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 Rahmawati¹*, Yova Ruldeviyani², Puja Putri Abdullah³, and Fathurahman Ma’ruf Hudoarma⁴
¹−⁴Faculty of Computer Science, Universitas Indonesia
Jawa Barat 16424, Indonesia
Email: ¹rina.rahmawati11@ui.ac.id, ²yova@cs.ui.ac.id, ³puja.putri@ui.ac.id, ⁴fathurahman.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].
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 sources 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.

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.

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

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 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 $\Delta$ as a metric associated with the dimensions (%). Based on the results, some business
rule criteria can be met. The details are explained in the analysis phase.

\[ \Delta = 1 - \frac{N_i}{N_t} \]  

(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 that fulfills 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%), 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
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-
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>
<thead>
<tr>
<th>Table</th>
<th>Column</th>
<th>Sum of Business Rules</th>
<th>Fulfilled</th>
<th>Not Fulfilled</th>
</tr>
</thead>
<tbody>
<tr>
<td>TB16</td>
<td>C16D1-C16D5</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TB17</td>
<td>C17D1-C17D23</td>
<td>4</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td></td>
<td>4</td>
<td>24</td>
</tr>
<tr>
<td>Average</td>
<td></td>
<td></td>
<td>90.68%</td>
<td></td>
</tr>
</tbody>
</table>

The 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
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.
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 processing 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.

<table>
<thead>
<tr>
<th>Root Cause</th>
<th>Impact</th>
<th>Cost</th>
<th>Risk</th>
<th>Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Technical</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TP1, TP2, TP3, TP4, TP5, TP6</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>15</td>
</tr>
<tr>
<td>TP7-a</td>
<td>5</td>
<td>3</td>
<td>4</td>
<td>12</td>
</tr>
<tr>
<td>TP7-b</td>
<td>5</td>
<td>5</td>
<td>4</td>
<td>14</td>
</tr>
<tr>
<td>Operational</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>OP1</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>15</td>
</tr>
<tr>
<td>OP2</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>15</td>
</tr>
<tr>
<td>OP3</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>12</td>
</tr>
<tr>
<td>OP4</td>
<td>4</td>
<td>1</td>
<td>5</td>
<td>10</td>
</tr>
</tbody>
</table>

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, timeliness, 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.

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
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 Customs and Excise of the Republic of Indonesia for data collection assistance.

REFERENCES


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


APPENDIX

The Appendix can be seen in the next page.

<table>
<thead>
<tr>
<th>Root Cause Analysis</th>
<th>Data Quality Activities with DMBOK</th>
</tr>
</thead>
<tbody>
<tr>
<td>Technical</td>
<td></td>
</tr>
<tr>
<td>I. Define High-Quality Data</td>
<td>Define the impact of completeness of data in making passport issuance decisions</td>
</tr>
<tr>
<td>II. Define a Data Quality Strategy</td>
<td>Define completeness of data in getting billing payments</td>
</tr>
<tr>
<td>III. Define Scope of the Assessment</td>
<td>Define critical data and business rules to be achieved</td>
</tr>
<tr>
<td>IV. Perform Initial Data Quality Assessment</td>
<td>Define data and business rules to be achieved</td>
</tr>
<tr>
<td>V. Identify and Prioritize Improvement</td>
<td>Define data and business rules to be achieved</td>
</tr>
<tr>
<td>VI. Develop Deploy Data Quality Operations</td>
<td>Define data and business rules to be achieved</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Operational</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>OP1 Disconnected storage from the backend server</td>
<td>Define the loss of storage on the sub-mission of applications</td>
</tr>
<tr>
<td></td>
<td>Know the infrastructure needs to ensure the level of application availability</td>
</tr>
<tr>
<td></td>
<td>Develop solutions to the causes of downtime in infrastructure</td>
</tr>
</tbody>
</table>

**TABLE A1**

**TECHNICAL AND OPERATIONAL DATA PROBLEMS AND SOLUTIONS.**

<table>
<thead>
<tr>
<th>Root Cause Analysis</th>
<th>I. Define High-Quality Data</th>
<th>II. Define a Data Quality Strategy</th>
<th>III. Define the Scope of Initial Data Quality Assessment</th>
<th>IV. Perform Initial Data Quality Assessment</th>
<th>V. Identify and Prioritize Improvement</th>
<th>VI. Develop and Deploy Data Quality Operations</th>
</tr>
</thead>
<tbody>
<tr>
<td>OP2</td>
<td>Define the impact of the module being down, stuck, or not running normally</td>
<td>Develop SOPs for handling problems</td>
<td>-</td>
<td>-</td>
<td>Compile troubleshooting solutions for problems on the module</td>
<td>Monitor the data sync module and scheduler. Troubleshoot if there is a problem with the module.</td>
</tr>
<tr>
<td>OP3</td>
<td>Define the impact of direct access of the development team and operational team to the database</td>
<td>Data change history needs to be consistently documented</td>
<td>Identify critical data for which changes should be recorded</td>
<td>-</td>
<td>Compile audit trail installation on the database</td>
<td>Install audit trail on database.</td>
</tr>
<tr>
<td>OP4</td>
<td>Define the impact in case of force majeure</td>
<td>Create DRP for the M-Passport application</td>
<td>-</td>
<td>-</td>
<td>Arrange activities in DRP for the M-Passport application</td>
<td>Implement activities in DRP for the M-Passport application.</td>
</tr>
</tbody>
</table>