifying a person as an Indonesian citizen. Attributes that meet this accuracy dimension are defined as business rules. During its measurement, real-time validation is carried out with the NIK web service when applying for a passport. For the accuracy dimension of three business rules, all fulfill and score 100%. The results, which are in the form of a dimensional accuracy chart, are shown in Fig. 5.

The consistency dimension measurement aims to ensure no inconsistency between the M-Passport database and the immigration back-office selected by the applicant. The selected attributes included in the business rules of this dimension are TB17 and TB18. However, these attributes can be changed by the verificator during the interview process. Based on the analysis, 28 attributes meet the criteria of the consistency dimension, as shown in Table V. These attributes are also defined as business rules.

For the consistency dimension of 28 business rules, the average percentage is 90.68%. Figure 6 shows that the business rules that fail to meet the

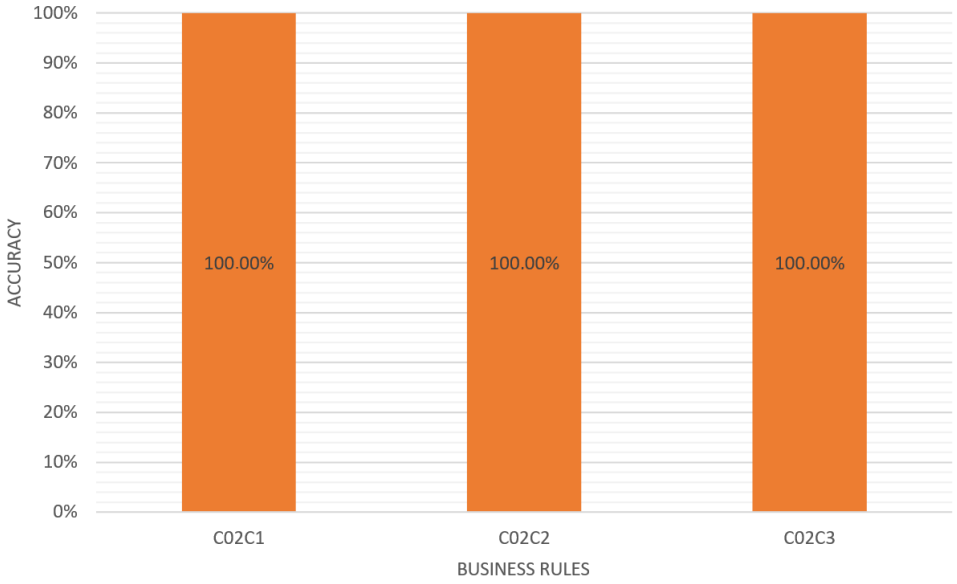

Fig. 5. Accuracy dimension chart.

TABLE V
RESULT OF VALIDITY MEASUREMENT.

Table	Column	Sum of Business Rules	
		Fulfilled	Not Fulfilled
TB16	C16D1-C16D5		5
TB17	C17D1-C17D23	4	19
Total		4	24
Average		90.68%	

value expectation of 100% are C16D1/family address (77%), C16D2/family gender (46%), C16D3/family name (46%), C16D4/family civil status (46%), C16D5/family nationality (98%), C17D1/applicant's official address (98%), C17D2/applicant's domicile address (98%), C17D3/applicant's email (99%), C17D5/applicant's passport name (86%), C17D6/applicant's full name (84%), C17D8/applicant's phone number (94%), C17D9 (91%)/applicant's occupation, C17D10/applicant's national identity issued date (94%), C17D11/applicant's place of birth (92%), C17D13/applicant's official district (99%), C17D14/applicant's domicile district (99%), C17D15/applicant's official subdistrict (99%), C17D16/applicant's domicile subdistrict (99%), C17D17/applicant's official zip code (99%), C17D18/applicant's domicile zip code (99%), C17D19/applicant's official province (99%), C17D20/applicant's domicile province (99%), C17D21/applicant's civil status (99%), and C17D23/applicant's birth date (99%). During the interview process, the information is cross-checked before it is updated. It tends to occur due to an error when filling in the data (TP5). The column that is changed the most by the verificator is the applicant's family information.

The measurement of the timeliness dimension aims to ensure that information is available in accordance with the time expectations. It is carried out based on the arrival schedule, such as the composite attribute date of quota and opening hours for the interview session selected by the applicant. In measuring this attribute, the timestamp on one of the database tables in the immigration back-office selected by the applicant is compared to the passport issuance office (data text/C13E1).

Additionally, a check is also made on the file uploaded by the applicant (data file/C13E2). For the timeliness dimension with two business rules, the average percentage is 78.52%. Figure 7 shows the attributes that do not fulfill the expected value of 100%. It has C13E1 (78.37%) and C13E2 (78.67%). It is due to the data synchronization module's failure to send requests to the destination immigration office (TP6).

The measurement of the uniqueness dimension aims to ensure no duplicate data in an application transaction. The selected attributes are name, date of birth, national identity number, quota date, opening hours, and service code. The measured business rule is in one application, and these six attributes need not be duplicated. The percentage that fulfills its requirements

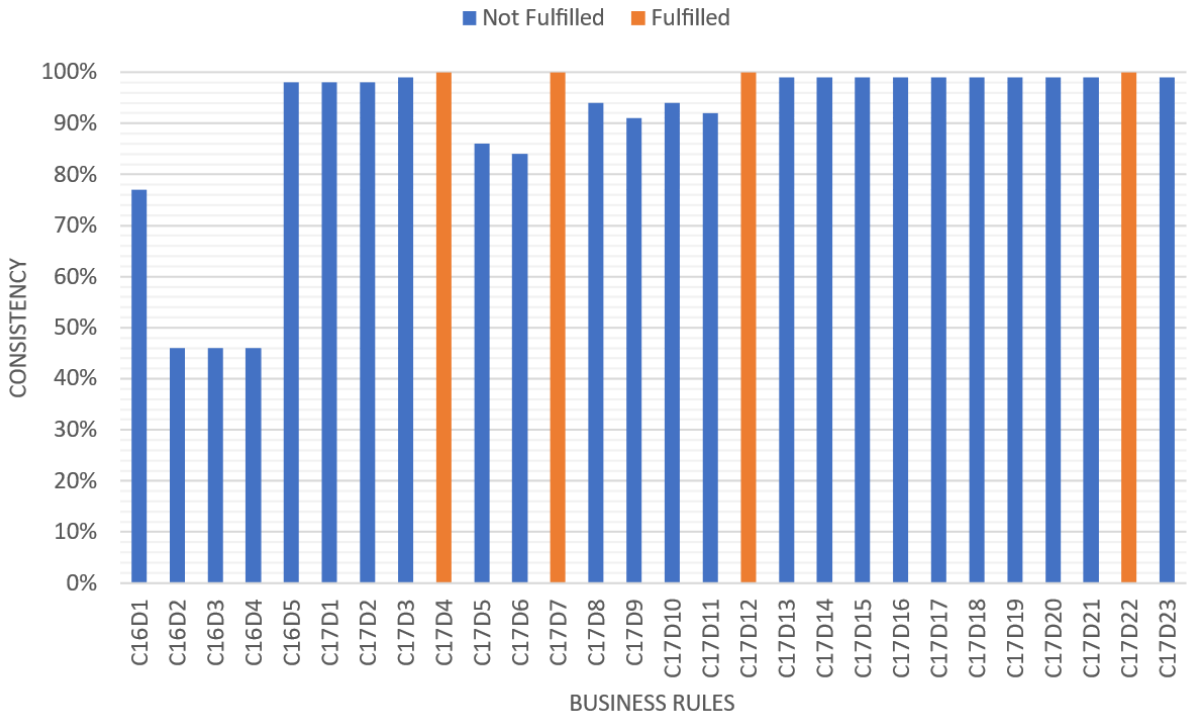

Fig. 6. Consistency dimension chart.

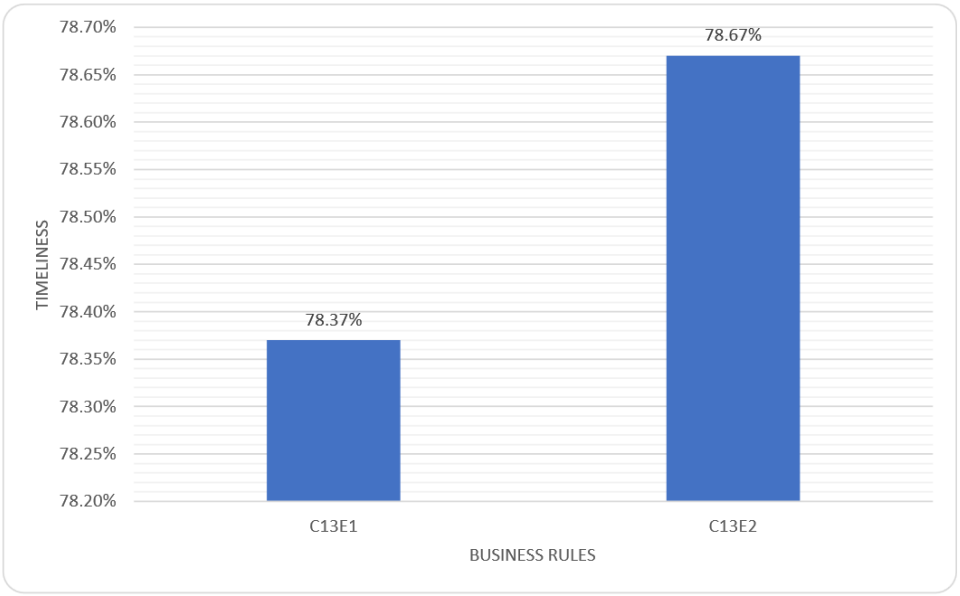

Fig. 7. Timeliness dimension chart.

is 99.98% due to application bugs in processing applications, resulting in data duplication (TP7). The results are represented in a graphic dimension in Fig. 8.

Table VI is a summary of the data quality measurement. The average percentage of business rules that meet the requirements of the completeness, validity, accuracy, consistency, timeliness, and uniqueness dimensions are 99.20%, 99.41%, 100%, 90.68%, 78.52%, and 99.98%, respectively.

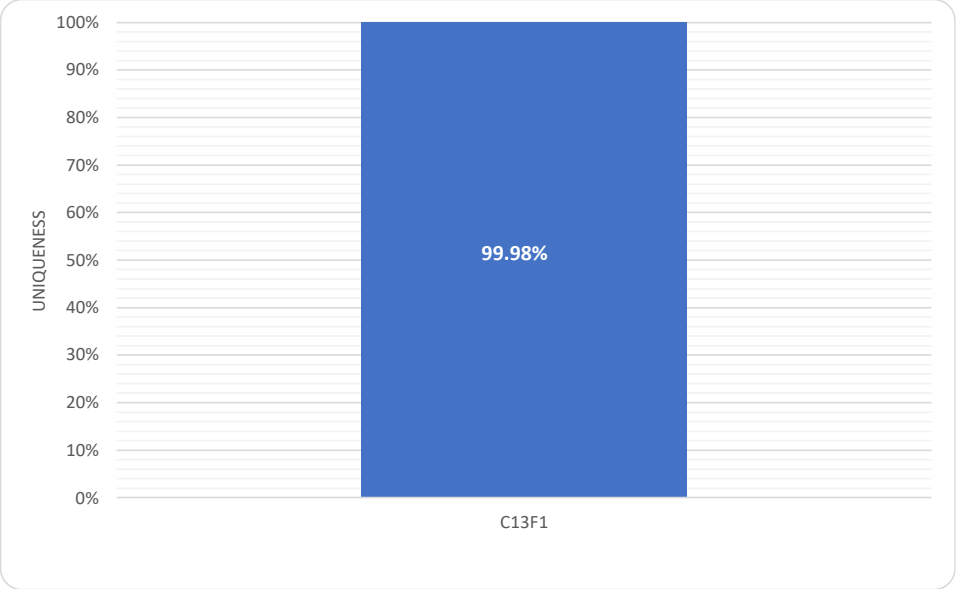

Fig. 8. Uniqueness dimension chart.

TABLE VI
DATA QUALITY MEASUREMENT SUMMARY.

Dimension	Business Rules	Average (%)
Completeness	81	99.20
Validity	66	99.41
Accuracy	3	100.00
Consistency	28	90.68
Timeliness	2	78.52
Uniqueness	1	99.98

D. Improvement Phase

After obtaining the analysis results from the measurement stage, the research proceeds to the improvement phase. As explained earlier in Fig. 2 in point 4A, the initial stage involves formulating problems detected both from the technical and operational management sides based on the measurement results, interviews, and observations in the case study application environment. First, technical data problems are issues encountered on the technical side of data quality obtained from the analysis's results involving the earlier six dimensions. Second, operational data problems refer to issues encountered on the operational side of the data obtained from the interview results with structural officials and staff of the organization and observations in the case study of the application environment.

In the technical problems, there are system bugs in saving data (TP1), processing payment billing (TP2), generating cover letter to immigration office (TP3), validating input data (TP4), and sprocessing the application (TP5). User errors can also occur in filling in data (TP6). Then, there is a failure in data synchronization module to send the request file to the destination immigration office (TP7).

In the operational problems, the applicant cannot apply due to down storage in the application (OP1). The inability to generate a cover letter for the immigration office is detected after receiving a report from the applicant. Data synchronization failure is detected after the applicant arrives for an interview due to incompetence in monitoring the data synchronization module (OP2). All parties, both staff and vendors, can access and make changes to data in the production database. However, no log records data changes (OP3). In addition, M-Passport has no disaster recovery plan to anticipate force majeure (OP4).

The existing technical and operational problems are formulated, which can be seen in more detail in Table A1 in Appendix. Table A1 explains how to improve data quality according to activities in the data quality knowledge area in DMBOK for each root cause. Activities carried out in improving data quality in DMBOK are defining high quality data, defining strategy, defining scope of initial assessment, performing initial assessment, identifying and prioritizing improvement, and developing and deploying data quality operations.

The research continues with the selection of solutions using the scoring method as explained in Fig. 2 in point 4B by analyzing the impacts, costs, and risks involved, as shown in Table VII. The following are impact, costs, and risk assessment criteria: 1 for the

TABLE VII
DATA SOLUTIONS SCORING.

Root Cause	Impact	Cost	Risk	Score
Technical				
TP1, TP2, TP3, TP4, TP5, TP6	5	5	5	15
TP7-a	5	3	4	12
TP7-b	5	5	4	14
Operational				
OP1	5	5	5	15
OP2	5	5	5	15
OP3	4	4	4	12
OP4	4	1	5	10

lowest and 5 for the highest. After calculating the score on the solution selection, the possible solution for technical problems is obtained by improving the application code. For operational data problems, the best solution is to monitor the data synchronization module and application infrastructure.

With the various alternative solutions explained in Fig. 2 in point 4C, developers need to make improvements to the application's frontend source code to ensure that the critical data that the applicant must fill in are completely filled (TP3), ensure that the timestamp logic of the application matches the order in which they occur (TP4), and guarantee that the id reference attribute has a value according to the reference table (TP4). They also need validation and instructions to minimize errors in filling out free text fields (TP6).

On the application's backend, developers must improve the application logic to prevent duplicate requests (TP5). Developers must also improve the payment scheduler to enable the submitted applications to receive billing payments immediately (TP2) and improve the scheduler to generate cover letters so that applicant who has paid will receive a cover letter from the immigration office (TP1).

From the operational side, monitoring must be carried out on the M-Passport infrastructure to enable the easy handling of any form of storage problem and prevent failure during the application process (OP1). Monitoring of the data synchronization module and the scheduler must also be conducted to ensure application failures in sending data, issuing billing codes, and generating cover letters immediately before receiving a report from the applicant (OP2). Meanwhile, the M-Passport applicant is expected to fill in the actual data to prevent the application from experiencing obstacles during the verification process. Providing incorrect information is an immigration violation, as appointed in Article 126(c) of Law no. 6 of 2011 concerning Immigration [1], which will result in rejection and non-refundable payment.

According to the research findings, the TDQM method is capable of exploring problems in applications related to data quality. Meanwhile, DMBOK can be used to develop strategic solutions that leverage data quality management strategies to ensure that data fulfills the needs of consumers. Furthermore, organizations are expected to implement these recommendations to improve the quality of their data and turn it into a useful and valuable asset.

V. CONCLUSION

In conclusion, the quality of the inputted data must be considered to ensure the validity of passport data as an official state document. It is important to note that input errors will increase the time needed for the interview process. Hence, the data verification procedure becomes inefficient. The research objective is to assess the data quality of M-Passport for an organization to take deliberate actions to enhance data quality.

Data quality is assessed in the research to determine the M-Passport application's issues. The research is conducted using the TDQM method with six dimensions of data quality in DMBOK. It consists of completeness, validity, accuracy, timeliness, uniqueness, and consistency. After all process has been conducted, the average measurement value of completeness, validity, accuracy, consistency, timelessness, and uniqueness is 99.20%, 99.41%, 100%, 90.68%, 78.52%, and 99.98%, respectively. The results show that timeliness and consistency are the lowest dimensions in meeting business rules. It indicates that organizations need to focus more on improving the existing data quality in these dimensions. The technical recommendations to improve the application code from the frontend are completeness, validity, and consistency. Those needed for the backend and scheduler are the uniqueness and completeness dimensions. Meanwhile, the best solution for operational data problems is to monitor its sync module and application infrastructure. Then, the formulation of strategic solutions that apply quality management techniques to data is conducted based on activities in the DMBOK.

The strength of the research is the use of the TDQM method which is clearly described in each phase. Combining the improvement phase in TDQM with the data quality management in the knowledge area in DMBOK, the research can be a guide for developing strategic solution recommendations to improve data quality in the case study. The method applied in the research can also be used in other organizations to improve their data quality. The data quality of M-Passport can be improved when the DGI carries out the research recommendations. It will enable the verification process, interviews, and issuance of passport

documents at the immigration office to be faster, more efficient, and less time-consuming.

For academic purposes, this document can be utilized as a resource for developing ways to enhance data quality using DMBOK and TDQM. The research can be sharpened by utilizing the analytical hierarchy method in the scoring process at the improvement phase. Thereby the prioritization of recommendations can be determined more objectively. However, the research is limited by the inability to test the given recommendations. Therefore, further researchers need to be conducted on implementing these recommendations and periodically reassessing the data quality

Acknowledgement

The authors gratefully acknowledged the financial support of the Ministry of Communication and Information Technology of the Republic of Indonesia through the Master Scholarship Program for 2021. The authors were also indebted to the Directorate General of Immigration of the Ministry of Law and Human Rights of the Republic of Indonesia for data collection assistance.

References

[1] Pemerintah Indonesia, "Undang-undang (UU) No. 6 Tahun 2011 tentang Keimigrasian," 2011. [Online]. Available: https://peraturan.bpk.go.id/Home/Details/39140/uu-no-6-tahun-2011

[2] ——, "Instruksi Presiden (INPRES) Nomor 3 Tahun 2003 tentang Kebijakan dan Strategi Nasional Pengembangan E-Government," 2003. [Online]. Available: https://peraturan.bpk.go.id/Home/Details/147277/inpres-no-3-tahun-2003

[3] L. L. Pipino, R. Y. Wang, J. D. Funk, and Y. W. Lee, *Journey to data quality*. The MIT Press, 2016.

[4] B. Moses, L. Gavish, and M. Vorwerck, *Data quality fundamentals: A practitioner's guide to building trustworthy data pipelines*. O'Reilly Media, 2022.

[5] R. Zhang, M. Indulska, and S. Sadiq, "Discovering data quality problems: The case of repurposed data," *Business & Information Systems Engineering*, vol. 61, pp. 575–593, 2019.

[6] J. Bicevskis, Z. Bicevska, A. Nikiforova, and I. Oditis, "An approach to data quality evaluation," in *2018 Fifth International Conference on Social Networks Analysis, Management and Security (SNAMS)*. Valencia, Spain: IEEE, Oct. 15–18, 2018, pp. 196–201.

[7] I. N. P. Pradnyana, D. J. Pradipta, and Y. Ruldeviyani, "Measurement of export data quality using Task-Based Data Quality (TBDQ): Case study of the Directorate General of Customs and Excise," in *2020 12th International Conference on Information Technology and Electrical Engineering (ICITEE)*. Yogyakarta, Indonesia: IEEE, Oct. 6–8, 2020, pp. 114–119.

[8] T. Hongxun, W. Honggang, Z. Kun, S. Mingtai, L. Haosong, X. Zhongping, K. Taifeng, L. Jin, and C. Yaqi, "Data quality assessment for online monitoring and measuring system of power quality based on big data and data provenance theory," in *2018 IEEE 3rd International Conference on Cloud Computing and Big Data Analysis (ICCCBDA)*. Chengdu, China: IEEE, April 20–22, 2018, pp. 248–252.

[9] O. Reda, I. Sassi, A. Zellou, and S. Anter, "Towards a data quality assessment in big data," in *Proceedings of the 13th International Conference on Intelligent Systems: Theories and Applications*, 2020, pp. 1–6.

[10] DAMA International, *DAMA-DMBOK: Data management body of knowledge*. Technics Publications, 2017.

[11] 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," in *2019 Fourth International Conference on Informatics and Computing (ICIC)*. Semarang, Indonesia: IEEE, Oct. 16–17, 2019, pp. 1–6.

[12] D. Y. Siregar, H. Akbar, I. B. P. A. Pranidhana, A. N. Hidayanto, and Y. Ruldeviyani, "The importance of data quality to reinforce COVID-19 vaccination scheduling system: Study case of Jakarta, Indonesia," in *2022 2nd International Conference on Information Technology and Education (ICIT&E)*. Malang, Indonesia: IEEE, Jan. 22, 2022, pp. 262–268.

[13] S. Ma'mun, A. N. Hidayanto, A. S. Alwadain, D. I. Inan, and N. Paoprasert, "Educational data quality management: Lessons learned from a public university in Indonesia," *International Journal of Innovation and Learning*, vol. 30, no. 2, pp. 201–219, 2021.

[14] A. Arief, M. Kartiwi, and I. Jaswir, "Dataanalytics of Fourier-Transform Infrared Spectroscopy (FTIR) for halal and non-halaladulterations," in *2020 8th International Conference on Cyber and IT Service Management (CITSM)*. Pangkal, Indonesia: IEEE, Oct. 23–24, 2020, pp. 1–5.

[15] R. Y. Wang, "A product perspective on total data quality management," *Communications of the ACM*, vol. 41, no. 2, pp. 58–65, 1998.

[16] T. King and J. Schwarzenbach, *Managing data*

quality:A practical guide. BCS Learning & Development Limited, 2020.

[17] R. Y. Wang and D. M. Strong, "What data quality means to data consumers," *Journal of Management Information Systems*, vol. 12, no. 4, pp. 5–33, 1996.

[18] J. Bicevskis, A. Nikiforova, Z. Bicevska, I. Oditis, and G. Karnitis, "A step towards a data quality theory," in *2019 Sixth International Conference on Social Networks Analysis, Management and Security (SNAMS)*. Granada, Spain: IEEE, Oct. 22–25, 2019, pp. 303–308.

[19] D. McGilvray, *Executing data quality projects: Ten steps to quality data and Trusted Information (TM)*. Academic Press, 2021.

[20] M. Mohammed, J. R. Talburt, S. Dagtas, and M. Hollingsworth, "A zero trust model based framework for data quality assessment," in *2021 International Conference on Computational Science and Computational Intelligence (CSCI)*. Las Vegas, USA: IEEE, Dec. 15–17, 2021, pp. 305–307.

[21] O. Azeroual, G. Saake, and M. Abuosba, "Data quality measures and data cleansing for research information systems," *Journal of Digital Information Management*, vol. 16, no. 1, pp. 12–21, 2018.

[22] C. Cichy and S. Rass, "An overview of data quality frameworks," *IEEE Access*, vol. 7, pp. 24 634–24 648, 2019.

[23] M. F. G. León and J. Dewulf, "Data quality assessment framework for critical raw materials. the case of cobalt," *Resources, Conservation and Recycling*, vol. 157, pp. 1–12, 2020.

[24] L. Vanbrabant, N. Martin, K. Ramaekers, and K. Braekers, "Quality of input data in emergency department simulations: Framework and assessment techniques," *Simulation Modelling Practice and Theory*, vol. 91, pp. 83–101, 2019.

[25] M. Yalaoui and S. Boukhedouma, "A survey on data quality: Principles, taxonomies and comparison of approaches," in *2021 International Conference on Information Systems and Advanced Technologies (ICISAT)*. Tebessa, Algeria: IEEE, Dec. 27–28, 2021, pp. 1–9.

APPENDIX

The Appendix can be seen in the next page.

Cite this article as: R. Rahmawati, Y. Ruldeviyani, P. P. Abdullah, and F. M. Hudoarma, "Strategies to Improve Data Quality Management Using Total Data Quality Management (TDQM) and Data Management Body of Knowledge (DMBOK): A Case Study of M-Passport Application", CommIT Journal 17(1), 27–42, 2023.

TABLE A1
TECHNICAL AND OPERATIONAL DATA PROBLEMS AND SOLUTIONS.

Root Cause Analysis	Data Quality Activities with DMBOK					
	I. Define High-Quality Data	II. Define a Data Quality Strategy	III. Define the Scope of the Initial Assessment	IV. Perform Initial Data Quality Assessment	V. Identify and Prioritize Improvement	VI. Develop and Deploy Data Quality Operations
Technical						
[TP1] Completeness: System bug in saving data	Define the impact of completeness of data in making passport issuance decisions	Identify required data, business rules, and data quality standards			Develop deep code repair solutions to ensure data completeness	Fix the code on the frontend
[TP2] Completeness: System bug in processing payment billing	Define the impact of failed applications to get billing payments	Make Service-Level Agreement (SLA) for the issuance of billing codes so that the applicant is not disadvantaged			Develop solutions to ensure applications get billing payments	Fix the code on the payment scheduler
[TP3] Completeness: System bug in generating cover letter to the immigration office	Define the impact of not generating a cover letter to the immigration office	Prepare Standard Operating Procedures (SOP) that make it easier for applicants to come to the office once payment has been received	Define critical data and business rules to be achieved	Measure data quality in the completeness dimension periodically	Develop a solution to ensure the applicant gets a cover letter to the immigration office	Fix the code on the scheduler to generate a cover letter
[TP4] Validity: System bug in validating input data	Define the impact of data validity in making passport issuance decisions	Identify required data, business rules, and data quality standards	Define critical data and business rules to be achieved	Measure data quality in the validity dimension periodically	Develop code repair solutions related to data validity problems	Fix the code on the frontend
Accuracy	No problem found	-	-	-	-	-
[TP5] Uniqueness : System bug in processing application	Define the impact of data duplication on the quality of production data	Identify required data, business rules, and data quality standards	Define critical data and business rules to be achieved	Measure data quality in the uniqueness dimension periodically	Develop code fix solutions related to data duplication problems	Fix the code on the backend
[TP6] Consistency: User error in filling in data	Define the impact of high data inconsistency on the immigration office's operations	Identify required data, business rules, and data quality standards	Define critical data and business rules to be achieved	Measure data quality in the consistency dimension on a periodically	Compile application design improvements that minimize data entry errors	Fix the code to minimize data entry errors
[TP7] Timeliness: Data synchronization module failing to send the requested file to the destination immigration office	Define the impact of delays in sending text and file data on service operations at the immigration office	Create SLA for sending text data and files that must be adhered to so as not to harm the applicant	Define critical data and business rules to be achieved	Measure the timeliness of data transmission on a regular basis; analyze the causes of delays in sending data	Develop preventive solutions, so there are no delays in data transmission	[a] Fix the text and file data synchronization module
					Develop corrective solutions for manual data transmission	[b] Build a manual synchronization module that officer executable
Operational						
[OP1] Disconnected storage from the backend server	Define the impact of the loss of storage on the submission of applications	Set the M-Passport infrastructure to be high-availability; develop SOPs for handling problems	Know the infrastructure needs to ensure the level of application availability	Measure the level of availability to find out the current conditions	Develop solutions to the causes of downtime in infrastructure	Fix the cause of downtime; monitor the infrastructure of the production environment

Root Cause Analysis	Data Quality Activities with DMBOK					
	I. Define High-Quality Data	II. Define a Data Quality Strategy	III. Define the Scope of the Initial Assessment	IV. Perform Initial Data Quality Assessment	V. Identify and Prioritize Improvement	VI. Develop and Deploy Data Quality Operations
[OP2] No available monitoring of the data sync module and document generator scheduler yet	Define the impact of the module being down, stuck, or not running normally	Develop SOPs for handling problems	-	-	Compile troubleshooting solutions for problems on the module	Monitor the data sync module and scheduler. Troubleshoot if there is a problem with the module
[OP3] No available audit trail on the database yet	Define the impact of direct access of the development team and operational team to the database	Data change history needs to be consistently documented	Identify critical data for which changes should be recorded	-	Compile audit trail installation on the database	Install audit trail on database
[OP4] No available Disaster Recovery Plan (DRP) yet	Define the impact in case of force majeure	Create DRP for the M-Passport application	-	-	Arrange activities in DRP for the M-Passport application	Implement activities in DRP for the M-Passport application