Educational data quality management: lesson learned from a public university in Indonesia

Syukron Ma'mun* and
Achmad Nizar Hidayanto

Faculty of Computer Science,
Universitas Indonesia,
Kampus UI Depok, 16424, Indonesia
Email: syukron.m@ui.ac.id
Email: nizar@cs.ui.ac.id
*Corresponding author

Ayed Said Alwadain*

Computer Science Department,
Community College,
King Saud University,
P.O. Box 2454, Riyadh 11451, Saudi Arabia
Email: aalwadain@ksu.edu.sa

Dedi Iskandar Inan*

The University of Papua,
Manokwari West Papua, 98314, Indonesia
Email: d.inan@unipa.ac.id

Naraphorn Paoprasert

Department of Industrial Engineering,
Faculty of Engineering,
Kasetsart University,
Bangkok, 10900, Thailand
Email: fengnpp@ku.ac.th

Abstract: The researchers conducted this study at one of the faculty in a public university in Indonesia. The purpose of this study was to evaluate the data quality of the human resource information system (HRIS). The data was analysed based on the Ministry of Research, Technology, and Higher Education Regulation, No. 16, 2016. This study began with a questionnaire to determine the maturity level of the data quality management. Then, to determine whether the HRIS criteria could be measured, the researchers conducted interviews. Based on the analysis, the researchers found that the agreement between the HRIS application and the criteria measures was 36.5% for completeness, 69.5% for validity, and 14.3% for the accuracy dimensions. The last dimension, novelty, could not be determined as no information was available on data change time. The researchers suggested that the faculty need to implement data quality management.

Keywords: data quality management; data quality management maturity level; total data quality management; TDQM; educational data; learning and innovation data quality; Indonesia.

Reference to this paper should be made as follows: Ma'mun, S., Hidayanto, A.N., Alwadain, A.S., Inan, D.I. and Paoprasert, N. (xxxx) 'Educational data quality management: lesson learned from a public university in Indonesia', *Int. J. Innovation and Learning*, Vol. X, No. Y, pp.xxx–xxx.

Biographical notes: Syukron Ma'mun obtained his Bachelor's in Information Technology from the Universitas Mercu Buana and Master's in Information Technology from the Universitas Indonesia. He is working in the education field as a staff. His research interests are related to data quality and information technology adoption.

Achmad Nizar Hidayanto is a Professor in Information Systems at the Faculty of Computer Science, Universitas Indonesia. He is also the Vice Dean for Resources, Venture, and General Administration. He received his Bachelor's, Master's and PhD in Computer Science from the Universitas Indonesia in 1999, 2002, and 2008 respectively. His research interests are related to e-health, information systems/information technology management, ecommerce, e-government, information systems security, and information systems/information technology adoption.

Ayed Said Alwadain is an Assistant Professor at the Computer Science Department at the Community College, King Saud University, Saudi Arabia. He received his PhD in 2014, from the Queensland University of Technology in Australia. In his research, he focuses on enterprise architecture, service management and engineering, business process management, requirement engineering and big data. He has published his work at many international conferences and journals such as *Data & Knowledge Engineering* and the *International Journal of Intelligent Information Technologies*.

Dedi Iskandar Inan is a Senior Lecturer in Information System Field at the Faculty of Engineering, Universitas Papua, Indonesia. He now serves as the Chief of Computer and Information Technology at the Universitas Papua. He completed his PhD in Information System from the University of Technology Sydney, Australia in 2018. His research interests in information system are ranging from mHealth, knowledge management, e-government and technology adoption and its continuance usage. During his PhD, he also worked closely with the State Emergency Service (SES) New South Wales, Australia to harness information system technology to contribute disaster management resilience endeavours. He published and presented in various and reputable information system and disaster management journals, and conferences such as ISF, *IJDRR*, PACIS, ACIS, etc.

Naraphorn Paoprasert is an Assistant Professor at the Department of Industrial Engineering, Kasetsart University, Thailand. She received her PhD from the Department of Industrial Engineering, University of Wisconsin-Madison, USA. Her past research studies have been focusing on decision analysis and game theory, risk analysis, system simulation, process improvement, and economics analysis. She has been teaching for almost six years. She has applied many teaching strategies to improve student learning outcomes.

This paper is a revised and expanded version of a paper entitled 'Data quality assessment on university: a case study on human resources information system at the faculty of psychology XYZ university' presented at 8th Engineering International Conference, Semarang, Indonesia, 16 October 2019.

1 Introduction

For an organisation, data is an asset that provides benefits for business (Wijayanti et al., 2018), digital forms of information are data about information technology as it has stored (D. International, 2017). Based on the purpose of data usage, data quality is the level at which data is based on consumer expectations, and poor data quality can hinder the success of an organisation's business (Loshin, 2011). Bad data will have an impact on data quality and can affect the organisation's technical, operational, and strategy areas. A good data quality cycle has five stages, namely: the goal of the organisation is to understand the impact of weak quality data, take measurements of the quality of the data against the business data that need to determined, build an action plan and improve the quality of the data for the results of both phases, implement an improvement strategy data quality, and monitoring data quality continuously becomes the last phase (Loshin, 2011).

Research on data quality is mostly carried out in various aspects, such as a survey of the dimensions of data quality by Sidi et al. (2013) from University Putra Malaysia about decision making in an organisation based on data and information obtained from data analysis.

An overview of data quality frameworks by Alpen et al. (2019) from Universität Klagenfurt about the importance of achieving and maintaining high data quality standards is widely recognised by practitioners and researchers, from data quality to big data quality by Batini (2015) of the University of Milano-Bicocca about the relationship between data quality and some relevant research coordinates in big data.

Awareness of the importance of maintaining the quality of data held by an organisation becomes the basis for evaluating the quality of data of an organisation. However, in evaluating the quality of data in an organisation, it cannot be generalised the need for quality data from each dimension for each organisation (Sidi et al., 2013). Every organisation should be aware of the importance of always maintaining data quality continuously, and this awareness becomes the basis for evaluating data quality. However, every organisation cannot generalise the needs of data quality from every dimension (Woodall et al., 2013).

These faculty is in a public university environment, which refers to government regulations (Yudhoyono, 2012; Kemendikbud, 2014; Kemenristekdikti, 2016, 2015), which is about the responsibility of each university or organiser of higher education for the completeness, correctness, accuracy and novelty data on the organisation of higher education (Kemenristekdikti, 2016). The rules are binding for the university, so it automatically binds this faculty also to present valuable and quality data following specified criteria. However, based on observations often occur, the case that the data quality is not following the expected standard.

Related research carried out by Herawati (2016) to improve the higher education database (PDDikti) quality, the result that there are problems in PDDikti data quality. Another research carried out by Wijayanti et al. (2018), about data quality assessment with a case study at the institute of statistics. The results in the completeness meet 21.15%, in the validity fulfilled is 57.90%, in the accuracy meet 45.45%. Another research carried out by Sabtiana et al. (2017) about data quality for a case study in BPS, Bengkulu province, needs to improve all areas but more specifically in performance management, technology, and data governance.

These studies encourage the authors to conduct data quality research in one of the faculties from a public university in Indonesia with the requirements expected by PDDikti. The way taken is to assess the quality of the data available in a human resources information system. Focus this research is to assess the quality of the data carried out with a subjective approach to user needs as well as an objective approach. Models built using combinations like this result in better evaluations (Bertoni et al., 2009). The researchers use a total data quality management (TDQM) framework. TDQM provides a general framework to facilitate understanding of the data improvement approach through data quality management. TDQM defines four areas in the data quality challenge, namely: systems, processes, procedures, and policies (Mosley et al., 2010). TDQM is a structured and comprehensive approach to organisational management in improving data quality (Wang, 1998). In improving data quality, TDQM stages include definition, measurement, analysis, and improvement (Wang et al., 2012).

In the first part of this research is an introduction to research, then reviews related study literature, research hypotheses, explains the research methodology, then presents an analysis of the research results, and ends with a conclusion from the results of the study.

2 Literature review

2.1 Data quality

From a research standpoint, various fields, including statistics, management, and computer science, had handled data quality, at the end of the 1960s, preceded by statisticians who investigated issues related to data quality considering duplication of data in statistics. Then a mathematical theory was proposed, in the 1980s, followed by management researchers who focused on how the control system detects and eliminates data quality problems. Early in the 1990s, computer scientists began to consider problems in defining, measuring, and improving the quality of data stored in databases, data warehouses, and legacy systems (Batini and Scannapieca, 2006).

The definition of data quality is data that has quality if it meets the intended use requirements. In other words, the quality of the data depends as much on the intended use as the data itself to meet the intended use. The data must be accurate, timely, relevant, complete, understood, and trusted (Olson, 2003).

2.2 Dimensions of data quality

In carrying out its activities, the organisation needs to define dimensions of data quality in order to be able to measure and analyse data quality. Each organisation determines the dimensions of data quality needed based on the context of its business needs to determine the impact of weak quality data on costs, reputation, compliance with regulations (Wijayanti et al., 2018). The Indonesian Ministry of Education Research, Technology, and Higher Education in Indonesia establishes the dimensions of data quality in higher education are completeness, truth, accuracy, and novelty (Kemenristekdikti, 2016).

2.3 Data quality measurement

Data quality measurement is part of data quality management where the data quality methodology itself consists of two activities, namely assessment, and improvement. Use a data quality methodology to support data quality management consists of the planning, implementation, and control activities of data (D. International, 2017). According to Bertoni et al. (2009), in TDQM, methodological activities related to data quality consist of four stages. Namely: defining data quality dimensions, measuring data quality based on a defined dimension matrix, analysing measurement results to identify the source of problems in the data, and the last is to improve data quality based on the results of the analysis (Bertoni et al., 2009). TDQM approaches in quality management in all processes in the organisation (Batini and Scannapieca, 2006).

2.4 Comparative perspectives

In using data quality measurement methodology, the authors compare several methodologies (Batini et al., 2009; Chichi and Rass, 2019) as in Table 1.

Table 1 Compare TDQM and DQA

Criteria	*TDQM*	*DQA*
Acronym	Total data quality management	Data quality assessment
Main component	Cycle: define, measure, analyse, improve and focus: information product	Subjective and objective data quality assessment, comparative analysis, root cause analysis, actions for improvement
Type of data	Structured, semi-structured	Structured
Assessment and improvement	Comprehensive also from an implementation perspective	Makes a distinction between subjective and objective quality metrics

From the comparison of Table 1, the method used in this study is TDQM. Because TDQM is a structured and comprehensive approach (Mosley et al., 2010) in improving data quality.

3 Research hypothesis

The hypothesis is carried out to determine the target of the expected level of maturity of data quality management. The hypothesis is carried out by examining the regulations relating to the administration of education; here are some rules that are examined:

1. Law of the Republic of Indonesia No. 12 of 2012 concerning higher education
2. Ministry of Education and Culture Republic of Indonesia No. 62 of 2016 concerning the higher education quality assurance system

3 Ministry of Research, Technology and Higher Education Regulation of the Republic of Indonesia No. 44 of 2015 concerning national higher education standards

4 Ministry of Research, Technology, and Higher Education Regulation of the Republic of Indonesia No. 61 of 2016 concerning the higher education database

5 BAN-PT Regulation No. 59 of 2018 concerning guidelines for preparation of self-evaluation reports, guidelines for preparation of higher education performance reports, and assessment matrices in higher education accreditation instruments

6 matrix of accreditation forms and self-evaluation of higher education institution accreditation

7 matrix of study program accreditation instrument evaluation program accreditation

8 statute of the university

9 university strategic plan

10 university internal quality assurance system.

From these rules, there are point rules that can be adjusted to the assessment characteristics of the Losin's data quality maturity model. Another reason for adopting the Losin's data quality maturity model is because an instrument that contains a checklist for measuring the maturity of data quality management is provided. This instrument is beneficial for researchers in assessing the maturity of the quality of data management. Also, Losin's data quality maturity model has been successfully used to measure the maturity of data management quality at government authorities in Indonesia.

These rules are used as the basis for determining the maturity level of data quality management, the following results of the review of these rules, for expected data quality the target is 3, for data quality dimension the target is 2, for information policy the target is 4, for data quality procedure the target is 3, for data management the target is 3, for data standardisation the target is 2, for technology the target is 3, and the last for performance management the target is 4. Next, to measure the expected maturity of data quality management, data collection instruments are needed. Data collection is done by using the level of maturity of data quality management, which is compiled based on Losin's data quality management maturity model with the following questions.

3.1 Expected data quality

1 Data quality assurance activities at this faculty are active.

2 Can identify data quality expectations.

3 Have documentation of data quality expectations.

4 The ability of anticipation against data problems is still limited.

5 There is an overview of the data quality expectations, including certain data quality dimensions and certain data values.

6 Simple data errors can be identified and reported.

7 Dimensions of data quality have been identified and documented.

8 Data quality expectations set out in data quality rules (including data quality dimensions related to data values, formats, and descriptions).

9 Data validation is done by using established data quality rules.

10 Methods for assessing the impact of data quality on businesses have been explored.

11 Data validity has been checked and monitored.

12 An analysis of the impact of data problems on the delivery of education is common.

13 Prioritisation of fulfilling data quality expectations as a follow-up to the analysis of the impact of data problems on the business has been formulated.

14 Assessment of data quality is carried out routinely and on a scheduled basis.

15 Guidelines for data quality have been established.

16 Compliance with data quality rules associated with employee performance targets.

17 The target of improving the quality of data has been set in accordance with the capabilities of this faculty.

18 Control on data validation has been integrated with the business processes of conducting education.

3.2 Data quality dimension

1 The faculty can measure data quality.

2 Data quality issues have been addressed.

3 The characteristics of data quality problems have been identified.

4 Common dimensions in data quality measurement have been used in the faculty.

5 The faculty can enforce measurements of data quality using data quality rules.

6 Data quality expectations set out in the data quality rules (including data quality dimensions and rules related to data values, formats, and descriptions).

7 The faculty can validate model values, and exchange data using predefined data quality rules.

8 The faculty has a simple report related to the measurement of data quality.

9 The quality dimension of the wench has been harmonised with the business impact.

10 The faculty has a data quality report matrix.

11 A data steward is notified whenever there is a data problem.

12 SLA for the quality of virgin has been established.

13 SLA for data quality is always monitored.

14 Data quality has become again from the sustainable development (SDLC) life cycle.

3.3 Information policy

1 Information policies are informal.
2 Information policy has been documented.
3 Data corrective actions were taken by many staff and coordinated.
4 The organisation tries to consolidate data in one source.
5 Privacy policies and restrictions on data use already exist but are still rigid.
6 Basic policy regarding the mechanism for handling data problems has been established.
7 Guidelines for achieving data quality management objectives are available.
8 Best practices have been adopted from data quality practitioners.
9 Data quality SLA has been designed to see compliance with information policies.
10 Information policies have been established and are coordinated throughout the organisation.
11 The history of data exchange is recorded in detail.
12 Data quality management is based on an information policy.
13 One of employee performance achievements is driven by information policy.
14 Data quality SLA have been established and are used to see compliance with information policies.
15 There is an automatic notification when there is a non-compliance with the information policy.
16 Employees obey by themselves have been set (self-governing).

3.4 Data quality procedure

1 Handling data problems is done in a systematic way.
2 There is coordination between work units in data improvement measures.
3 The root causes of data problems can be identified.
4 The same data problem occurs repeatedly.
5 The faculty can trace incomplete data issues.
6 The faculty can trace syntax problems or invalid data structures.
7 Analysis of the root causes of data is possible using data quality rules and simple validation rules.
8 Inspection procedures have been established and documented to check the accuracy and validity of the data.

9 Data quality management is carried out at the work unit level and at the faculty level as a whole.

10 Data validation is done automatically, and only problematic data is checked manually.

11 Available procedures for making backup data.

12 Virgin quality is actively monitored.

13 The control mechanism of the data is designed and implemented into the application or information system supporting the implementation of education in the faculty.

14 Data errors can be identified as early as possible.

15 The process of improving data follows the definition of good governance.

16 There is validation for data exchange.

17 The data validation process can be audited.

18 Control over data quality applies to the entire organisation.

19 Data quality measures published to each data user.

20 Managing data quality transparently.

3.5 Data management

1 At the faculty, there has been a discussion related to data quality management.

2 Data problems are the responsibility of all work units, not just information technology units.

3 Stewardship data available.

4 Responsibility for improving data is assigned on an ad hoc basis.

5 Best practices are collected and shared with all data users.

6 There is a working group to design and make program recommendations and governance policies.

7 The working principles of data quality are already in the development stage.

8 The organisational structure for monitoring data governance has already been defined.

9 Work principles and policies related to data governance have been documented.

10 Stewardship program has been run, and all organisations have the same view of the program.

11 Data management operational procedures have been defined.

12 There is a data management committee consisting of various work unit representatives at the faculty.

13 The data governance committee meets regularly.

14 Have an SLA for operational quality data management.

15 Each work unit uses the same governance framework.

16 There are measurable indicators as a form of control over the application of data governance.

17 Data quality performance matrices are always evaluated for improvement.

18 Awards are given to staff who successfully meet data quality targets.

3.6 Data standardisation

1 There is a standardisation of data.

2 There is uniformity in appearance for similar data.

3 There is a data definition.

4 The definition of data uses the terminology commonly used in education.

5 Reference data have been identified.

6 The data elements have been identified from the organisation's information needs.

7 The data source certification process begins to be identified.

8 Metadata standards are managed in all work units.

9 Guidelines for standard exchange formats have been identified.

10 There are data standards and metadata management.

11 Data format and structure standards have been set for all data elements.

12 The data exchange scheme has been set.

13 The data source certification process starts.

14 A master data reference has been set.

15 Data exchange standards are managed through a process of monitoring data standards.

16 The data standards supervisory board oversees data standards and their achievements.

17 Master data management is completely implemented to manage master data.

18 Taxonomies for data standards have been defined and implemented.

19 There are technical rules that are in accordance with established data standards.

20 The data standardisation process is done automatically.

3.7 Technology

1 The routine work of the information technology unit is carried out in a planned (not ad hoc) manner.

2 The information technology unit does not avoid complaints about data issues.

3 Available tools to check data quality objectively.

4 Available data standardisation, pursing data and repairing data.

5 Available technology to find, match and connect data.

6 Standard procedures are available for using data checking and quality improvement applications.

7 Validation based on business rules has been carried out.

8 Technology components for implementing data validation, checking and reporting on data quality are available.

9 The technology component is standard for all lines of organisation in terms of service and implementation.

10 Correction of data is carried out automatically based on guidance from the policies and rules of the administration of education.

11 Analysis of the impact of data problems supported by reporting and dashboard applications.

12 Non-technical use can dynamically define and modify data quality rules and data dimensions.

3.8 Performance management

1 If a data problem is found, the impact of the problem can be identified.

2 Areas of administration of education affected by data problems can be identified.

3 Data profiles are used to identify data errors in the education implementation process.

4 A framework for analysing the impact of data problems is already available.

5 Data quality service components are available and can detect data errors early.

6 Data quality service components have been identified.

7 A problem tracking system is available and records every problem and the solution.

8 Data quality matrices are displayed in managerial reports.

9 An audit has been carried out based on compliance with data quality rules.

10 There is regular reporting on data quality management.

11 Available dashboard for monitoring the performance of data quality management.

12 There is a distribution of access rights to data and information according to the employee's role.

13 The contribution of data quality components to the delivery of education can be well described.

14 Organisational performance in overall data quality management can be improved through the modification of policies and related rules.

4 Research methodology

This research is in the form of a case study conducted at the faculty from a public university in Indonesia, using quantitative methods in measuring the quality of data and continued using qualitative methods to analyse the measurement results. The assessment is done using structured query language (SQL) queries directly on the human resource information system (HRIS). A system built by the Faculty of Psychology to manage all faculty data, using a data sample from 2015 to 2019.

As explained earlier, this study covers the first three stages of the TDQM methodology. The steps of this research adopt the previous research steps as follows:

- Analyse and understand the scheme from the HRIS database.
- Identification of essential data to be measured based on government regulations regarding documents (Kemenristekdikti, 2016).
- Extract and explore the data needed from the HRIS database.
- Determine dimensions of data quality that refer to dimensions set by the government, including completeness, correctness, accuracy, and novelty. From each dimension, criteria are defined based on business rules for abnormal data viewing on the system.
- Measurement of data quality using direct queries.
- Analysing the results of measurement of data quality and analysis of the causes of an anomaly in data quality found in the system. Interviewing IT staff who oversee the system done to analysis of the causes of problems with data quality.

5 Results and discussion

Currently, every decision taken is not based on data. The stakeholders in the organisation are not sure of the data quality they owned. This study aims to determine the maturity level of data quality management and recommend data quality management. Therefore, initial information is needed relating to the maturity level of data quality. In this case, the researchers made a questionnaire form to determine the maturity level of data quality management. The researchers use Loshin's (2011) data quality management maturity model by adopting the answer choices in the questionnaire assessment process software – the capability maturity model. The questionnaire was distributed to five respondents, namely: head of IT, head of archives, head of public relations, coordinating centre for educational administration, and general manager.

In the questionnaire for measurement expected data quality, there are 18 questions, for measurement data quality dimension there are 14 questions, for measurement information policy there are 17 questions, for measurement data quality procedure there are 20 questions, for measurement data management there are 18 questions, for measurement data standardisation there are 20 questions, for measurement technology there are 12 questions and for measurement performance management there are 14 questions. The following is the level of data quality maturity based on the form of the contents.

Table 2 Results of measurement of data quality management maturity (see online version for colours)

Measurement	*Current*	*Target*
Expected data quality	2.4	3
Data quality dimension	1.6	2
Information policy	2.8	4
Data quality procedure	2.5	3
Data management	2.3	3
Data standardisation	1.7	2
Technology	3	3
Performance management	3.6	4

Note:The coloured cells is show component achieved the target.

The values in Table 2 come from questionnaires filled out by five respondents; there are four answers, namely, ‘yes, no, not applicable, do not know’, if the respondent chooses the ‘yes’ answer, the respondent will continue to fill in the level of maturity on a scale of 1–5 for each question characteristic. Each component has a different number of questions. Table 2 shows that out of eight components, only one component, namely technology that has achieved the target, while seven others are below the target, makes the gap between the expectation with the current level of maturity.

The researcher’s analysis of the results shown in the components below the target occurs because stakeholders in the organisation do not yet know the importance of data quality in supporting organisational success, so there is a need to increase stakeholder awareness of data quality. There is no information policy on data quality needs that are adjusted to the organisation’s internal and external needs; data quality assessment has not been done using established data quality rules and their impact on education implementation; data quality business rules have not been defined related to data standardisation.

The dimensions used in measuring data quality are the dimensions of completeness, validity, accuracy, and novelty. Researchers conducted interviews to measure data quality with five related faculty staffs from general, education, research, collaboration, and technology departments. These interviews to determine what can be measured with the dimensions that have been determined. From the result of the interviews, it was found that these dimensions can be made as the context or categories to measure the data quality available in the HRIS.

The system consists of a total of 84 tables, but only four tables that were used as the subject of measurements because they represent the core data of the system. In the first table consists of 69 attributes with a record number of 458, the second table consists of

22 attributes with a total of 5,242 records, the third table consists of 13 attributes with a total of 36,001 records, and the fourth table consists of 54 attributes with the number record of 8,251. The measurement was carried out using direct queries on data stored on MySQL version 10.1.39. This measurement was done by seeing whether the data in the system meets the business rules of each dimension specified or not. From the measurement results in this study, there are criteria for business rules that can be met, and there are several criteria that cannot be met.

The results can be seen in Table 2, which can be seen in full, the suitability of completeness dimensions is 36.5%, the suitability of validity dimensions is 69.5%, the suitability of accuracy dimensions is 14.28%. While in the latest dimension, we cannot measure it. The results of measuring the quality of each dimension and analysing data quality problems will be discussed in more detail in the next discussion.

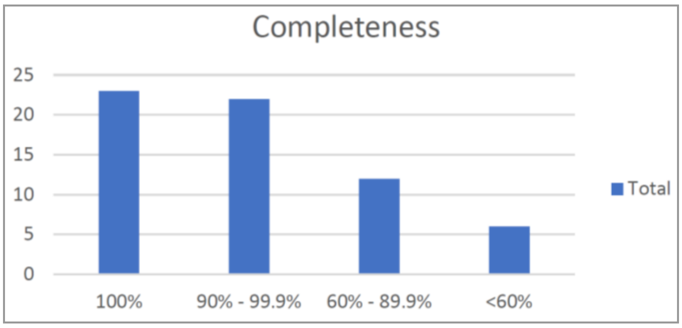
Table 3 Results from data quality measurements

Dimension	*Business rule*		*Match*		*Not match*	
	Total	*%*	*Total*	*%*	*Total*	*%*
Completeness	63	100	23	36.5	40	63.5
Correctness	23	100	16	69.5	7	30.5
Accuracy	7	100	1	14.28	6	85.72
Novelty	-	-	-	-	-	-

5.1 Dimension completeness measurement

The completeness dimension is measured by seeing whether an attribute in a table has a value or not. If an attribute contains null or an attribute contains empty spaces (“ ”), then the attribute is grouped into attributes that do not meet the completeness criteria. From Table 2, we can see that from the 63 attributes that assessed, only 23 attributes met the criteria for completeness dimensions, or equivalent to 36.5%, which were fulfilled by HRIS. Data that does not meet the criteria comes from employment data, including ID card number, class, starting date of work, e-mail, cellphone number, address, mother's name, department, attendance, and attributes related to the date.

Figure 1 Results measurement of dimension completeness (see online version for colours)



5.2 Measurement of validity dimensions

For the validity dimension, we can see from Table 2. We can see there are 16 out of 23 suitable criteria or equivalent to 69.5%. The problems encountered in this dimension are cellphone number, identity card number, account number, the start date of work, retirement date, active retention, and inactive, which has no value or value but does not conform to the rules.

The following list is the business rules for measuring the validity dimension:

- employee ID number cannot contain strange letters or characters
- the identity card number cannot contain strange letters or characters
- cellphone numbers cannot contain strange letters or characters
- account numbers cannot contain strange letters or characters
- names cannot contain numbers or strange characters
- mother's name cannot contain numbers or strange characters
- names in accounts cannot contain numbers or strange characters
- place of birth cannot contain numbers or strange characters
- start date of work must be after the date of birth
- retirement date after start date of work
- the coefficient category cannot contain strange letters or characters
- presence cannot contain strange letters or characters
- the unit price of the university scheme cannot contain strange letters or characters
- rewards cannot contain strange letters or characters
- the unit price of the faculty scheme cannot contain strange letters or characters
- active retention cannot contain strange letters or characters
- inactive retention cannot contain strange letters or characters.

5.3 Measurement of accuracy dimensions

To find out that the quality of HRIS data has met the accuracy dimension by comparing the existing data in the HRIS system with the applicable rules. In this dimension, there are seven criteria, and HRIS can only meet one criterion or equivalent to 14.28%, which can be fulfilled is the age of workers when starting work that is above 17 years. In this dimension, the remaining six are not according to the rules. Following are the business rules for measuring the accuracy dimension:

- date of birth format is year-month-date (yyyy-mm-dd) and not 0000-00-00
- start date of work format is year-month-date (yyyy-mm-dd) and not 0000-00-00
- the retirement date format is year-month-date (yyyy-mm-dd) and not 0000-00-00

- the difference between the start date of work and date of birth is at least 17 years
- the unit price of the university scheme rewards is the result of the equivalent X index of the university scheme credits X unit price of the university scheme
- the unit price of the faculty scheme rewards is the result of (credits X index of the faculty scheme X unit price of the faculty scheme)/should be present
- incentive attendance for educational staff is a minimum of eight hours, and the lecturer is a minimum of six hours.

Figure 2 Results measurement of validity dimensions (see online version for colours)

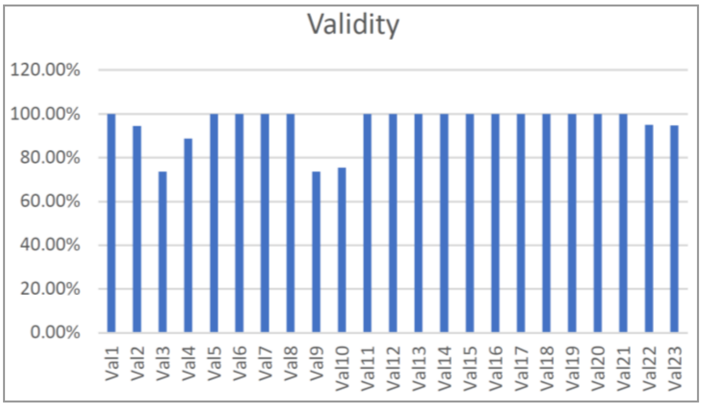
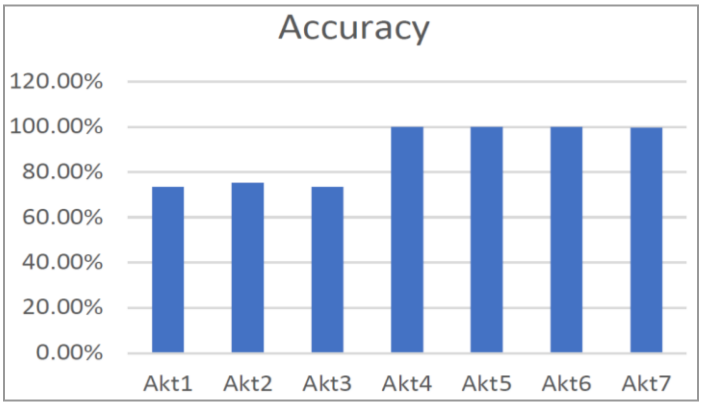


Figure 3 Results measurement of accuracy dimensions (see online version for colours)



5.4 Measurement of novelty dimensions

This dimension cannot be measured, because the tables in HRIS do not contain information about the last time that the data was updated.

5.5 Discussion

Based on the results of data quality measurements conducted in this case study with using the TDQM framework and referring to the dimensions of government regulations, the following results are obtained: for completeness, dimensions met 23 of 63 attributes that meet the criteria for completeness dimensions or equivalent to 36.5%, for validity dimensions there are 16 out of 23 attributes that meet the validity dimension criteria or equivalent to 69.5%, for accuracy dimensions, there are 1 of 7 attributes that meet the dimension criteria or equivalent to 14.28%, the latest dimension cannot be done, because there is no information about the last time to make changes data.

The problem arises because, at the time of data measurement, there were blank values (zero values), invalid data entry, data duplication, and unavailability of the latest information to update the data. The researcher also analyses the cause of the problem from the organisational policies and the HRIS system itself.

The findings of this study are recommendations for improvement in data quality management as an effort to improve the quality of educational data. These recommendations can then be used by making policies related to data quality management, and ultimately impacting stakeholders in decision making based on data.

Similar studies have been conducted at PDDikti data quality, data quality assessment at the institute of statistics, and data quality in BPS Bengkulu province and the results that are not much different. These finding can be generalised because the data quality is not monitored regularly due to a lack of data quality management in the organisation. So, the level of maturity of the case study sites is not in accordance with the expected standards. With this research, organisations can find out the condition of the data quality at this time and can determine the steps for future improvement so that the data quality becomes in accordance with applicable rules and can provide more benefits.

Research's academic contributions based on comparison with previous studies can be a reference for similar research in determining the expected level of maturity, by comparing the rules that apply in organisations with the characteristics of Loshin's data quality maturity model. Conduct questionnaires to measurement and conduct interviews to determine what can be measured with the dimensions that have been determined.

The innovation about data quality management is reflected through a series of continuous processes that start from determining parameters to determine the level of data quality, analysing data quality, identifying data anomalies, and ensuring that the information produced meets the needs of all data consumers in the organisation.

6 Conclusions

The results of assessing the quality of the data at this Faculty by using the data stored in HRIS show that in each dimension, the measurement is incompatible, the data is incomplete, the data is incorrect, inaccurate, and cannot be ascertained with the latest data. This incident occurred because this faculty has not implemented data quality management.

Further research can be developed from this finding by taking multiple case study samples not only from one faculty but several faculties to get the generalised results across faculties in the same university.

Acknowledgements

This study was supported by PIT9 Research Grant No. NKB-0005/UN2.R3.1/HKP.05.00/2019 Universitas Indonesia.

References

Alpen, A., Bank, V.G. and Stefan, R. (2019) *An Overview of Data Quality Frameworks*, Universität Klagenfurt.

Batini, C. (2015) *From Data Quality to Big Quality*, Italian National Institute of Statistics.

Batini, C. and Scannapieca, M. (2006) *Data Quality Concept, Methodologies, and Techniques*, Springer, Heidelberg.

Batini, C. et al. (2009) 'Methodologies for data quality assessment and improvement', *ACM Computing Survey*, Vol. 41, No. 3, pp.1–52.

Bertoni, M. et al. (2009) 'A case study on the analysis of the data quality of a large medical database', *20th Int. Work. Database Expert Syst.*, pp.308–312.

Chichi, C. and Rass, S. (2019) 'An overview data quality framework', *IEEE Access*, Vol. 7, pp.24634–24648.

D. International (2017) *The DAMA Guide to the Data Management Body of Knowledge*, 2nd ed., pp.591–600, Technics Publications, LLC, New Jersey.

Herawati, P. (2016) *Strategi Peningkatan Kualitas Data:Studi Kasus Pada Pangkalan Data Pendidikan Tinggi Kementerian Riset, Teknologi, dan Pendidikan Tinggi*, Universitas Indonesia.

Kemendikbud (2014) *Ermendikbud No 50 Tahun 2014 Tentang Sistem Penjaminan Mutu Pendidikan Tinggi*, pp.1–7.

Kemenristekdikti (2015) *Surat Edaran Dirjen Dikti No. 226/E4.5/2015 Tentang Pangkalan Data Pendidikan Tinggi*, Kementerian Riset, Teknologi dan Pendidikan Tinggi, Jakarta.

Kemenristekdikti (2016) *Permenristekdikti No.61 Tahun 2016 Tentang Pangkalan Data Pendidikan Tinggi*, Kementerian Riset, Teknologi dan Pendidikan Tinggi, Jakarta.

Loshin, D. (2011) *The Practitioner's Guide to Data Quality Improvement*, Elsevier, Burlington.

Mosley, M., Brackett, M.H., Earley, S. and Henderson, D. (2010) *The DAMA Guide to the Data Management Body of Knowledge*, pp.1871–1874, Technics Publications, LLC, Bradley Beach, New Jersey.

Olson, J.E. (2003) *Data Quality: the Accuracy Dimension*, pp.170–192, Morgan Kaufmann Publishers, San Francisco.

Yudhoyono, S.B. (2012) *Undang-Undang Republik Indonesia Nomor 12 Tahun 2012 tentang Pendidikan Tinggi*, pp.1–97.

Sabtiana, R., Yudhoatmojo, S.B. and Hidayanto, A.N. (2017) *Data Quality Management Maturity Model: a Case Study in BPS-Statistics of Kaur Regency, Bengkulu Province*, Universitas Indonesia.

Sidi, F., Panahy, P.H.S., Affendey, L.S. and Jabar, M.A. (2013) *Data Quality: a Survey of Data Quality Dimensions*, University Putra Malaysia.

Wang, R.Y. (1998) 'A product perspective on total data quality management', *Communications of the ACM*, Vol. 41, No. 2, pp.287–308.

Wang, R.Y., Ziad, M. and Lee, Y.W. (2012) *Data Quality*, Kluwer Academic Publisher, New York.

Wijayanti, W., Hidayanto, A.N., Wilantika, N., Adawati, I. and Yudhoatmojo, S.B. (2018) *Data Quality Assessment on Higher Education: a Case Study of Institute of Statistics*, Universitas Indonesia.

Woodall, P., Borex, A. and Parlikad, A.K. (2013) 'Data quality assessment: the hybrid approach', *Information & Management*, Vol. 50, No. 2013, pp.369–382.