



Assessing the Effectiveness of System for Managing Integrated Data of Higher Public Institutions

Rahayu Ahmad¹ , Suzilah Ismail¹ , Fauziah Baharom¹ ,
Nur Haryani Zakaria¹ , Haslina Mohd¹ , Rohaida Romli¹ , Yuhanis Yusof¹ ,
Mohamed Ali Saip¹ , Juhaida Abu Bakar¹ , Wan Hussain Wan Ishak¹ ,
Mohd Hasbullah Omar¹ , and Salwati Badroddin²

¹ Universiti Utara Malaysia, 06010 Sintok, Kedah, Malaysia
rahayu@uum.edu.my

² Strategic Planning Division, Ministry of Higher Education, 62200 Putrajaya, Malaysia

Abstract. Effective data management in higher education institutions is essential for maintaining national higher education quality. Centralizing and integrating this data within a unified database enhances accuracy in estimates, improves decision-making, and aids in developing better strategies. To support this, a system was developed to collect and integrate data from public higher education institutions, and it has undergone continuous refinement to ensure reliability and efficiency. This study evaluates the effectiveness of the integrated system and assesses user satisfaction among public and private higher education institutions. Two data collection methods were employed for this study: surveys and interviews. The primary method involved surveys with 77 respondents, while the supplementary approach included interviews with eight officers involved in data management and strategic planning across various public higher education institutions. Overall, participants expressed satisfaction with the integrated system, noting that it complements internal institutional data and is a valuable resource for strategic planning. However, several areas for improvement were identified, including the need for better integration with other systems, the development of a data dictionary, and automated data mapping to address issues in the data transmission process.

Keywords: Data Integration · Portal · Strategic Planning · Data Quality

1 Introduction

Managing data in higher education institutions is crucial to ensuring national higher education quality is consistently upheld, particularly in fostering competitive human capital. There is a significant amount of academic data in higher education institutions. This includes data about students, staff, graduates, research and publications, and institutional information, among other things. Higher education data is heterogeneous since it comes from various institutions and is kept in distinct locations on a range of platforms. To enable a comprehensive and methodical examination, it is essential that this

data be maintained and integrated into a centralized database. Integrated data analysis can enhance accuracy in estimates, more precise decision-making, and formulation of better strategies [1]. Among the challenges during data integration are missing, erroneous, or incomplete data, even with the abundance of support tools available to aid with data integration [2]. To ensure data quality, a developed system has been developed which collects data from public higher education institutions. Since its creation in 2006, the system has continuously improved to guarantee reliable and efficient functioning. Therefore, it is imperative to evaluate the effectiveness of the system and how the system is used in strategic planning of the Public Higher Education Institutions.

2 Literature Review

2.1 Heterogeneous Data Management and Integration

Nowadays, most organizations focus on data because it is one of their valuable assets. This perception of data as an asset requires organizations to commit to improving data quality. If the data managed is of good quality, more benefits will be obtained [3]. Generally, data can be processed to produce value that can be used by stakeholders. Therefore, data needs to be managed well and effectively so that the value obtained from the data can benefit the organization.

Poor data quality can affect the productivity of an organization. However, in general, the management of large-sized data is very complicated and complex. Large data processing can be processed in batch processing and stream processing. Batch processing is performed based on data analysis in a certain period of time and is adopted when there are no constraints involving response time. On the other hand, streaming processing is suitable for real-time feedback applications [2]. In addition, one of the challenges faced in managing large-sized data is that if the data involved is heterogeneous, the data is stored in various platforms separately. The data needs to be integrated to enable a holistic analysis.

However, integrating heterogeneous data is challenging and requires an iterative process [4]. Among the challenges is when performing data extraction, which involves the process of pulling data from data sources. This process is the first step that needs to be taken when extracting data. It is a complicated and time-consuming process because the data source uses different formats, structures, and types. Therefore, after the data is extracted and before it is integrated into the destination system, the data needs to be transformed to make it consistent and match [5].

In addition, data integration also involves managing unstructured data. Therefore, the data lake approach (data pool) must be implemented for this purpose. According to [4], a data lake (data pool) is a central storage repository supporting large quantities and data types, which involves structured and unstructured data. Nevertheless, metadata development is necessary to avoid ambiguity to ensure the data lake implementation is done effectively. Without any metadata, dumping all the data into a data lake will result in the data being unable to be used to its best potential. In addition, according to [4], when collecting data from heterogeneous systems, there are three sources of data errors: data entry errors, data type incompatibility, and semantic incompatibility in business entity definitions.

Furthermore, according to [6], data quality is vital in continuously managing and maintaining data. If the data is not managed correctly, some data errors, such as missing values, spelling errors, inaccurate data, incorrect formats, inconsistencies, etc., may occur. These small mistakes can cause data to become unusable and create poor data quality issues, leading to more severe problems. It is well known that poor-quality data can cause problems such as increased cost and time required to repair the data.

In addition, good data management regarding data security must also be emphasized, especially when sharing data with third parties [7]. This is followed by the need for comprehensive data governance, data collection, and data usage that continues to grow. Therefore, excellent and efficient data management by using certain technologies or methods, such as through a data management system that takes data security into account and using data analytics methods, can help the management of an organization to ensure data security [7], obtain helpful information, helping them manage data management better, make transparent, and accurate decisions, also perform data management.

Apart from data security issues, data governance needs to emphasize critical success factors (Critical Success Factors (CSF)) that can determine the direction of data governance in an organization. [8] through their research related to the financial data governance model, has identified seven main core categories that are CSFs for data governance: 1) employee data efficiency, 2) transparent data processes and procedures, 3) data tools and technologies that are flexible, 4) standardized data policy, 5) developed/acquired data roles and responsibilities, 6) precise inclusive data needs, 7) focused and visible data strategy. They also found that creating successful data governance requires high attention from stakeholders and the need to establish data roles and responsibilities. They were followed by focusing on actions related to the underlying data-focused strategy to create a successful data governance program.

An Integrated System for Managing Public Higher Institutions Data

In an attempt to address the issues regarding data management and integration, an integrated system was developed. The integrated system for integrating and improving data quality from Public Higher Education Institutions (PHEI) was established in 2006. It is an application system that runs on the web and consists of multiple parts: The e-Lounge System, the Web Analytics System, and the Information Portal.

The e-Lounge System component was created to integrate and improve data quality over time. This integrated system also helps prevent potential disagreements over the legitimacy and quality of the data given. This system facilitates data quality validation after data collection to guarantee no errors before the data is stored.

Another component of the Web Analytics System enables the preparation of reports for specific modules in higher education institutions. Four (4) modules are included: the Student Module, Staff Module, Research, Development and Commercialization (R&D&C) Module, and Institutional Information Module. Since its development, this system has undergone several revisions. Hence, it is imperative to evaluate the efficacy of the integrated system and user satisfaction among public and private higher education institutions.

3 Methodology

For this study, two approaches to data collection were employed. The primary data collection uses surveys, while the complementary approach executes interviews.

3.1 Survey

The survey was distributed to all officers in PHEI who were experienced in using the integrated system. The target respondents (population) consist of 161 coordinators of all PHEI. Using a purposive sampling approach, the survey link was distributed to a social media group of all coordinators and officers in PHEI. To increase the response rate, reminders were sent occasionally to the group, and 77 respondents participated in the survey.

The survey is composed of several parts: Part A: Demographics. This section contains four questions: the institution, the respondent's role in sending data to the integrated system, the modules involved by the respondent, and the duration of experience using the system. Part B: Infrastructure. This section also contains questions about respondents' infrastructure problems when using the integrated system. These include Internet access problems, hardware/device problems, application/software problems, and technology/computer skills issues. Part C: Usability of the Integrated System. This section has several questions adapted from [9] that measure system quality, service quality, information quality, and satisfaction.

3.2 Interviews

An interview was conducted with eight officers involved in managing data and strategic planning in various Public Higher Education Institutions to determine the system's effectiveness. Each interview session lasted from 20–30 min. The Interview guide questions are as follows:

1. How do higher education institutions perceive the usefulness and function of the integrated system?
2. How do higher education institutions utilize data from the system in strategic planning or policy formulation and framework of action plans?
3. What is the level of support from senior management towards the implementation and management of the system?
4. Are there any suggestions for improvement?

3.3 Data Analysis for Survey

A descriptive SPSS software study was conducted to determine user satisfaction with the integrated systems. The analytical report is included in the upcoming section.

Based on Fig. 1, a large portion of respondents (65%) were technical officers and coordinators of the integrated system at each PHEI. Specifically, the participants consisted of 30% coordinators, while 35.56% were technical officers. Around 28.8% of respondents are data owners. Hence, the respondents were relevant to this study since they manage data transmissions and use the integrated system.

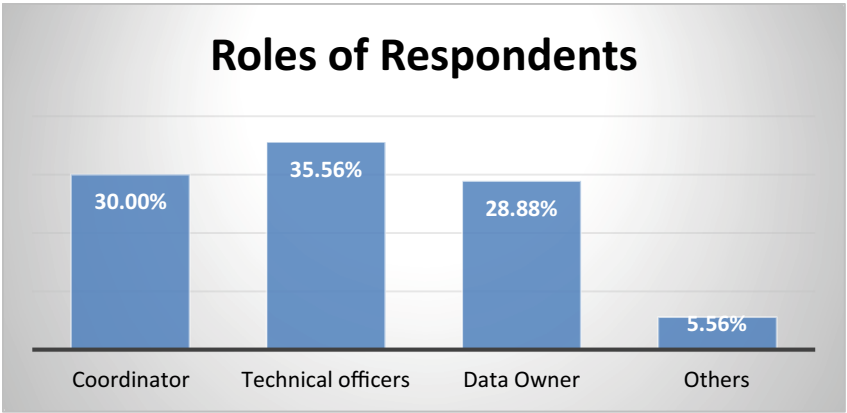


Fig. 1. Distribution of respondents based on their roles.

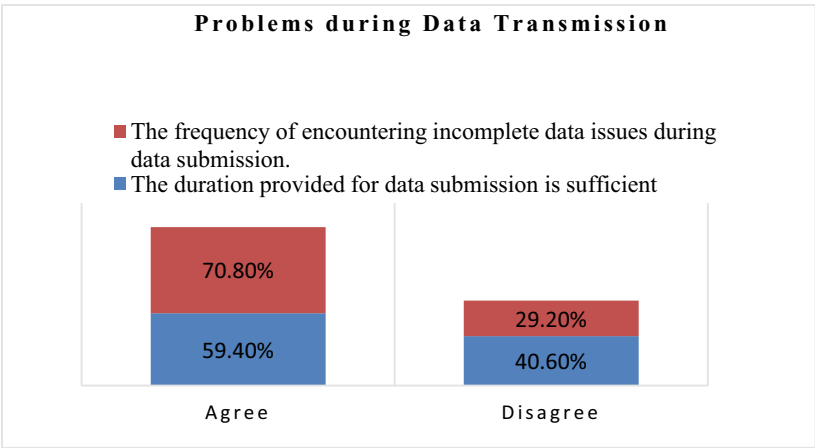


Fig. 2. Problems during Data Transmission.

The study results indicate that 70.8% of respondents encountered issues with incomplete data during data submission. This problem is partly related to the insufficient data submission period (40.6%) (Fig. 2).

Perception of System Quality. Most respondents strongly agreed that the integrated system is easy to use and has all the required features, as reflected by the mean score in Fig. 3.

Perception of Service Quality. Similarly, most respondents agreed that the integrated system provided support through manual, workshop, documentation and online help (Fig. 4).

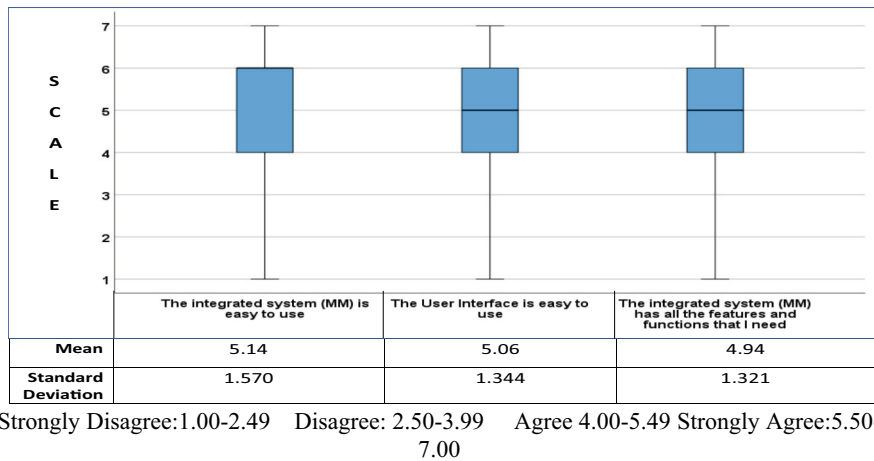


Fig. 3. Perception of System Quality.

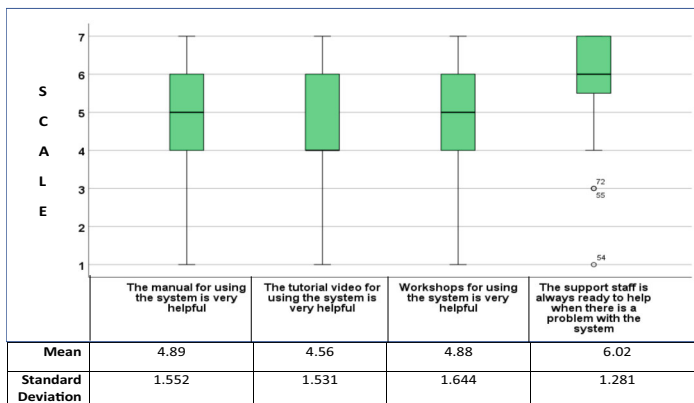
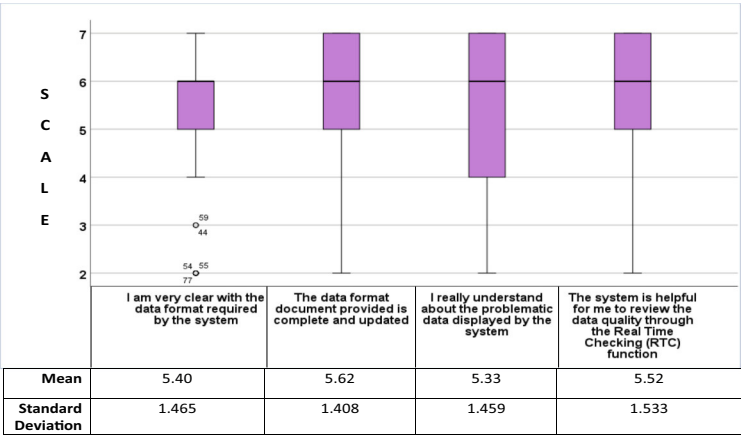


Fig. 4. Perception of Service Quality.

Perception of Information Quality. Many respondents had positive perceptions about the completeness and accuracy of the data provided by the integrated system, as demonstrated in the figure below. Most of them strongly agree that they understand the data format required by the system and that the data format provided is complete and updated. They also strongly agree that they know the problematic data displayed by the system (Fig. 5).

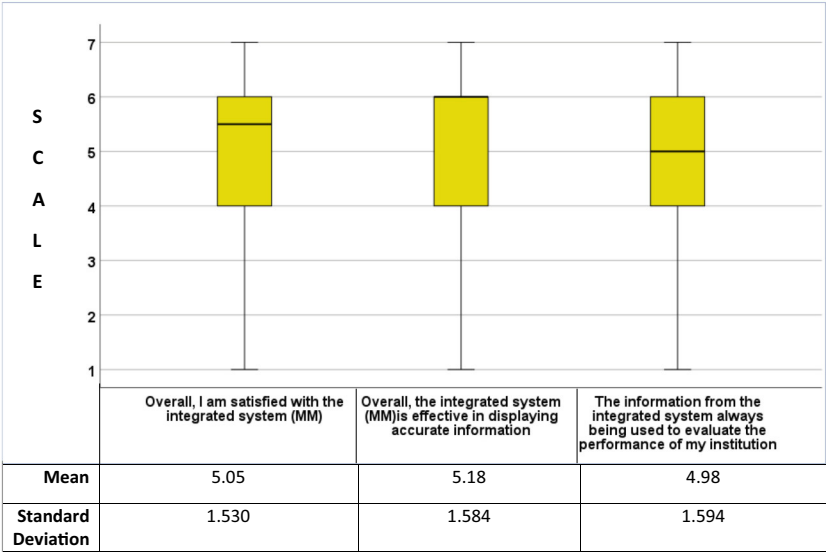
Satisfaction with the Integrated System. In conclusion, most respondents were satisfied with using the integrated system and perceived it to be easy to use. The system



Strongly Disagree:1.00-2.49 Disagree: 2.50-3.99 Agree:4.00-5.49 Strongly Agree:5.50-7

Fig. 5. Perception of Information Quality.

provided the required documentation and help using it, and it provided complete and accurate information (Fig. 6).



Strongly Disagree:1.00-2.49 Disagree: 2.50-3.99 Agree:4.00-5.49 Strongly Agree:5.50-7

Fig. 6. Satisfaction towards the Integrated System.

Data Analysis for Interview. Data analysis was conducted with two coders. Open coding analysis was undertaken to identify emergent themes on the a) usefulness

of the integrated system, b) the utilization of the system for strategic planning, c) recommendations

4 Results and Discussion

4.1 Facilitating Analysis and Trends

The integrated system's usefulness is perceived as a platform that integrates data sources for all Public Higher Education Institutions (PHEI). The data allow a more profound study to spot trends and deal with graduate and student-related problems. For example, a participant commented:

"In our university, we combined data from this system and our internal database to make analysis on graduate employability"

P3

4.2 Catalyst for Development of Local Data Warehouse

The existence of this integrated system also serves as a catalyst for the development of a local data warehouse with structures that align with the integrated system. Awareness of data quality and data management systems inside PHEI improves when all are mandated to send educational data to the integrated system. As a result, IPTA implements follow-up measures to enhance data repositories and create internal systems. As highlighted by the following participant:

"The integrated system inspires us, and we have created our data warehouse to facilitate data transmission to the integrated system"

P1

Some respondents said that IPTA's senior management constantly places a high priority on data quality. To improve the quality of data administration, internal IPTA system integration is also used:

"We created our own local data warehouse, and it has reduced a lot of data validation issues when transmitting data to the integrated system"

P4

4.3 Input for Action Plan

Every six months, there will be a national Data Quality meeting where every PHEI will be rated and receive their start rating based on several data quality metrics. As a result of this rating based on the integrated system data, some Universities took further initiatives to devise their action and strategic plan as commented below:

"After each data quality meeting and the release of star rating, we will hold a meeting led by the vice-chancellor, with participation from data owners, to discuss the performance and the quality of data provided to the integrated system"

P1

Action plans were devised based on the rating performance of data supplied to the integrated system. These plans include enhancing data governance procedures and devising methods to improve data quality. Certain PHEIs have created integration systems and implemented Real-Time Checking (RTC) features in line with the integrated system's requirements to detect erroneous data early on before it is submitted to the integrated system.

4.4 Minimal Utilization

However not every PHEI optimized the data and report provided by the integrated system. Certain PHEIs only supply data to the integrated system to fulfil the reporting requirements. These IPTAs don't use this information for internal strategy planning. Additionally, any needs related to strategic planning depend increasingly on extensive internal IPTA data sources. Moreover, insufficient knowledge among senior management of the significance and accessibility of data provided by the integrated system has led to inadequate utilization of the information. Additionally, any needs related to strategic planning depend increasingly on extensive internal IPTA data sources. Moreover, insufficient knowledge among senior management of the significance and accessibility of MyMohes and SKPG data has led to inadequate use of the information.

4.5 Recommendations for Improvement

Some of the respondents highlighted the issue of frequent changes in data format that need to be supplied to the integrated system. To address this issue, a standardized data dictionary will be provided for all the PHEIs. A few of them also commented that they lacked technical competencies since they were new users. Hence, responsible officers must attend technical workshops periodically and consistently communicate with the system's technical team. There were also suggestions for creating integration with other systems and developing a standardized data model. Additionally, users requested access to dynamic data and fields in real time. An automated operational definition mapping between PHEI and the data in the integrated system needs to be implemented to facilitate this demand. This will lead to more efficient data transmission processes.

5 Conclusion

The integrated system is generally beneficial in providing data about students, staff, research, publications, and commercialization as inputs for creating institutional action plans by most Public Higher Education Institutions (PHEI). Institutions have initiated the development of internal systems and data warehouses that align with the integrated system format to facilitate more efficient data management and transmission. It is believed that the integrated system's data complements the internal institutional data and becomes an important resource for developing strategic planning. However, several improvements in data dictionaries and automated data mapping can resolve some issues during data transmission.

References

1. Zipkin, E.F., Inouye, B.D., Beissinger, S.R.: Innovations in data integration for modelling populations. *Ecology* **100**(6), 1–3 (2019)
2. Varadharajan, C., et al.: Challenges in building an end-to-end system for acquisition, management, and integration of diverse data from sensor networks in watersheds: lessons from a mountainous community observatory in East River. Colorado. *IEEE Access* **7**, 182796–182813 (2019)
3. Mahanti, R.: Data quality: Dimensions, measurement, strategy, management, and governance. Quality Press (2019)
4. Wang, L.: Heterogeneous data and big data analytics. *Autom. Control Inf. Sci.* **3**(1), 8–15 (2017)
5. Le Sueur, H., Bruce, I.N., Geifman, N.: Masterplans Consortium: the challenges in data integration—heterogeneity and complexity in clinical trials and patient registries of Systemic Lupus Erythematosus. *BMC Med. Res. Methodol.* **20**, 1–5 (2020)
6. Azeroual, O., Schöpfel, J.: Quality issues of CRIS data: an exploratory investigation with universities from twelve countries. *Publications* **7**(1), 14 (2019)
7. Gupta, R., Gupta, I., Singh, A.K., Saxena, D., Lee, C.N.: An IoT-centric data protection method for preserving security and privacy in cloud. *IEEE Syst. J.* **17**(2), 2445–2454 (2022)
8. Karkošková, S.: Data governance model to enhance data quality in financial institutions. *Inf. Syst. Manag.* **40**(1), 90–110 (2023)
9. DeLone, W.H., McLean, E.R.: Information systems success: the quest for the dependent variable. *Inf. Syst. Res.* **3**(1), 60–95 (1992)
10. DeLone, W.H., McLean, E.R.: The DeLone and McLean model of information systems success: a ten-year update. *J. Manag. Inf. Syst.* **19**(4), 9–30 (2003)