

Data Governance Framework for Big Data Implementation with a Case of Korea

Hee Yeong Kim
College of Business
Hankuk University of Foreign Studies
Seoul, Korea
heeykim0@naver.com

June-Suh Cho
College of Business
Hankuk University of Foreign Studies
Seoul, Korea
jscho@hufs.ac.kr

Abstract - Big Data governance requires a data governance that can satisfy the needs for corporate governance, IT governance, and ITA/EA. While the existing data governance focuses on the processing of structured data, Big Data governance needs to be established in consideration of a broad sense of Big Data services including unstructured data. To achieve the goals of Big Data, strategies need to be established together with goals that are aligned with the vision and objective of an organization. In addition to the preparation of the IT infrastructure, a proper preparation of the components is required to effectively implement the strategy for Big Data services. We propose the Big Data Governance Framework in this paper.

The Big Data governance framework presents criteria different from existing criteria at the data quality level. It focuses on timely, reliable, meaningful, and sufficient data services, focusing on what data attributes should be achieved based on the data attributes of Big Data services. In addition to the quality level of Big Data, the personal information protection strategy and the data disclosure/accountability strategy are also needed to achieve goals and to prevent problems.

This paper performed case analysis based on the Big Data Governance Framework with the National Pension Service of South Korea. Big Data services in the public sector are an inevitable choice to improve the quality of people's life. Big Data governance and its framework are the essential components for the realization of Big Data service.

Keywords: *Big data, Data governance, Data governance framework, Case analysis.*

I. INTRODUCTION

In Korea, Big Data is a major concern for both the government and enterprises. In addition, IT-based marketing strategies are more actively implemented in South Korea than any other countries around the globe, as the South Korean government is leading the disclosure of data and supporting private enterprises to utilize the disclosed information to start new commercial services. However, there are some cases where Big Data solution providers advertise the exaggerated contents, which may cause some unreasonable expectations. This hyped expectation only led to paying too much money for the introduction of solutions, but the service effects fell far short of meeting the expectations. The heads of the IT organizations cannot be too careful in starting projects to introduce Big Data solutions due to the overblown expectations of the CEOs. From a viewpoint of an IT expert,

more attention should be paid to how to operate Big Data solutions after the introduction.

From a viewpoint of leaders of IT organizations, there are issues that need to be addressed right away with the introduction of Big Data solutions, such as security against leakage of personal information and accurate calculation of financial resources needed for a heavy investment in infrastructure. Apart from the IT-related issues, the introduction of Big Data solutions requires the estimation of an organization's effect against investment and the formulation of a strategy to achieve it. Especially in terms of the reliability of analysis results, if there is a difference between predicted values provided by Big Data solutions and actual results, the decision making about the introduction of solutions will be challenged.

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

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

As research methods, we first look at what kinds of issues can possibly occur in relation to Big Data. And, we will also explore solutions to these issues in terms of data quality and data governance. As Big Data is more large-scaled and risky than the operation of the existing structured data, it is necessary to establish a system for the introduction and operation to recognize risk factors and to make an attempt to relieve and remove such risk factors. In this study, we try to improve the success of overall implementation, including the

introduction of big data solutions, stable service operation, and improvement of effects through the elimination of such risks.

II. APPLICATION POSSIBILITY OF BIG DATA

Big data can be more meaningfully utilized by the public sector than by private enterprise or individuals. It is because it makes it possible to utilize the data provided by the government and public institutions, to analyze the trends of the country and society, to enhance public services, and ultimately improve the quality of life. According to McKinsey, in the case of the US, the utilization of Big Data in healthcare, public administration, retail, manufacturing and personal information sectors can further improve productivity by 1%, which can be translated into economic effects worth at least \$ 100 billion to \$ 700 billion by sector[15]. The Economist Magazine also expresses a viewpoint similar to McKinsey's analysis that data will serve as a new raw material for business at almost the same level as capital or labor, and Big Data will help identify business trends, solve crimes, and prevent diseases[13].

In the United States, there are cases in which Big Data is utilized for public safety by introducing the collection, analysis, and forecasting system of Big Data of the whole population for the purpose of crime prevention and counter-terrorism centering on the Department of Homeland Security. In the case of South Korea, the government has made various efforts to utilize Big Data in the public sector, and the National Informatization Strategy Committee is using the Big Data to improve the efficiency of the government operation, the transparency of the state and the provision of customized national services. (In Korean) "in October 2011. In addition, in April 2012, the Korea Information Technology Agency (KISA) launched the "Big Data National Strategy Forum" to support smart government operations and strategic planning. The Korean government and public institutions have already proposed a public service scenario using Big Data. New knowledge that can be found from Big Data has an unlimited potential, and various service application scenarios can be endlessly created in future[14].

III. BIG DATA'S ISSUES

A. Personal Information Leakage and Privacy Infringement

The first issues to be considered in relation to Big Data are the leakage of personal information and the infringement of privacy. Due to a ubiquitous use of social media platforms, personal information is disclosed to the public by users themselves, but sometimes it is revealed by their friends and acquaintances regardless of their intentions to do so. Such disclosed information includes text messages, photos, video data, etc. and someone with an undesirable intention might use it to infringe on the privacy of a certain person by using Big Data analysis technology. Especially, unintended information can be created without being known to 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 as this data is objective and has a quantitative nature of being created at a large scale, it has a high application value and can provide meaningful implications if this information is connected with time[4]. And various scenarios can be analyzed if they are connected with time information.

If personal information is used for business purposes, it can be said that there is no problem because each individual has consented to the collection and use of personal information, but the responsibility for the risk that the data may be used out of an undesirable and malicious intention cannot be avoided. In the past when Big Data solutions were not technically realized, even when personal information about individuals were leaked, the subsequent ripple effects were not significant, because there was no technology available to conduct in-depth analysis and to infer more data by linking it with the related data. However, due to the 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 the crime prevention effect and enhancing the security system when it is used to arrest criminals and for the sake of national security, it cannot exclude a possibility that an innocent citizen's personal information might be used not by the government but by a group or organization for a malicious purpose.

In Europe, the term to refer to the protection of personal data is 'personal data protection'. The categories of information which are subject to personal data protection include not only demographic characteristics such as name and age but also they tend to include various types of unstructured data such as pictures, photos, and images that can be used to describe or identify individuals[7]. A type of information that is most likely to cause privacy problems is personal information or location information that is uploaded to social media. This information belongs to the subject of consumer analysis or business analysis, and to the target of the government's monitoring[2]. The data belonging to this domain is used to set up a target group for an advertisement from a viewpoint of marketing research, or become the subject of analysis to arrest spies or prevent crimes from a viewpoint of national security[18]. [18] expressed concern about the seriousness of a possible breach of privacy by citing that the recent mobile communication devices are mounted with location information functions, through which geographical information continues to be generated. Foursquare or Qype are the services that perform the functions that use this location information, and users can use personalized online services in return for disclosing their location. And the 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 spread of smartphones, the commercial availability of such location information becomes increasingly important, whereas the risks of personal information leakage and privacy infringement are also on the rise.

B. Concern about the Appearance of Big Brother

Concern about the appearance of the Big Brother is an issue that is constantly raised together with the invention of computers. Although companies and public institutions that have a system which can inquire or utilize a large amount of personal information data can provide convenience through services utilizing Big Data. Paradoxically, it means that citizens or customers' private lives are monitored by someone else. This will become more serious if the information is monopolized by a specific organization and the services utilizing the information becomes concentrated on a specific organization. The seriousness gets more intensified because the ability to collect, analyze and visualize data using Big Data solutions is expected to develop to a more advanced level. However, it is necessary to disclose a vast amount of the government's data for the utilization of Big Data for the sake of the public interest, to look for measures for individuals, enterprises and the government to utilize Big Data, and to devise an institutional vehicle to prevent a Big Brother type of data monopoly through mutual check and balance[8].

The vast amount of data that the government possesses is controlled by an independent department responsible for data management, and those enterprises that execute services for the public have their own department and staff in charge of the management of various kinds of data including personal communication information and credit information. They are responsible for the maintenance of data, the protection of personal information, the authority to determine the scope of data disclosure, and the quality of disclosed data. The directions for the preparation of institutional vehicles are the determination of what kinds of data it will disclose to what extent, the realization of common interests, and the prevention of side effects[8]. The related laws and regulations should be established in harmony with the existing laws and regulations relating to information and 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 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 a collision with these existing laws, new laws relating to the sharing and use of information for Big Data need to be established, or the existing laws need to be revised.

C. Responsibility for Data Quality and Service

In case that the public data is disclosed for big data and used for commercial business purposes by private organizations such as corporations or research institutes, problems can arise in decisions that are made based on the analysis results, if there is a problem with the reliability of the concerned data. It cannot be guaranteed that the source data has a 100% complete data quality, and, what is more, if we discuss who is responsible for the problems, it will make it difficult to disclose any data. The issues of data coherence and consistency are common issues not only in Big Data but also for those ordinary organizations that process a large amount of data. However, in the case of the quality problem of the source data, an adequate level of quality assurance is necessary, as the purposes and interests between organizations can be different. But it is not easy to determine which level is an adequate level. Unless such appropriateness level and responsibility are not determined in advance, it can cause various disputes or lead to an endless debate on who is responsible for the results.

Apart from the quality problem of data managed by public agencies, the data that are searched and extracted from the Internet can have a quality problem. Fundamentally, data that can be openly searched on the Internet has already been circulated for a long period of time, and as there is inaccurate information, we cannot tell what is right, or there may be intentionally misleading data, such as spams. Therefore, there is a problem of how to evaluate the quality of such data and how to determine which data is useful for utilization[1]. As the data contained in the social media may have various kinds of jokes or distorted information, such issues as how much we can trust this data and how to utilize the visualized content in decision making as a way of utilization can occur.

Apart from the reliability of the source data, there may be problems concerning the data analysis rules of Big Data. The purpose of the Big Data analysis is to discover a data rule that the source data has and to utilize it to set up a strategy or predict a trend[13]. This analysis process is mechanically performed by software, but there may be some cases where the software analysis process does not work properly. The review and verification of such problems are the preconditions for deriving appropriate results from big data analysis. Apart from data quality, when new services utilizing Big Data are provided, problems concerning the performance and reliability of the services can occur. In terms of ordinary IT services, there can be issues arising relating to the guaranty of an appropriate level of response times and the provision of a stable service during the peak time when a server is flooded with simultaneous accesses. In the case of Big Data, there can be some cases which require more resources for analysis and visualization, and if the resources are not prepared sufficiently, the service latency increases too much, or the system can be discontinued if it is overwhelmed by too many requests for the service.

IV. DATA GOVERNANCE

The management domain of data has broadened from the operation database of the existing OLTP to the data warehouse or data mining of the OLAP, and we are faced with a situation where it is now expanding into Big Data. From the standpoint of data governance, the scope of control has been expanded further, so has the scope of its necessity. [17] thinks that the purpose of data governance is to effectively define, manage, and share corporate data from a viewpoint of enterprises. [12] suggested that the purposes of data governance are 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. The introduction of technology without a responsible organization or policy preparation can increase risks. Data governance is needed to provide and share accurate and complete information about the current status with the inside stakeholders of an organization, but is also needed for the demands of outside stakeholders.

Data governance, from a viewpoint of data, enables an organization to realize a systematic data standardization and integrated management, an efficient management of application systems for data, the establishment of related organizations and processes, policy formulation, and establishment of business processes. However, to establish an effective data governance system, it needs to be set up in connection with corporate governance, IT governance, and ITA / EA from a company-wide perspective[11].

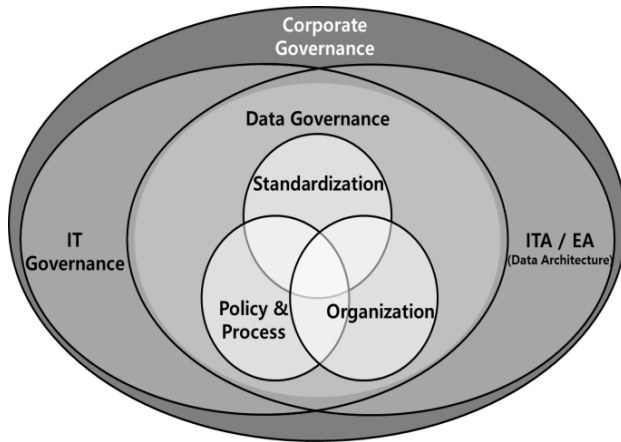


Fig. 1 I. Relationship between Data Governance and Corporate Governance, IT Governance and the ITA/EA

To perceive data as an important asset of an organization and to create new value through it, it is necessary to appoint organizations and managers who are responsible for managing the quality of data. The responsible departments and managers play the roles of establishing policies and processes for data in accordance with data governance, of standardizing data based on these policies and processes, and of managing data by guidelines. Data standardization, policies & processes, and organizations are the important

components of data governance[11]. These components pursue the same data attributes as ordinary data: accessibility, availability, quality, consistency, security, and auditability[16].

Data governance requires a framework and a load map as a strategy for overall data transparency and use, and suggests data collection strategies and methods to support data integration and corporate information management programs[9]. The definition of access, control, and accountability of data can be specified by a framework for data governance. The data governance framework is utilized as a framework for determining the disclosure scope of data, defining the responsibility for data quality, and providing the standards for a stable data service. The data governance framework enables a systematic organization of thoughts and communications about complex and ambiguous concepts[19].

Among the components of data governance, the standards make the definition of an organization's data standards and of a master data management and a standard model for them, and they also define the related technologies and tools. The policy & process is to suggest policies necessary for all processes from data creation to disposal, including access and transfer of data, and supervision and evaluation of these processes, and is to ensure that these processes are implemented in accordance with those policies. The organization is to define the roles and responsibility of the staff, to educate technology and methods needed to deal with data through education and trainings, and to manage the organization to encourage it to change in a developmental direction. The data integration infrastructure is a technical component that automates processes based on systematic support and guarantees data quality[16].

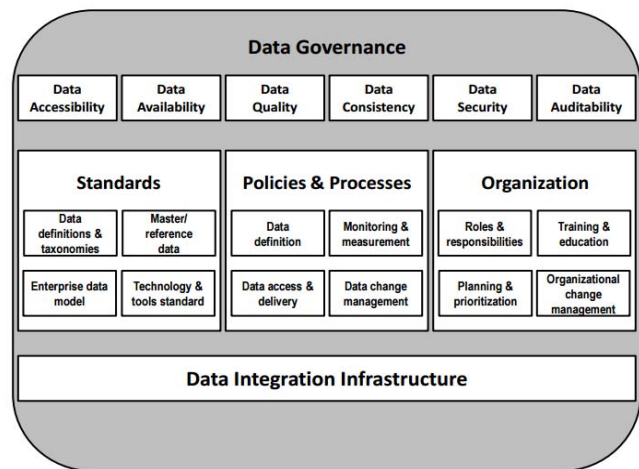


Fig. 2 Data Governance Framework

Apart from Big Data, improving the quality of data is not easy. There may be many reasons for this, but the most representative ones are a complex business relationship, a lack of data integration, an absence of data quality

management policies, and limited time and resources. A rapid growth of an organization and the simultaneous occurrence of a project to develop an information system to support it can cause a duplication of data and inconsistency between data. The integrated consolidation of these data is possible not with a one-time event of a large-scale reorganization but with a continuous effort of an organization. However, as this task is not highly valued in reality, a vast amount of data is duplicate managed by multiple organizations and the overlapping data are slowly losing consistency. If a migration process is done due to the necessity for data integration, the problem concerning decision making about which data is correct among inconsistent data is constantly being raised.

V. DATA GOVERNANCE FRAMEWORK FOR BIG DATA

A. *The limits of general data governance for Big Data implementation*

General data governance focuses on safe management of data because it does not have a particular system in mind. In other words, no matter what system is developed, there is no problem with providing faithful services from the perspective of the data. However, it is not enough to implement big data services. What is the best way to realize Big Data services while protecting personal information and privacy, preventing the 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 the existing ordinary data, in addition to unstructured and semi-structured data, and should deal with precise and minute data such as mechanical sensor results or log information that is automatically generated by programs. And it is highly likely that the source of generation and the ownership of data become uncertain. However, it is necessary to develop an approach for quality management adequate to the characteristics of Big Data[3]. [3] suggest a quality management approach that is different from the existing data management methods, in that it ignores data users' errors and sometimes skips verifying the validity of individual data.

The data quality in Big Data emphasizes data attributes different from those of the existing data analysis systems with OLTP (Online Transaction Processing) characteristics or those of data analysis methods through data warehouses. As seen above, [16] data governance has the basic data attributes of the IT field that are pursued without consideration of Big Data. The data attributes of Big Data should be defined differently from what is suggested by [3]. This is because it can be difficult to achieve the goal that Big Data pursues if the attributes of ordinary data are applied without consideration of the unique characteristics of Big Data. In this study, the data attributes of Big Data are classified into timeliness, trustfulness, meaningfulness and sufficiency.

First, timeliness means that the data should be prepared in a timely manner to be adequate to an analysis purpose of Big Data. This is because if the timing is missed, it might be possible to derive analysis values but becomes almost impossible to achieve the goal of Big Data. Second, reliability means how much the source data can be trusted and whether it can prove the validity of analysis results. In Big Data, it is highly likely that there is not much time to verify the minute accuracy of data values and to evaluate their consistency and reliability. Third, in the case of the trustfulness of source data, if a reliability level is sufficiency to make a probabilistic inference about analysis results, it is possible to respond to the risk of fallible decision making. Lastly, meaningfulness is a question of whether the data to be used for Big Data can provide meaning as a topic appropriate to the purpose of analysis. If one processes and analyzes a large amount of meaningless data, the subsequent results to be visualized into the resultant values cannot provide any meaning. For example, if the source data is about air humidity among the data relating to weather forecasting, meaningfulness means that it is possible to provide a meaningful result by identifying the correlation between humidity and rain probabilities. sufficiency means whether it is possible to provide enough data to an extent that a business domain and purpose of an organization can be achieved. Sufficiency can be interpreted as whether it can have time and space attributes enough to derive analysis results. If it is possible to derive analysis results as values for future prediction, it will be considered sufficiency.

B. *Big Data Governance Blueprint*

The blueprint of the Big Data Governance Framework presented by this study is shown in Figure 3. Unlike general data governance, Big Data governance needs to create new analytics and values. Clearly, the aim is to clarify what values the Big Data service will need to be brought to light. Based on the purpose of the declaration, a strategy should be formulated to achieve the goals of personal information protection, data quality, and data responsibility, and to ensure that each phase of audit and control is carried out in accordance with the collection, process, analysis and visualization phases. Because, unlike conventional data governance, where erroneous predictions or indications are indicated by output, one should be able to identify which stage was problematic.

The goal that data governance should pursue to realize 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 should be strategically approached in the implementation process to avoid any side effects such as personal information leakage or privacy breaches. If necessary, protection devices should be devised, and it should be possible to provide a systematical protection. New standards for data quality of Big Data rather

than the attributes of data quality in the existing data governance need be presented, and therefore, it is necessary to establish new quality standards through new definitions for the data which are created without a particular purpose or control and which has an uncertain ownership[3]. To achieve the goal of Big Data, it is necessary to assess whether appropriate data are prepared and whether they can be analyzed. Depending on the quality level of data, the level of future predictability may vary, and decision makers may be able to consider various factors at the same time. Timeliness is a criterion that considers the timing of decision making as an important factor, and if the right timing is missed due to excessive efforts to improve the trustfulness to provide a more accurate predictability, the decision making becomes meaningless. However, together with satisfaction with timeliness, all types of decision making must consider trustfulness. The meaningfulness and sufficiency attributes of data should be taken into consideration, for instance, which kind of meaningful data will be used for analysis of social changes and the behavioral patterns of citizens and how much data will be good enough to make the most of the analysis results.

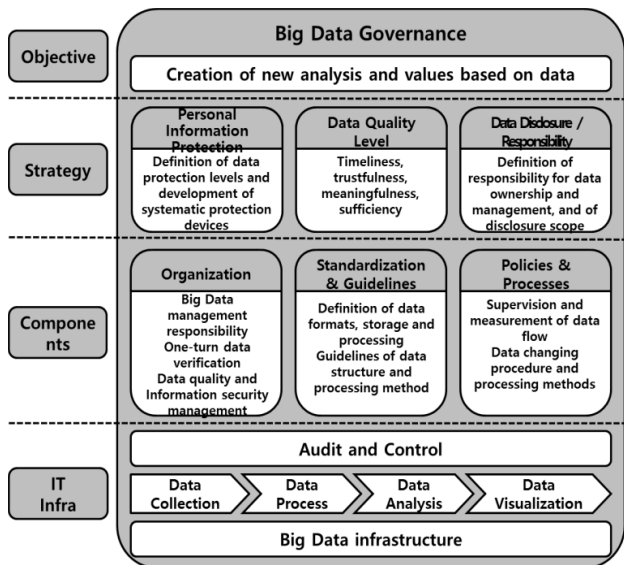


Fig. 3 Framework for Big Data Governance

In terms of data disclosure scope and data management responsibility, the data maintenance management and accountability should be defined clearly, even if they are not related to personal information, to prevent data monopolization and to achieve the goal of Big Data. It also needs to help determine the level of presumed trustfulness of data analysis results. Although the data governance for Big Data has almost the same components as the existing data governance, it is necessary to form an organization required for an expanded application to the domain of Big Data, to

establish standards and guidelines for Big Data, and to prepare for data flow and management processes for Big Data. The implementation of a strategy can be realized by preparing the components required to execute data governance. The IT infrastructure for Big Data will require new solutions and computing resources. If it is equipped with a proper big data governance, it is possible to make decisions to ensure the selection of appropriate solutions and the maintenance of balanced and appropriate service levels about investment in IT infrastructure.

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

C. Case Study and Analysis of Big Data Systems from a viewpoint of Big Data Governance Framework with the National Pension Service of South Korea

The National Pension Service (NPS) of South Korea is the four largest pension fund worth 440 trillion won in assets, following Japan's GPIF (the 1st), Norway's GPF (the 2nd), and the Netherlands' ABP (the 3rd). The NPS is expected to disclose the national pension data in 2016 by South Korea's 'Government 3.0' policy and 'Public Data Use Promotion Project'. The disclosed data can be used for Big Data services, and will contribute to creating new business opportunities and public services such as traffic information and hospital search.

As of the end of 2013, the NPS possesses 20 terabytes of data and 74.7 billion cases of data, and if we look into the types of data, 3.4 billion cases of data are related to subscribers' management; 28 billion cases are related to insurance fee charges and payments; 4.5 billion cases are related to salary decision[13]. The NPS already applied Big Data analysis techniques to CRM for analysis of trends in the public opinion. The purpose of the public opinion trend analysis is to analyze the flow and changing trend of the public opinion and to respond to its possible influence on consumers. The public opinion trend analysis of the NPS is conducted by integrating the internal customer consultation information and various kinds of external unstructured data. Given that the national pension service is targeted at the entire population, it is necessary to identify the changing trend of the public opinion in a sensitive manner. It is part of an effort to alleviate the public's distrust about the national pension fund and to improve job efficiency through expanded communication to respond more actively to customers' changing demands[5]. And it also has another purpose of enhancing job efficiency by applying the analysis results about customers' needs to customer-tailored campaign.

The goal of the NPS's Big Data service is to enhance the value of services and to implement scientific administration. This is a service to realize the NPS' vision of "National Pension for a Successful Life". The NPS is operating the service under a dual organization structure: the Information Protection Department and the Personal Information Protection Department. The Personal Information Protection Department is responsible for establishing and operating of the personal information protection regulations and enforcement rules, and internal management plans. South Korea has strengthened the related laws by having the leakage of personal information subject to criminal penalty. It is different from the United States and most other countries, in that these countries impose civil fines for privacy violations. In South Korea, leaked personal information is often used by criminal organizations located in China and the Philippines for voice phishing and threats, which emerges as a major social issue. For the prevention of cyber infringement and response to it, the NPS set up the Information Protection Department as a separate organization that is responsible for the improvement of security vulnerability, the operation of a security system, and the control service of cyber threat.

Regarding to data disclosure, the NPS is pursuing openness, sharing, communication and cooperation by the South Korean government's 'Government 3.0' policy (the NPS's official website, 2016), and discloses 96 sample items of information and 124 unique items of information. The sample items are a list of the standardized information possessed by public institutions that helps the public to grasp the objects of publication at a glance. The unique items are a list of information which are autonomously disclosed by public institutions to help the public to see various kinds of information possessed by the public institutions.

It is necessary to examine the NPS' data quality level in terms of timeliness in accordance with the Big Data governance framework. The NPS monitors the performance of the service in real time by introducing a system performance management tool, and conducts an improvement work by selecting abnormal transactions[10]. Reliability is not easy to verify, but the dedicated department for data quality management performs a regular verification and improvement work by the company-wide data standards and quality standards. In terms of the significance and sufficiency of data, the NPS' Big Data service already meets the qualifications of meaningfulness and sufficiency, as almost all citizens make contributions to the national pension fund.

The data processes of the NPS's Big Data service consist of four steps: Extract, Transform, Cleansing, and Load. A data flow analysis system is constructed for management. The analysis system of data flow is used to manage the data flow in a structured form that can be traced through a schematized diagram[10]. The enterprise-wide metadata management system performs the procedural management for data standards and integrated data structure management and conducts the management of database alignment in terms

of structure consistency. The full-scale Big Data service of the NPS is launched in 2016.

VI. CONCLUSION

A system for Big Data services can be constructed with the introduction of Big Data solutions. It is also possible to derive those analysis results that meet the goal of the organization and to provide meaningful information to the target customers. However, if you do not consider enough the strategic factors that need to be taken into consideration in advance from a viewpoint of Big Data governance, such as personal information and privacy violation, which are major issues in Big Data, the risk of problem occurrence increases. If Big Data analysis techniques become widely spread, possible ways to utilize knowledge information will be actively sought after, endless debates about data ownership and big brother will be constantly arising. A project to introduce Big Data solutions requires a bigger investment than the existing IT solution introduction projects. The ripple effects of the expected problems including the risk of personal information leakage, the harms caused by information monopoly, and the importance of decision-making based on analysis results are far greater than those of other solutions.

Big Data requires new standards different from data quality standards established based on the existing structured data, which are timeliness, accuracy, sufficiency and meaningfulness. Data disclosure and accountability issues as well as policy making in consideration of the above two factors are the most overlooked factors in projects, which need to 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 due to a policy failure. A pilot project for the realization of Big Data can be conducted prior to the introduction to evaluate its technical possibility, but it is necessary to assess what kinds of interests are correlated in the processes of data collection and disclosure of analysis results.

In this study, we propose a Big Data governance framework to prevent a policy failure by summarizing policy considerations in addition to technical success possibility. For a successful execution of Big Data, a governance framework can be used as a basis for analyzing the risk factors in advance and preventing the risk of problem occurrence. In this case study, we examined the actual cases from a viewpoint of a Big Data governance framework. In addition, we discussed various possibilities which might occur in those cases that were expected to launch Big Data services. A Big Data governance framework can be used as part of an effort to prevent future problems and makes it possible to analyze the criteria that can be taken into consideration. If goals are established and strategies are formulated by a Big Data governance framework and if the components of the Big Data governance are prepared in the preparation process of a Big Data service, it is possible to

overcome an expected difficult crisis and to achieve the goal of the Big Data service that is intended to be realized. The Big Data in the public sector is an evitable choice for improving the quality of people by one notch. A Big Data governance and a subsequently required framework are important factors that must be considered for the realization of Big Data services.

REFERENCES

- [1] Bizer, C., Boncz, P., Brodie, M. L., Erling, O., "The Meaningful Use of Big Data: Four Perspectives – Four Challenges", SIGMOD Record, Vol. 40, No. 4, pp.56-60, 2012.
- [2] Boyd, D., Crawford, K., "Six Provocations for Big Data", SSRN eLibrary., 2011.
- [3] Edjlali, R. & Friedman, T., "Data quality for big data: Principles remain, but tactics change" Gartner, 2011.
- [4] Gantz, J., David R., "Extracting value from chaos." IDC iVIEW 1142: 9-10, 2011.
- [5] Jung, Y., "Big Data Revolution and Media Policy Issues," KISDI Premium Report, 12(2), 2012.
- [6] Lee, S., "National Pension Service of Korea, preparing the base of Big Data Analysis Consulting and Pilot is On," Digital Daily, 2013.11.23.
- [7] Kim, B., "Smart Era, Data Protection Strategy and the Law II", Korea Academy Information, Kyunggi-do(Korea), 2012.
- [8] Kim, H. Y., "Data Governance for Big Data and Medical Information Service practice", Information System Audit and Control, Vol. 25, pp.18-27, 2013.
- [9] Kim, S., "A Case Study of Implementation Data Governance for Enterprise Architecture," Journal of Information Technology and Architecture, Vol. 8. No. 3, pp.255-265, 2011.
- [10] Kim, S., "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, 2013.
- [11] KIPA, "The importance of data governance and data quality management", SW Industry trend, 2013 Oct.
- [12] Kumar, S., "Data governance: An approach to effective data management." White paper, Satyam Computer Services, Ltd. 2008.
- [13] Lee, J., "Data Big bang, the trend of Big data" Journal of Communications & Radio Spectrum, vol. 47, pp.43-55, 2012.
- [14] Lee, K. Y., Nam, G. H., Sim J., Cho, G., Ryu, W., "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, 2012.
- [15] Manyika, J., Chui, M., Brown, B., Bughin, J., Dobbs, R., Roxburgh, C., & Byers, A. H., Big data: The next frontier for innovation, competition, and productivity, 2011.
- [16] Panian, Z., "Some Practical Experiences in Data Governance", World Academy of Science, Engineering and Technology, 2010.
- [17] Power, D., "A 4-D Approach to Data Governance", Information Management, May, pp.29-30, 2011.
- [18] Smith, M., Szongott, C., Henne, B., Voigt, G. V., "Big Data Privacy Issues in Public Social Media", 2012 6th IEEE International Conference, June, 18-20, 2012.
- [19] Thomas, G., "The DGI data governance framework." The Data Governance Institute, Orlando, FL, 2006.