

Data governance framework for big data implementation with NPS Case Analysis in Korea

Hee Yeong Kim

June-Suh Cho

Hankuk University of Foreign Studies

Keywords

Big data, data governance, data governance framework, case analysis

Abstract

Information services based on Big Data analytics require data governance that can satisfy needs for corporate governance. While existing data governance focuses on data quality but Big Data governance needs to be established in consideration of a broad sense of Big Data services such as analysis of social trends and predictions of change. To achieve goals of Big Data services, strategies need to be established with alignment to the vision of the corporation. For successful implementation of Big Data services, there is needed a framework to enable initiation of a Big Data project as a guide and method. We propose the Big Data Governance Framework to facilitate successful implementation in this study.

Big Data governance framework presents additional criteria from existing data governance focused on data quality level. The Big Data governance framework focuses on timely, reliable, meaningful, and sufficient data services. The objective of Big Data services is what data attributes should be achieved based on Big Data analytics. In addition to the quality level of Big Data, personal information protection strategy and data disclosure/accountability strategy are needed to prevent problems.

This study conducted case analysis about the National Pension Service (NPS) of South Korea based on the Big Data Governance Framework we propose. Big Data services in the public sector are an inevitable choice to improve quality of life of people. Big Data governance and its framework are essential components for the realization of Big Data services' success. In case-analyses, we identified vulnerabilities or risk areas, and we hope that these case studies will be used as major references to implement Big Data services without problems.

Corresponding author: June-Suh Cho

Email addresses for corresponding author: jscho@hufs.ac.kr

First submission received: 27th August 2017

Revised submission received: 30th October 2017

Accepted: 27th November 2017

1. INTRODUCTION

Big Data is a major concern for the government and companies in Korea. The South Korean government has a new policy to open its data for use in the quality of life of people and to launch new commercial services of companies for economic growth. However, there are few successful Big Data cases and just introductions of data solutions of IT are popular. Data solutions are needed for Big Data services but vendors advertise exaggerated content, which may cause unreasonable expectations. This hyped expectation would be too costly for introduction of solutions, but service effects could fall far short of meeting expectations of stakeholders. Heads of IT organizations should be cautious in starting projects to introduce Big Data solutions due to overblown expectations of CEOs. From the perspective of IT experts, more attention should be devoted to how to manage Big Data solutions after the introduction.

There are issues that must be addressed with the introduction of Big Data solutions, such as security against leakage of personal information and accurate calculation of financial resources needed for heavy investment in infrastructure. Apart from IT issues, the introduction of Big Data solutions requires estimation of an organization's effectiveness against investment. If there is a significant difference between predicted values provided by Big Data solutions and actual results, decision-making about the introduction of solutions will be challenged.

From the perspective of companies, it is necessary to consider if investment in Big Data can maintain competitiveness and create new business opportunities. In contrast, from the perspective of the public sector, it is necessary to consider disclosure of data and its scope for the sake of public interests, the private organization's commercialization of disclosed data, and achievement of the public's benefits. And even if the review is conducted sufficiently for the sake of public interests, it is necessary to devise strategies on how to prepare for change management and how to plan successful projects.

In this study, we will first look at various issues and risks relating to expected strategies at the start of projects for implementation of Big Data or relating to service operation, after projects are completed. And we will describe those issues and risks through case analysis from the perspective of data governance. After that, we will produce implications in achieving a successful Big Data project and maximizing its effectiveness.

As research methods, we first look at what kinds of issues can possibly occur in relation to Big Data. We will also explore solutions to these issues in terms of data quality and data governance. It is necessary to establish a system for introduction and operation to recognize risk factors and to relieve and remove such risk factors. In this study, we improve the success of overall implementation, stable service operation, and effects through elimination of such risks.

2. Application possibility of big data

Big data services can be more meaningfully used by public sectors than by private companies or individuals. To analyze trends of the country and society, to enhance public services, and ultimately improve quality of life are possible by government and public institutions using data. According to McKinsey, in the case of the United States, use of Big Data in healthcare, public administration, retail, manufacturing and personal information sectors can further improve productivity by 1%, that can be translated into economic effects worth at least \$1-\$7 billion USD (Manyika, et al., 2011). *The Economist* magazine expresses a perspective like McKinsey's analysis that data will serve as new raw material for business at almost the same level as capital or labor, and Big Data will facilitate identification of business trends, solve crimes, and prevent diseases (Lee, 2012).

In the United States, there are some cases in which Big Data is used for public safety by introducing a prediction system based on Big Data. In the case of South Korea, the government has made various efforts to use Big Data services, and the National Informatization Strategy Committee is using Big Data analytics to improve efficiency and effectiveness of government operations, transparency of public administration, and provision of customized national services.

The Korea Internet and Security Agency (KISA) launched the "Big Data National Strategy Forum" twice to support smart government operations and strategic planning in October 2011 and in April 2012. The Korean government and public institutions have proposed a public service scenario using Big Data services. New knowledge that can be found from Big Data analytics has unlimited potential, and various service application scenarios can be endlessly created in the future (Lee, et al., 2012).

3. Big data's issues and risks

3.1 Personal Information Leakage and Privacy Infringement

Due to ubiquitous use of social media platforms, personal information is disclosed to the public by users, but sometimes it is revealed by their friends and acquaintances regardless of their intention. Such disclosed information includes text messages, photos, video clips, etc. Anyone without a desirable intention may use Big Data technology to infringe on the privacy of a certain person. Unintended information can be created without the knowledge of the concerned person, which includes location information, search information created while surfing on the Internet, traveling route information, and time and place information of phone calls. These types of information are called 'Digital Shadow' and this data has a quantitative nature of being created on a large scale. Digital shadow data has high application value and can provide meaningful implications if this information is connected with time (Gantz & David, 2011). Various scenarios are possible with time and place data combined and it can be used for negative purposes.

If personal information is used for business purposes, there is no problem because everyone has consented to collect and use his or her information. But risk of using personal information with malicious

intention cannot be avoided. In the past when Big Data solutions were not technically realized, even when personal information about individuals were leaked, subsequent ripple effects were insignificant, because there was no technology available to conduct in-depth analysis and infer more information by linking it with related data. However, due to advancement of technology, it becomes possible to produce a 3D personal profile and to predict a person's intention. Although such Big Data technology can contribute to boosting crime prevention and enhance security systems when it is used to arrest criminals and for national security, it cannot exclude a possibility that an innocent citizen's personal information may be used by the government as well as a group or organization for a malicious purpose.

In Europe, the term referring to personal information leakage is "personal data protection". Categories of information subject to personal data protection include demographic characteristics such as name and age as well as various types of unstructured data such as pictures, photos, and images that can be used to describe or identify individuals (Kim, 2012). A type of information that is most likely to cause privacy problems is personal or location information uploaded to social media. This information belongs to the domain of consumer analysis or business analysis, and to the target of the government's monitoring (Boyd & Crawford, 2011). Data belonging to this domain is used to identify a target group for an advertisement from a perspective of marketing research, or become the subject of analysis to arrest spies or prevent crimes from a perspective of national security (Smith, et al., 2012).

Smith, et al. (2012) expressed concern about the seriousness of a possible breach of privacy by citing that recent mobile communication devices are mounted with location information functions, through which geographical information is generated. Foursquare is the service that conducts functions that use location information, and users can be served personalized online favorites in return for disclosing their location. Location information will be openly available on Google Maps or Yelp and then will be linked to social networking services such as Facebook Places and Google Latitude. Due to the rapid proliferation of smartphones, commercial availability of such location information becomes increasingly significant, whereas risks of personal information leakage and privacy infringement are also on the rise.

3.2 Concern about data monopoly

Concern about data monopoly is an issue that is constantly raised with the invention of computers and development of the Internet. Companies store customers' information and analyze it to provide customers' convenience but paradoxically, it means that customers' private lives are monitored by someone else. This will become more serious if the information is centralized and monopolized by a specific organization. The seriousness is more intensified because the ability to collect, analyze and visualize data using Big Data technologies is expected to develop to a more advanced level. However, it is necessary to disclose a vast amount of the government's data for use of Big Data services for the sake of public interest. Companies and government want to use Big Data technologies and to devise an institutional vehicle but there is needed some role or organization to prevent data monopoly through mutual checks and balances.

The vast amount of data that government store is controlled by an independent department responsible for data management, and those companies that execute services for the public have their department and staff in charge of managing data. They are responsible for maintenance of data, protection of personal information, authority to determine scope of data disclosure, and quality of data. The directions for preparation of institutional vehicles are the determination of kinds of data it will disclose and to what extent, the realization of common interests, and prevention of side effects (Kim, 2013).

Related regulations should be established in harmony with existing laws relating to communication and personal information protection. Currently, South Korea has many laws relating to Big Data, including the Constitution, the Act on Personal Information Protection, the Act on Act on Promotion of Information and Communication Network Utilization and Information Protection, the Act on Use and Protection of Credit Information, the Act on the Consumer Protection in Electronic Commerce, the Act on Protection of Communications Secrets, and Act on Informatization Promotion, etc. To avoid conflict with these existing laws, new laws relating to sharing and use of information for Big Data services must be established, and existing laws must be revised.

3.3 Responsibility for Data Quality and Service

In case that public data is disclosed for Big Data services and used for commercial business purposes by private organizations such as companies, risks can be high in decisions made based on analysis results, if there is a reliability problem of the concerned data. It cannot be guaranteed that the source data has 100% complete quality, and, if we discuss who is responsible for data quality problems, it will make it difficult to disclose data. Issues of data coherence and consistency are common issues in Big Data services as well as in data processes of organizations. However, in the case of the quality problem of source data, an adequate level of quality assurance is necessary, as purposes and interests between organizations may be different. But it is not easy to determine which level is adequate. Unless such appropriateness level and responsibility are determined in advance, it can be the cause of various disputes or an endless debate on who is responsible for results of a poor decision based on Big Data analytics.

Apart from the quality problem of data managed by public agencies, data searched and extracted from the Internet would have quality problems. Fundamentally, data that can be openly searched on the Internet has been circulated for a lengthy period, and as there is inaccurate information, we cannot discern what is accurate, or there may be intentionally misleading data. Therefore, there is a problem of how to evaluate quality of such data and how to determine which data is useful (Bizer, et al., 2012). As data contained in social media may have various kinds of humor or distorted information, such issues as to the extent to which we can trust data and how to use visualized content in decision-making as a way of use can occur.

There may be problems concerning data analysis rules of Big Data from reliability of source data. The purpose of Big Data analysis is to discover a data rule that source data has and to use it to develop strategy or predict a trend (Lee, 2012). This analysis process is mechanically conducted by software, but there may be some cases wherein software analysis process does not work properly. Review and verification of such problems are preconditions for deriving appropriate results from Big Data analysis. When new services using Big Data are provided, performance or reliability is another problem. In terms of IT services, there can be issues arising relating to guaranty of an appropriate level of response time and provision of a stable service during peak time when a server is flooded with simultaneous access. There can be some Big Data service cases requiring more resources for analysis and visualization, and if resources are not prepared sufficiently, service latency increases too much, or the system can be shut down if overwhelmed by too many requests for services.

4. Data governance

Management domain of data has broadened from operation database of existing OLTP to data warehouse or data mining of OLAP, and we are faced with a situation wherein it is expanding into Big Data. From the perspective of data governance, the scope of control has been expanded, and so has the scope of its necessity. Power (2011) posits that the purpose of data governance is to effectively define, manage, and share corporate data from a perspective of enterprises. Kumar (2008) suggests that the purpose of data governance is data accountability, security improvement, reduction of overall costs, consistency between data and business functions, and provision of quality. Data governance is not a technical application but about policies, organizations, standards and guidelines. Introduction of technology without responsible organization or policy preparation can increase risks. Data governance is needed to provide and share accurate and complete information about current status with stakeholders.

Data governance, from a perspective of data, enables an organization to realize systematic data standardization and integrated management, an efficient management of application systems for data, establishment of related organizations and processes, policy formulation, and establishment of business processes. However, to establish an effective data governance system, it must be in connection with corporate governance, IT governance, and ITA/EA (Information Technology Architecture/Enterprise Architecture) from a company-wide perspective (KIPA, 2008). Corporate governance is defined as "the set of processes, customers, policies, laws, and institutions affecting the way a corporation is directed, administered or controlled" (Monks & Minow, 2008). IT governance is a mature discipline like corporate governance and International Organization for Standardization (ISO) has established IT governance processes (Panian, 2010). IT governance defined as "the leadership and organizational structures and

processes that ensure that the organization's IT sustains and extends the organization's strategies and objectives" (IT Governance Institute, 2003). Information Technology Architecture (ITA) is a high-level blueprint to view the IT assets to operate data consumed by users as information. ITA would be aligned to Enterprise Architecture (EA) and EA sometimes is changeable with ITA to make effectiveness and efficiency with process re-engineering or IT solution introduction for Big Data services. Data governance is not a full subset of IT governance (Wende, 2007) and IT governance is neither.

It is necessary to appoint a manager or team responsible for quality of data responsibility to perceive data as a crucial asset of an organization and to create new value through data processing and analytics. Responsible departments and managers play key roles in establishing policies and processes for data in accordance with data governance. Standardizing data and processes based on these policies and guidelines are added roles of managers. Data standardization, policies & processes, and organizations are key components of data governance (KIPA, 2008). These components pursue the same data attributes as ordinary data: accessibility, availability, quality, consistency, security, and auditability (Panian, 2010).

Data governance requires a framework as a strategy for overall data transparency and use and the data governance framework suggests data collection strategies, process methods to support data integration and information management (Kim, S., 2011). Definition of access, control, and accountability of data can be specified by a framework for data governance. Data governance framework is used as a framework for determining disclosure scope of data, defining responsibility for data quality, and providing standards for a stable data service. The data governance framework enables a systematic organization of thoughts and communications about complex and ambiguous concepts (Thomas, 2006). Among components of data governance, "Standardization" defines an organization's standard data model with related technologies and tools. "Policy & Process" suggests policies for processes from data creation to disposal, including access and transfer of data, supervision and evaluation of these processes. "Policy & Process" ensures that these processes are implemented in accordance with those policies. "Organization" defines roles and responsibilities of staffs, to educate and train them about technology and process. "Organization" makes methods needed to manage data and to encourage staffs in changing the organization's direction. Data integration infrastructure is a technical component that automates processes based on systematic support and guarantees data quality (Panian, 2010).

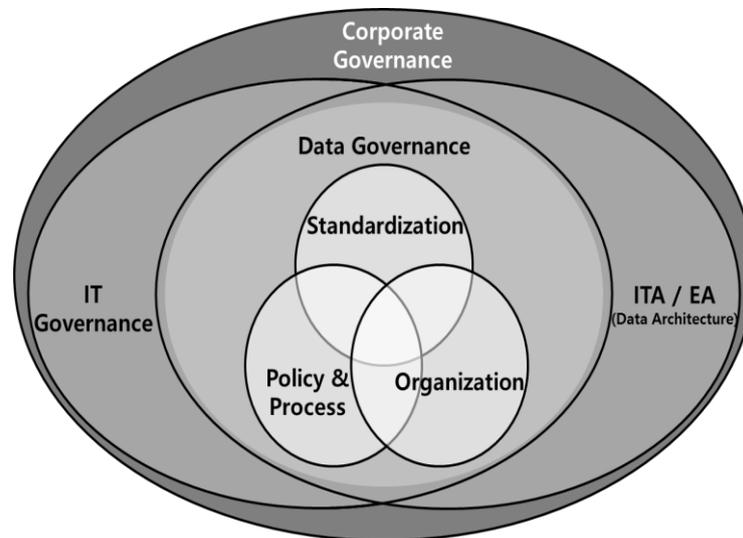


Fig. 1 the Scope of Corporate Governance and Data Governance¹

Apart from Big Data issues, improving quality of data is another issue that is difficult. There may be many reasons for this, but the most representative one is a complex business relationship. A lack of data integration, an absence of data quality management policies, and limited time and resources are

¹KIPA (Korea Internet and Security Agency), (2008) "The importance of data governance and data quality management", *SW Industry trend*, Oct.

reasons for failure of data quality. Rapid growth of an organization and simultaneous occurrence of a project to develop an information system to support it can cause a duplication of data and inconsistency among data. Integrated consolidation of these data is possible not with a one-time event of a large-scale reorganization of databases but with continuous effort of an organization's specialized team. However, as this task is not highly valued in reality, a vast amount of data is duplicate managed by multiple departments and teams in changing organizations. Overlapping data are slowly losing consistency. If a migration process is conducted due to necessity for data integration, the problem concerning decision-making about which data is viable among inconsistent data is constantly raised. Decision-making and business intelligence rely on quality of data (Shankaranarayan, et al., 2003; Price & Shanks, 2005).

Figure 2 is the Data Governance Framework of Panian (2010) and quality of data. It is viable base to enlarge the scope of Big Data services. Big Data services need quality of data but must be open to corporate governance for successful services implementation to achieve the objective of the organization.

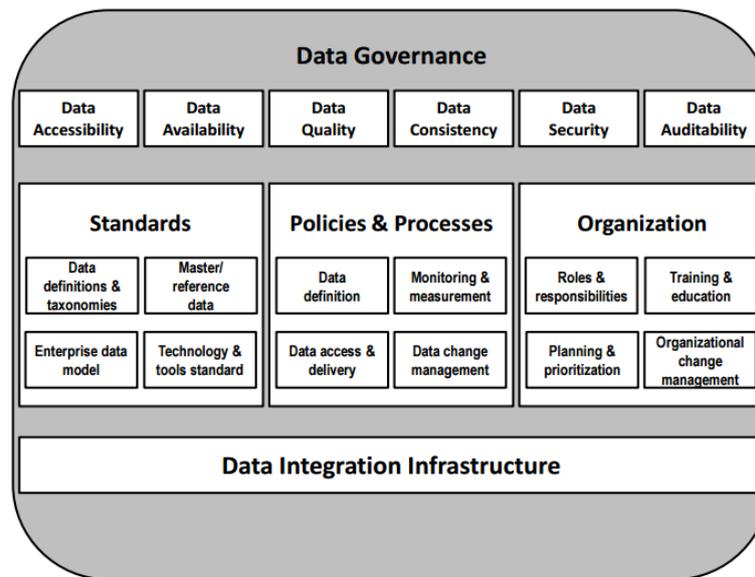


Fig. 2The Data Governance Framework of Panian²

5. Data Governance framework for big data

5.1 The limits of data governance for Big Data implementation

Academic research about Big Data governance is in infancy. Data governance focuses on safe management of data because it does not have a particular system in mind. No matter what system is developed, there is no problem with providing faithful services from the perspective of data. However, it is not enough to implement Big Data services with various analytics. What is the best way to realize Big Data services while protecting personal information and privacy, preventing monopoly of information, and guaranteeing appropriate data quality? It is to establish a system that can prepare an infrastructure for Big Data services based on data governance and provide stable services through a balance of responsibility and authority. Big Data should store and process a much larger amount of data than existing ordinary data, in addition to unstructured and semi-structured data. Big Data services should address precise and minute data such as mechanical sensor results or log information automatically generated by sensors or embedded systems but there is uncertainty regarding the source of data generation and ownership of data. However, it is necessary to develop an approach for quality management adequate to characteristics of Big Data (Edjlali & Friedman, 2011). Edjlali and Friedman (2011) suggest quality management approach for Big Data that is different from existing data management methods that ignore data users' errors or sometimes skip verifying validity of individual data.

²Panian, Z., (2010) "Some Practical Experiences in Data Governance", *World Academy of Science, Engineering and Technology*.

Data quality in Big Data emphasizes data attributes different from those of existing data analysis systems with OLTP (Online Transaction Processing) characteristics. As seen above, Panian's data governance framework (2010) has basic data attributes of the IT field. Data attributes of Big Data should be defined differently from what is suggested by Edjlali and Friedman (2011) because it can be difficult to achieve the goal of Big Data services. Results of Big Data services have no meaning if attributes of ordinary data are applied without consideration of unique characteristics of Big Data. In this study, data attributes of Big Data are classified into timeliness, trustfulness, meaningfulness and sufficiency.

First, timeliness means that data should be prepared in a timely manner to be adequate for an analysis purpose of Big Data. If timing is missed, it may be impossible to achieve the goal of Big Data services. Second, trustfulness means how much source data can be trusted and if it can prove validity of analysis results. It is likely that there is not much time to verify accuracy of data values and to evaluate their consistency. Trustfulness must be verified after Big Data services and if there is a fault in data processing of Big Data analysis, maintenance efforts would be needed. Third, though trustfulness level is sufficient to make a probabilistic inference about analysis results, it is possible to respond to risk of fallible decision-making. Meaningfulness is a question of if data used for Big Data can provide meaning as a topic appropriate for analysis. If one processes and analyzes a large amount of meaningless data, subsequent results to be visualized into resultant values cannot provide meaning. For example, if source data is about air humidity among data relating to weather forecasting, meaningfulness means that it is possible to provide a meaningful result by identifying correlation between humidity and rain probabilities. Last, sufficiency means if it is possible to provide adequate data to an extent that a business domain and purpose of an organization can be achieved. Sufficiency can be interpreted as if it can have time and space attributes adequately to derive analysis results. Sufficiency makes it possible to derive analysis results as values for future prediction.

5.2 The Big Data Governance Framework

The Big Data Governance Framework presented in this study is in Figure 3, and Big Data governance must create new analytics and resulted values along with objective of the organization. What values the Big Data service will reveal would be clarified based on purpose of the declaration. The strategy should be formulated to achieve the objective. Protecting personal information, preserving the level of data quality, and defining data responsibility are key strategies to winning. Failure of strategies is the main reason to suspend Big Data services because reliability of organization and trust from customers are no longer maintained.

Data processing must be controlled by independent audit authority of IT departments from collection of data to visualization results. Auditors must periodically ensure each phase of data processing and monitor possible risks in accordance with collection, processing, analysis and visualization phases. Audit results should identify which stage was problematic and enhance the process to block erroneous prediction or indication results of Big Data services.

The goal of Big Data services is to derive implications and create values with data through a new analysis method. These goals can vary depending on the nature and mission of organizations and can be more clearly defined in the planning phase of Big Data services. The goal of Big Data Governance Framework is to implement Big Data projects successfully to prevent side effects such as personal information leakage or privacy breaches. New standards for data quality of Big Data rather than attributes of data quality in existing data governance must be presented, and therefore, it is necessary to establish new quality standards through new definitions for data created without particular purpose or control and which has uncertain ownership (Edjlali & Friedman, 2011). To achieve the goal of Big Data services, it is necessary to assess if appropriate data are prepared and if data can be analyzed for precise results. Timeliness is a critical factor in making prudent decisions, and if the right moment is missed due to excessive efforts to improve trustfulness in providing more accurate predictability, decision-making is meaningless. However, all types of decision making must consider trustfulness and timeliness at same time. Meaningfulness and sufficiency attributes of data should be considered, for example, which kind of meaningful data will be used for analysis of behavioral patterns of citizens and how much data will be viable.

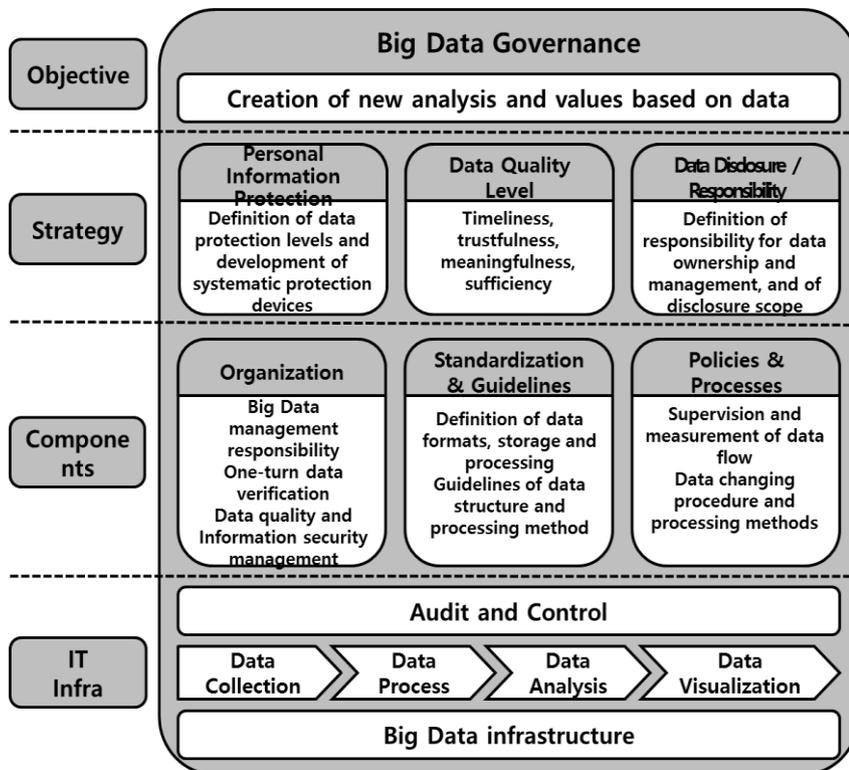


Fig. 3The Big Data Governance Framework

In terms of data disclosure scope and data management responsibility, data maintenance management and accountability should be defined clearly, even if unrelated to personal information. It is also needed to enable determination of the level of presumed trustfulness of data analysis results. Although data governance for Big Data has almost the same components as existing data governance of Panian (2010), it is necessary to form an organization required for expanded application to the domain of Big Data, to establish standards and guidelines, and to prepare for data flow and management processes. Implementation of a strategy can be realized by preparing components required to execute data governance. IT infrastructure for Big Data will require new solutions and computing resources. If it is equipped with proper Big Data governance, it is possible to make decisions to ensure selection of appropriate solutions and maintenance of balanced and appropriate service levels about investment in IT infrastructure.

Big Data infrastructure may vary in scale depending on business models and establishment strategies, but it must provide the basis to achieve the goal of Big Data services. The audit and control of a system is to monitor guidelines and procedures that must be maintained in terms of infrastructure and components. If monitoring results indicate that immediate reporting is needed, reporting will be conducted through procedural processes in accordance with policies and processes.

5.3 Case Study of the National Pension Service of South Korea

The National Pension Service (NPS) of South Korea is the 4th largest pension organization in the world and the volume of assets would be 440 trillion Korea won (approximately \$380 billion USD), following Japan's GPIF (1st), Norway's GPF (2nd), and the Netherlands' ABP (3rd). The NPS is expected to disclose national pension data by Korea government in line with its open policy. It is called 'Government 3.0'. South Korea has a new movement to open public data and it is called 'Public Data Use Promotion Project'. Disclosed data can be used for Big Data services to enhance the quality of life of people and would be manipulated to create new business opportunities such as traffic information services for citizens and hospital search services by smartphone apps.

As of the end of 2013, the NPS has stored 20 terabytes of data and 74.7 billion cases. Nearly 3.4 billion cases are related to subscription management, 28 billion cases are related to insurance payments,

and 4.5 billion cases are related to salary information (Lee, 2013). The NPS has applied Big Data analysis techniques to CRM for analysis of trends in the public opinion. The NPS can respond to customers on possible influences against negative opinions. Public opinion trend analysis of the NPS is conducted by integrating internal customer consultation information and various kinds of external unstructured data. It is necessary to identify the changing trend of public opinion in a sensitive manner because almost the entire population of Korea has subscribed to the NPS and newly recruited or business people are targeted for subscription as a mandatory option. It is the main part of an effort by the NPS to mitigate the public's distrust about the NPS and improve job efficiency through responding more actively to customers' changing demands (Jung, 2012). Analysis results about customers' needs are applied to customer-tailored campaigns and it is another effort towards achieving customers' satisfaction.

The goal of the NPS's Big Data services is to enhance value of services and implement efficient administration. This is the NPS's vision, "National Pension for a Successful Life". The NPS is operating the service under two departments: Information Protection Department and the Personal Information Protection Department. The Personal Information Protection Department is responsible for establishing and operating personal information protection regulations, enforcement rules, and internal management plans. South Korea has strengthened related laws such as enforcing leakage of personal information as a criminal penalty. It is different from most of the other countries in the world. Most countries impose civil fines for privacy violations. In South Korea, leaked personal information is often used by criminal organizations in China and the Philippines for voice phishing and threats, emerging as major social issues. For prevention of cyber infringement and response to it, the NPS established the Information Protection Department as a separate organization that responsible for improvement of security vulnerability, operation of a security system, and control service of cyber threats.

Regarding data disclosure, the NPS is pursuing openness, sharing, communication and cooperation by the South Korean government's 'Government 3.0' policy, and discloses 96 sample items and 124 unique items. Sample items are lists of standardized information stored by public institutions and enables the public to access objects of publication at a glance. Unique items are lists of information autonomously disclosed by public institutions to enable the public to access various kinds of information.

It is necessary to examine the NPS' data quality in terms of timeliness in accordance with Big Data governance framework. Performance of the real-time service is monitored by the performance management tool and abnormal transactions are selected to improve (Kim, 2013). Reliability is difficult to verify, but the dedicated department for data quality management verifies reliability regularly and improves company-wide data quality. In terms of significance and sufficiency of data, the NPS' Big Data services have been successful in meeting the qualification of meaningfulness and sufficiency, as practically all citizens make contributions to the national pension fund's information quality.

Data processes of the NPS's Big Data services consist of four steps: Extract, Transform, Cleanse, and Load. A data flow analysis system is constructed and is used to manage data flow in structured form that can be traced through a schematized diagram (Kim, 2013). Enterprise-wide metadata management system conducts procedural management for data standards and integration for data structure. The metadata system conducts management of database alignment in terms of structure consistency. The full-scale Big Data services of the NPS was launched in 2016.

6. Conclusion

Big Data services can be implemented easily with introduction of Big Data solutions. However, without sufficient consideration of strategic factors in advance from the perspective of data governance, risks would increase. Privacy problems, information monopoly, and responsibility of data quality are risks of Big Data services and would be obstacles to achieving meaningful analysis results to meet the goal of an organization or the needs of customers. Big Data adoption projects require more investment than previous IT solution adoption projects and would be risky to integrate legacy systems. Ripple effects of expected problems including risk of personal information leakage, harm caused by information monopoly, and the importance of decision-making based on analysis results are far greater than other solutions.

A Big Data project requires new standards different from data quality standards established based on existing data that are timeliness, accuracy, sufficiency and meaningfulness. Data disclosure and

accountability issues as well as policymaking in consideration of the above two factors are the most overlooked in projects, that must be addressed in an early phase of a Big Data project. If these are not determined until the end of the project, it is likely to face a crisis from policy failure. A pilot project for the realization of Big Data services can be conducted prior to evaluating technical feasibility, but it is imperative to assess kinds of interests among stakeholders correlated in processes of data collection and disclosure of analysis results.

In this study, we propose Big Data governance framework to prevent policy failure by summarizing considerations in advance to make a Big Data project successful. For implementation of Big Data services, Big Data governance framework can be used as a basis for analyzing risk factors and preventing problems. In this case study, we examined actual case in a perspective of Big Data governance framework. Additionally, we discussed various possibilities that may occur in projects that were expected to launch Big Data services. It is not easy to study big data governance based on actual cases, and it can be seen as a contribution of this study. Based on this study, we hope that the Big Data governance framework presented in many other cases will be further verified and developed. Big Data governance framework can be used as part of an effort to enable analysis of criteria that can be considered. Big Data services in the public sector are an evitable choice for improving the quality of life of people. For successful realization of Big Data services, data governance and subsequently required Big Data governance framework are critical factors that must be considered.

Acknowledgement

This study was supported by Hankuk University of Foreign Studies Research Fund of 2018.

References

- Bizer, C., Boncz, P., Brodie, M. L., Erling, O., (2012) "The Meaningful Use of Big Data: Four Perspectives – Four Challenges", SIGMOD Record, Vol. 40, No. 4, pp.56-60
- Boyd, D., Crawford, K., (2011) "Six Provocations for Big Data", SSRN eLibrary.
- Edjlali, R. & Friedman, T., (2011) "Data quality for Big Data: Principles remain, but tactics change" Gartner
- Gantz, J., David R., (2011) "Extracting value from chaos." IDC review 1142: 9-10.
- IT Governance Institute, Board Briefing on IT Governance, 2nd ed., Rolling Meadows, IL, pp.6-7.
- Jung, Y., (2012) "Big Data Revolution and Media Policy Issues," KISDI Premium Report, 12(2)
- Lee, S., (2013), "National Pension Service of Korea, preparing the base of BigData Analysis ... Consulting and Pilot is On," Digital Daily, 2013.11.23. <http://www.ddaily.co.kr/news>
- Kim, B. (2012) "Smart Era, Data Protection Strategy and the Law II", Korea Academy Information, Kyunggido(Korea)
- Kim, H. Y., (2013) "Data Governance for Big Data and Medical Information Service practice", Information System Audit and Control, Vol. 25, pp.18-27.
- Kim, S., (2011) "A Case Study of Implementation Data Governance for Enterprise Architecture," Journal of Information Technology and Architecture, Vol. 8. No. 3, pp.255-265
- Kim, S., (2013) "The Analysis of Data Governance model for Business and IT Alignment," Journal of The Korea Society of Computer and Information, Vol. 18, No. 7, pp.69-78.
- KIPA (Korea Internet and Security Agency), (2008) "The importance of data governance and data quality management", SW Industry trend, Oct.
- Kumar, S., (2008) "Data governance: An approach to effective data management." White paper, Satyam Computer Services, Ltd.
- Lee, J., (2012) "Data Big bang, the trend of Big Data" Journal of Communications & Radio Spectrum, vol. 47, pp.43-55.
- Lee, K. Y., Nam, G. H., Sim J., Cho, G., Ryu, W., (2012) "Construction of Knowledge Base for The Utilization of Big Data in Public Domain" Communications of the Korea Information Science Society 30(6), pp.40-46
- Manyika, J., Chui, M., Brown, B., Bughin, J., Dobbs, R., Roxburgh, C., & Byers, A. H. (2011). Big Data: The next frontier for innovation, competition, and productivity.
- Monks, R. A. G., Minow, N., (2008). Corporate Governance, Chichester, England: John Wiley and Sons Ltd, p.14.

-
- Panian, Z., (2010) "Some Practical Experiences in Data Governance", World Academy of Science, Engineering and Technology.
- Power, D., (2011) "A 4-D Approach to Data Governance", Information Management, May, pp.29-30.
- Price, R. & Shanks, G. (2005) "A semiotic information quality framework: development and comparative analysis", Journal of Information Technology, 2005(20), pp.88-102.
- Shankaranarayan, G., Ziad, M., Wang, R. Y., (2003) "Managing Data Quality in Dynamic Decision Environments: An Information Product Approach", Journal of Database Management, 14(4), pp.14-32.
- Smith, M., Szongott, C., Henne, B., Voigt, G. V., (2012) "Big Data Privacy Issues in Public Social Media", 2012 6th IEEE International Conference, June, pp.18-20.
- Thomas, G., (2006) "The DGI data governance framework." The Data Governance Institute, Orlando, FL (USA)
- Wende, K., (2007) "A model for data governance-organization accountabilities for data quality management", 18th Australasian Conference on Information Systems, Toowoomba. pp.417-424.
-