The Construction of Data Governance System in Vocational Colleges

Shouzhong Wang, Hongbin Gu, and Kongpeng Wei(✉)

Panjin Vocational and Technical College Zhengbang Road, Panjin, Liaoning, People's Republic of China
{wangshouzhong,guhongbin,kpwei}@pjvtc.edu.cn

Abstract. The information construction of many vocational colleges has accumulated a certain amount and different forms of data, which are in the form of relational data based on traditional business systems, log files of network devices and software, and external data related to vocational colleges. These data cannot be integrated, correlated, managed and analyzed through traditional business systems. It is necessary to use big data technology to govern these different types and sources of data, and on the basis of data governance, data integration, analysis and prediction are carried out to provide digital support for scientific management and decision-making of vocational colleges. In the field of data governance, what distinguishes vocational colleges from undergraduate institutions is not only the difference of data scope, but also the general lack of awareness, methods and talents of data governance in vocational colleges. This paper focuses on the awareness cultivation of data governance, data governance methods and data governance team building, and tries to study the methods and procedures from data governance strategic planning, data analysis procedure, data governance procedure, making data governance policy and establishing data governance structure, data management and data governance team formation.

Keywords: Data governance methods · data governance team building · system construction · vocational colleges

1 Strategic Planning for Data Governance in Vocational Colleges

In the field of data governance [1], what distinguishes vocational colleges from undergraduate institutions is not only the difference in the scope of data, but also the general lack of awareness, methods and talents of data governance in vocational colleges. In this paper, we focus on the methods of data governance and try to study the methods and procedures in terms of data governance strategic planning, data governance procedure, development of data governance policy, data management and data governance team formation.

A data governance strategy for vocational colleges [2] defines and articulates a comprehensive vision for the entire higher education institution and lays the basis for the organization to adopt capabilities which are related with or dependent on data. A precisely

C. F. Peng et al. (Eds.): EIMT 2023, AHSSEH 8, pp. 386–392, 2023.
https://doi.org/10.2991/978-94-6463-192-0_50

defined and sophisticated data governance strategy enables the gain of data governance to be made actionable for the vocational colleges. It defines the steps the organization should implement to be a "data-driven higher education institution". The data governance strategy contains guidelines to achieve the vision which is data-driven, guides the organization in selecting special business objectives, and serves as the start point for data-driven program among the vocational colleges.

In addition to realizing the benefits of competitive advantage, vocational colleges need a data governance strategy because it is beyond organization's boundaries. Without a data governance strategy, vocational colleges will have to handle lots of data-related activities that are possible to be initiated by different departments or faculties. Individual faculties or departments may initiate vocational colleges own data management, business intelligence, or analytics initiatives without considering the whole long-term strategic goals.

The driver for developing a data governance strategy should be a union of the CEO/CIO (Chief Executive Officer/CIO), and for vocational colleges, typically a combination of the dean and the vice president for information technology. This suggests that data is not only an enterprise asset, but also an IT asset for the entire organization.

2 Data Governance Procedure

Data governance procedures are defined procedures that organizations follow to ensure control of vocational colleges' data throughout its lifecycle. Because "big data" is a strategic asset, a variety of organizations need to establish controls. Data governance procedures in vocational colleges ensure that important data assets are seriously managed by the higher education institution and that the data can be believed in the procedure of making decision. Typically, the procedures used in data governance include responsibility for any negative events resulting from data quality [3].

In the last few years, the interest in data governance and its concomitant procedures has increased considerably, especially because of the increased demand for data confidentiality and data privacy in countries around the world. Therefore, the data governance procedure in vocational colleges requires not only the development of policies and the assignment of responsibilities across the higher education institution, it also requires ensuring that the higher education institution complies with local data regulations and laws.

There is a tight relationship between the data management procedure and the data governance procedure. The data governance procedure is involved in setting responsibilities and policies at the strategic level, while the data management procedure is concerned with implementing and monitoring those policies at the operational level.

The data governance procedure includes technology, organizational structure, and the people that are necessary to create consistency and appropriate dealing with an organization's data. The governance procedure includes technology, organizational structure, and the people necessary to create proper and consistent processing of an organization's data throughout the higher education institution. While goals may be different depending on the essence of the vocational colleges, the level of control required, and local regulatory requirements, for every organization some general data governance activities are the same.

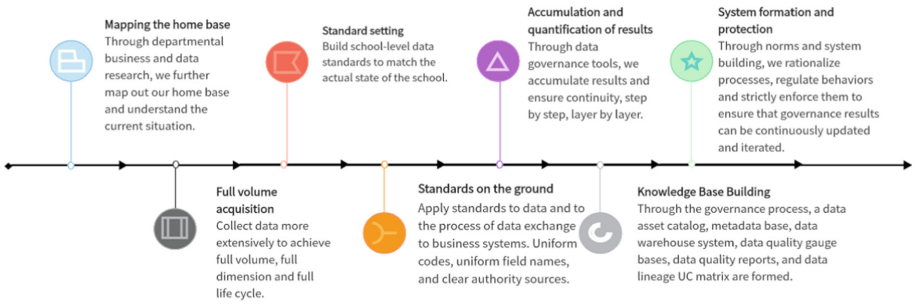

Fig. 1. Data Governance Procedure

To manage and guide all data quality activities consistent with the overall business strategy needs a data quality strategy. The strategic goals pursued by the data management procedure are included in the data quality strategy, how it aligns with overall functional scope of the higher education institution and the strategic business goals. In addition, it provides a description of its stakeholder engagement, which implies analysis and understanding of the role of data in the vocational colleges. The data quality strategy should be updated and reviewed at least every year.

There are legal requirements for citizen's data privacy [4] and in most cases, (auditable) reports and logs are required to ensure that the organization is in compliance with these laws and regulations. For vocational colleges, student and faculty data privacy should also be taken into account in the data governance procedure.

The data governance procedure can be visually illustrated in Fig. 1.

3 Develop Data Governance Policies and Establish Data Governance Architecture

Regulatory requirements and data quality policies need to be interpreted into many data governance policies and made publicly available to every person in an organization. These policy documents include the decisions of the higher education institution regarding the organization of data quality and account for the requirements for each individual. These policies are useful input for management decisions and can deliver worthy input for other business procedures.

The data governance procedure must define distinct responsibilities and roles that cross departmental frontiers in the vocational colleges. Role assignments must ensure authority, accountability and oversight, and the participatory of senior management and business management, and bolster desired behaviors in the use of data.

After developing data governance, the data governance architecture needs to be established. After analyzing the business systems and data sources of the vocational colleges, the vocational colleges' data governance architecture should be established.

4 Data Management

The data management procedure is an independent procedure and an important task after data governance. It secures the data's quality at the daily business level. This procedure enforces the strategic instructions of the data governance procedure.

The main goal of the data governance procedure is to assure data quality. The value that can be obtained by analyzing Big Data relies highly on the quality of the input data. Even with the most complex Big Data solutions, the general "garbage in garbage out" rule still works. If the data set is crashed or has errors, data analysis may lead to invalid conclusions or results.

So vocational colleges need data management procedures to continuously validate, update and cleanse higher education data. The data management procedure provides a practical and structured way to achieve the below points. Vocational colleges need a method to normalize vocational colleges' aspirations to measure the alignment of data quality with those aspirations.

Vocational colleges must be able to establish a base line of data quality levels in order to identify problems and analyze the underlying causes of data failures. Vocational colleges need to be able to transmit vocational colleges' level of confidence in the quality of vocational colleges' data.

The data management procedure is a pragmatic operational procedure (aligned with the strategic instructions of the data governance procedure) that watches data quality every day.

To assure and measure data quality through all the data lifecycle, vocational colleges assign performance and metrics indicators based on data quality dimensions appropriate to the vocational colleges' information needs. These metrics should be linked to the overall goals and objectives of the vocational colleges as defined in the Data Governance Policy.

These performance and metrics indicators can be documented in a balanced data quality score card. The creation of such a score card provides an effective way to constantly manage and monitor data based on key performance indicators.

Based on the performance and metrics indicators special in the aforementioned activities, the vocational colleges data needs to be monitored. Data sets can be indexed and monitored to measure the quality of the data based on the specified performance metrics through automation tools. The results can be described in a data quality scorecard again.

An important element of this procedure activity is the generation of alerts (and subsequent follow-up). If data is found to be corrupted or changed, alerts need to be generated.

The data management procedure' next activity is improving the vocational colleges' data set. For example, the balanced data scorecard from the previous activity may indicate that there are a number of repeated records in the dataset. The validation activity and data improvement is concerned with "cleaning up" the dataset to improve performance and metrics indicators.

The final activity of the data management procedure is to educate and communicate the vocational colleges' stakeholders to participate in the data management program actively. Through ensuring that systems are being used correctly and data management processes are followed, the quality of data at the higher education institution can be

notable improved. In a number of cases, faculty and staff do not understand the data structure and are unaware of the value of the data to the higher education institution.

Training programs can improve productivity, reduce user error, and increase adherence to key controls to improve this knowledge. Data quality practices and education addresses core data principles are finished by training of role-specified. Especially, why and how consumers use data must be understood by data collectors.

5 Data Governance Team Formation

The people whom are involved are data governance's the most important aspect. Many organizations sometimes spend not enough on the "people" side and too much time on the "data" while they have plans to turn data into value. In the short time that data science has been a part of the professional enterprise, many new roles have developed that are critical to big data's success. These roles include: big data scientists, big data analysts, big data analysts, and big data engineers.

A big data analyst [5] is a role that in order to discover business value involves processing, acquiring, and summarizing information from big data sets. Unlike a data scientist, a data analyst is much of a generalist. Big data analysts are required to know Python R, SQL, HTML, Javascript and C++. They need to have more knowledge of data visualization and storage systems, data retrieval and data warehousing using Hadoop-based analytics [7], ETL tools [6], and business intelligence concepts [8]. These passionate and persistent data miners typically have a strong background in mathematics, machine learning, statistics, and programming.

Big Data analysts are involved in data calculation [9] and data visualization. Data analysts must query databases if there is a request for data insight from a stakeholder. They are responsible for the scraped data, quality assurance and management. They must interpret the data and communicate the findings effectively.

A big data scientist [10] is a role that involves developing and deploying statistical models and algorithms to predict future outcomes and provide business value based on big data sets. The role of data scientist has gained much popularity in recent years and there is a high demand for this job role. The job role of a Big Data Scientist is an advanced role that requires a deep understanding of data processing operations and algorithms. A person in this role should be an expert in Python, R, SAS, MatLab, SQL, Hive, Pig, and Spark. Data scientists typically have advanced degrees in quantitative disciplines such as mathematics and statistics, and are proficient in big data techniques and analytics tools. The role of a data scientist is not just about the computation of data. It is about creating some valuable actionable insights about the data, understanding business challenges, and communicating vocational colleges' findings to vocational colleges. In addition, the role of a data scientist requires innovative thinking and problem-solving skills that are necessary to develop, design, and deploy algorithms that can retrieve value from big data.

The big data engineer is a role that builds, designs, and manages the underlying IT infrastructure necessary to extract value from big data sets. The data engineer ensures that the Big Data ecosystem in the higher education institution operates without fail for analysis by data analysts and data scientists. Big Data Engineers maintain, develop,

build and test highly scalable data management systems. Different from data scientists, who seek an exploratory and iterative way to reach a solution, data engineers seek a linear path. By integrating newer data management technologies, data engineers will improve existing systems. They will develop custom software components and analytical applications. Data engineers store and collect data, process it in batch or in real time, and make it available to data scientists for analysis through APIs.

6 Conclusion

Data governance in vocational colleges is a complex project that needs to be built in several aspects such as awareness, methods and team. For vocational colleges with data governance awareness, they can set up data governance teams based on existing information centers or data centers, adopt the latest big data technologies and methods, and give enough support to the teams to cope with the needs of data governance in vocational colleges in the era of big data. The question that needs to be further studied is how higher education institutions can improve their own data governance capabilities and the technical and business levels of their teams with the help of external forces, especially data governance companies.

References

1. Otto B. Data Quality Management: Framework and Approach for Data Governance [J]. Cdq Data Governance Data Architecture Datenqualittskriterien, 2007.Davis, A. R., Bush, C., Harvey, J. C. and Foley, M. F., "Fresnel lenses in rear projection displays," SID Int. Symp. Digest Tech. Papers 32(1), 934–937 (2001).
2. Minquan Y U, Liu Y, Zhao Y. Research on the Influence of Data Governance on Information Construction in Higher Vocational Colleges [J]. Journal of Beijing Institute of Economics and Management, 2018.
3. Hikmawati S, Santosa P I , Hidayah I . Improving Data Quality and Data Governance Using Master Data Management: A Review [J]. Universitas Gadjah Mada, 2021(3).
4. Warner J, Chun S A. A citizen privacy protection model for e-government mashup services [C]// International Conference on Digital Government Research. DBLP, 2008.
5. Quinn B E, Analyst S, Babineau W B , et al. The Convergence of Big Data Procedureing and Integrated Infrastructure.
6. Zhou H G, Zhou J C, Peng Y Q , et al. General Frame Design for ETL Tools [J]. Computer Applications, 2003.
7. Lee M, Jung H, Cho M . On a Hadoop-based analytics service system [J]. International journal of advances in soft computing & its applic, 2015.
8. Rouhani S, Asgari S , Mirhosseini S V . Review Study: Business Intelligence Concepts and Approaches. 2012.
9. Jelen B, Alexander M . Pivot Table Data Crunching [J]. 2005.
10. Affelt A. Acting on Big Data: A Data Scientist Role for Info Pros. 2014.