An Overview of Data Governance

Article · September 2016

1 author:

Zhang Ning
Guizhou University of Finance and Economics
3 PUBLICATIONS   **11** CITATIONS

Some of the authors of this publication are also working on these related projects:

perceived information quality View project

data governance View project

An Overview of Data Governance

Ning Zhang
School of Information Management
Nanjing University, China
Email: zn8124@126.com

Qin jianYuan
School of Information Management
Nanjing University, China
Email: yuanqj@nju.edu.cn

ABSTRACT

Nowadays, data and information affect the decision-making process on the various activities of an organization, as data-intensive decision making is being increasingly adopted by businesses, governments, and other agencies around the world, many leaders are aware of the importance of the data, while these organizations devise various ways to deal with the challenges such data brings, data processing, data quality and governance programs should not be underestimated. The recent proliferation of scholarship on data governance indicates the issue has come to be regarded. Researchers in many different fields have shown great interest in data governance, we first review how the people understand the data, and various definition of data governance, then examine the literature relating to framework, DQM, data lifecycle, privacy and security, compliance of data governance, next we explore the types of application areas of data governance based on the literature. The objective of this article is to provide an overview for data governance that can be used by researchers to focus on important data governance issues, and to demonstrate how the abstract framework established conditions for generalization, and how the understanding of these conditions inspired the new researches and applications.

INTRODUCTION

Today, there are kinds of data exist among organizations, for example, with the emergence of social media, which can generate online customer behavior data, such as records of the paths taken by users through a company website, call detail and data-usage records (also known as event detail records) from telecommunications corporations, and data files or transaction records or audio files captured by contact centers (e.g., devices that control client contacts through telephone calls), and Sensors and other facilities continuously and automatically produce data and event streams. Biometric data (such as fingerprints or retina scans) as well as medical images and health care data, while traditional mission-critical applications continue to generate terabytes of transaction data.

In a 2006 survey of 359 North American organizations that had deployed business intelligence and analytic systems, a program for the governance of data was reported to be one of the five success "practices" for deriving business value from data assets (Khatri, V. & C.V. Brown 2010). In the last century, still some organizations do not know what data

they have, how critical that data is, the sources that exist for critical data, or the degree of redundancy of their data assets (Levitin, A.V. & Redman 1998), however, data governance is an important topic for any organization that acknowledges the importance of their business data as a foundation to their success in recent years. It is an area of corporate management that looks at decision-making and authority for data-related matters.

UNDERSTANDING OF THE DATA

At present, more and more organizations realize the importance of the data, they believe that data can improve performance ,create value, enhance competitiveness as well as cut cost. So many surveys conducted by the institution or the conferences conveyed the idea.

The former investigating practice reveals that many organizations, if not all, lack an all-encompassing information governance policy (Economist Intelligence Unit, 2008), especially for external and free format information, and often the policies and processes they do have are not effective, however, they also desire to rely on the data to get the advantage.

Recently, nearly two-thirds of respondents in the 2012 Big Data study confirm that the use and the analysis of information (including Big Data) is creating a competitive advantage for their organizations (M. Schroeck, et al 2012). Another recently survey conducted by the International Association for Information and Data Quality (IAIDQ) (E. Pierce, et al 2012) also showed that more than 70% respondents considered data a strategic asset. According to the 2012 IBM CEO study, the ability of an organization to obtain value from data is strongly related with performance, where outperforming organizations are twice as good as underperformers at accessing and drawing insights from data.

Clearly, with the rapidly growth of digitized data inside and outside the organization, and with the increase of possibilities to access this data, organizations have become aware of the need for right use of their data, there is a great emphasis on data and deriving its value through effective governance.

THE DEFINITION OF DATA GOVERNANCE

Data governance (DG) refers to the overall management of the availability, usability, integrity, and security of the data used in an organization. Since the initial emergence of data governance as an important and fundamental issue to organizations, the data governance community and researchers have published several definitions of the data governance. Although the definition of data governance is still evolving, current usage describes this discipline as being a facilitator for managers to take control over all aspects of their data resource.

In the first place, some researchers argue that the data governance is different from information governance, Kooper (Kooper, et al 2011) stress that data governance and information governance are essentially different, he elaborated information governance is the set of activities aimed at establishing a normative foundation to facilitate and stimulate sense making interactions, which means that data governance focuses on data assets, information governance is related to interactions. However, other scholars may consider they are the same thing, M. Godinez (M. Godinez, et al 2010) considered the information governance is a discipline that governs company data assets throughout their life cycles.

There are many definitions, first steps have been taken to define such policies and processes from a compliance perspective (Donaldson & Walker, 2004; Kahn & Blair, 2004).Hereafter, some people stress data governance refers to who holds the decision rights and is held accountable for an organization's decision-making, for example, Thomas (Thomas

2006) set forth data governance is a system of decision rights and accountabilities, executed according to agreed-upon models which describe who can take what actions with what information, and when, under what circumstances, using what methods. Budi Laksono Putro (Budi Laksono Putro et al 2015) considered data governance is a decision-making and authority for matters relating to the data. Weber, K (Weber, K et al 2009) said that data governance specifies the framework for decision rights and accountabilities to encourage desirable behavior in the use of data, to promote desirable behavior, data governance develops and implements corporate-wide data policies, guidelines, and standards that are consistent with the organization's object, strategy, values, and culture.

Other scholars pay attention to policies, processes, technologies and responsibilities for the control and management of data. Fernandes said that data governance represents the enterprise policies or strategies that define the purpose for collecting data and governing the ownership and intended use of data (Fernandes L, O'Connor M 2009). Rosenbaum also agreed that data governance is the process by which responsibilities of stewardship are conceptualized and carried out, data governance establishes the broad policies for access, management, and permissible uses of data (Rosenbaum, S 2010).

In addition, some researchers hold the opinion that data governance is closely related with data assets, which means the great importance of digital data in the organization. Some scholars (e.g. Ladley 2012, M. Godinez et al 2010, Abrar Haider 2012, Aiken 2016) treated data governance is the way the company or institution manages its data assets, Only by reconceiving data as a strategic asset can organizations begin to address the new challenges.

From the above, we can find data governance contains many aspects, which formed from a convergence of procedures, technologies, processes, policies, responsibilities and decision making rights for the use of data in organizations.

THE RESEARCH OF DATA GOVERNANCE

Developments during the year in review strongly suggest that good data governance has become a necessary element of a well-managed company (Powe 2005).
The framework of DG

The framework or model plays a major role of implementing data management, is the important reference when performing data governance, a recently research group at the University of St. Gallen tested the various data governance frameworks in three projects with partners from different industries and requirements, the results demonstrated that the model is considered a very useful tool for organizing, communicating and coordinating accountabilities for DQM (Data Quality Management) in an organization-wide context.

Firstly, almost all published frameworks recommend that a data governance should undertake the maturity assessment, to establish the current state of data management and control, such as DataFlux, EWSolutions, Gartner, IBM, MDM Institute, ARMA International, and CMMI Institute's Data Management Maturity Model (Neera Bhansali 2013), however, here studies focused on the maturity of organizational data management practices are rare (Aiken 2016). A maturity model for data governance is a critical first step (S. Soares, T. Deutsch, S. Hanna & P. Malik 2010). Some frameworks have supporting (online) tools which can facilitate this process by directing an enterprise to assess all key aspects of data management. For example, Supplier Kalido (Data Governance Institute 2011) offers consultancy services and data management software, encouraging customers to first do a self-assessment of their data governance maturity with online tool, the criteria that cover three potential aspects (organization, process and technology) of data governance implementation,

based on the results of the assessment, an enterprise then fall into one of four data governance maturity stages: Application-Centric, Enterprise Repository-Centric, Policy-Centric, and Fully Governed. As the early step in the implementation of data governance, the purpose is to help the organizations assess their current state and realize the gaps, to make right efforts to reach the future state. (P, Malik 2013). Several common maturity assess models are given by the famous institutions, such as IBM's Information Governance Council Maturity Model (IBM 2007) and so on, based on these models, organization also add the other factors to formulate their own maturity assess model.

In terms of the different organizational structure, culture, politics and policies, researchers intend to use a common framework or approach. DGI (Data Governance Institute 2011) provided a single framework that is able to deal with data governance programs that focus on one or more data-related areas, no matter what the areas are, the framework proposes to establish to share the same universal objectives from enabling better decision making to ensuring transparency of process. Learned from the previous published IT governance framework by Weill & Ross (Weill & Ross 2004), Khatri & Brown (Khatri & Brown 2010) propose a data governance framework using five interrelated data decision domains: data principles, data quality, metadata, data access and data lifecycle, this framework become the important reference for the latter research. Both academic and practical sources provide data governance as a universal approach – one that fits all enterprises alike, however, Weber (Weber et al 2009) states that there should not one approach to data governance because of the distribution of accountabilities, so he proposes a flexible data governance model made up of several roles (executive sponsor, chief steward, business data steward, and technical data steward), decision areas (or tasks) and assignment of responsibilities, which stress data quality roles and their interaction with DQM (Data Quality Management) activities. The roles of the model are drawn together into a responsibility assignment matrix.

Data quality management

In the above framework, the data quality is an important element. Data quality management (DQM) focuses on the collection, organization, storage, processing, and presentation of high-quality data (Wende 2007). Data Quality Management specifies planning and implementation of techniques and processes to cleanse and purify existing data, and to ensure that whatever data is entered in the organizational information systems conforms to quality specifications (Abrar Haider 2012). In the early research, only role accountable in DQM, such as information product manager (Wang 1998), currently DQM approaches deal with more than one role, which related to several organizational levels, and their tasks and responsibilities.

Data quality is seen as the most important aspect influencing usability of data for business processes and reporting (Friedman and Smith 2011). Data quality in turn largely determines the effectiveness of the business processes, and also influences the reporting quality (Eppler and Helfert 2004). Due to the organizational compliance reporting, data quality becomes even more crucial (Nick Martijn, Joris Hulstijn & Mark De 2015), if low data quality is pervasive, costly and can cause high inefficiencies (Eppler and Helfert 2004; Fisher & Kingma 2001; Wang et al. 2001), if data is not kept 'clean' and the overall data quality is reduced, for example, in North America over $600 billion in lost revenue has been attributed to the data quality among companies (Friedman, Logan, & Newman 2008). Therefore, data of high quality are a prerequisite for fulfilling these changing business requirements and for achieving enterprise agility objectives (Newman & Logan 2006). Data governance is not just about improving quality of data alone. It is about managing the asset of

the organization so as to enable a continuous improvement-based learning progression (Abrar Haider 2012).

Data Life Cycle

Scholars consider that there should be different phrase of data governance because of various stages in organization. From this perspective, in the early stage of organization, data governance is not a competitive tool, so many organizations do not know the essential of the data, or the importance of data assets, the organization in the early phrase may not has the resource and desire to develop an effective data governance. (Neera Bhansali 2013).

By understanding how data is used, and how long it must be retained, organizations can develop ways to make usage patterns fit to the optimal storage media, thereby minimizing the total cost of storing data over its life cycle (Khatri & Brown 2010). By placing data according to the life cycle, complying with business needs, data can be used more effectively distributed across multiple resources, thus leading to improved storage utilization and reduced storage acquisition costs.

At the International Digital Curation Conference in 2010, during the university of Oxford long way in developing research data management (RDM) policy, they addressed the various phases of the research data lifecycle, data lifecycle were envisaged as discrete phases formerly, now regard it as more helpful to think in terms of a continuum, which can help data management interventions more appropriately to the way of working (James& Pau 2013).

Metadata Architecture

Metadata describes what the data is about and provides a mechanism for a concise and consistent description of the representation of data, it is needed to be addressed in data governance, in order to assure the data is interpretable, standardizing metadata provides the ability to effectively use and track information, helping interpret the meaning or "semantics" of data. In Khatri's opinion, as the environment for a business changes, the way an organization conducts business and consequently the associated data also changes. As such, there is a need to manage changes in metadata as well (Khatri & Brown 2010).

Different kinds of metadata play different role in the discovery, retrieval, collection ,management and analysis of data Singh et al defined metadata type are physical, domain-independent, domain-specific, and user metadata (Singh et al 2003), whereas Neera Bhansali classified metadata as business type, technical type, process type, and operation type , and do the case study of metadata use for data governance.(Neera Bhansali 2013). Several data governance framework have mentioned the metadata (Khatri & Brown 2010, Aiken 2016, Andi Ogier et al 2014, Abrar Haider 2012, Carolyn Begg 2012, Marco & Smith 2006, James& Pau 2013).

Security and privacy of data

security and privacy of data, relates to the protection of data stored via computer, server, or any other form of electronic media. In the most basic terms, we expect data will always be kept confidential, have integrity or accuracy, and be available when needed to provide them with service. The increase organizational heightened concern about privacy have created the need for enterprises to develop formal data governance strategies and programs. Example as the financial sector, compliance, data security, and trust protection, are all elements critically important to the sector.

The data Quality is moreover depends on the data privacy; it will be good in quality by how it is in privacy. The data quality mean that how our product is accurate & working relevant to the necessary purpose that is indicated . Data governance is an ongoing process of

monitoring, evaluating, and assessing data, its users, and database activity to better understand and control data risk (Neera Bhansali 2013), in addition to compliance with legal requirements of information security, most of the data governance also propose the security requirements, related to how personally identifiable information is secured and protected, and establish the role of information or data security officer to guarantee the secure activities such as data access, Powe advise that an organization should raise the standards reflected in its security policies (Powe 2005). Budi refer to data governance factors, mentioned data security is one of the factors (Budi et al 2015), Wende stressed that organization should establish policies and procedures to enforce security management (Wende K. 2007). Carolyn explore the awareness and attitude to many factors (involve security factor) of data governance within each enterprise (Carolyn Begg 2012).

Compliance:

Regulatory compliance, in terms of institutional compliance, is the process to ensure that laws and regulations that govern how business is conducted. Companies have to comply to certain laws, such as Sarbanes-Oxley for companies listed on the US stock exchange, or Solvency II for insurance companies (Eling et al. 2007), or the HIPAA (Health Insurance Portability and Accountability Act) for health insurance. These laws demand that companies demonstrate to the regulator that they are compliant, which requires evidence. For example, HIPAA requires an organization to maintain, publish, and educate its personnel on the primary components of the HIPAA regulation through the implementation of policies and procedures related to protecting a patient's protected health information. Regulatory compliance creates additional data requirements. It is not sufficient to supply evidence; an audit trail is also required (Jiang and Cao 2011). As researchers said, not meeting data requirements can lead to severe financial consequences.

Governance programs allow for the compliance to regulatory requirements. There exists no formalized process for proving for regulatory compliance with legal laws. Data governance initiatives may be aimed at achieving a number of objectives including offering better visibility to internal and external customers and compliance with regulatory laws.

THE APPLICATION AREAS OF DATA GOVERNANCE

Data Governance have applied in many industries, the most typical is in the enterprise, as the enterprises, they have a variety of data requirements and differing data focus, they all seem to have similar views on developing a data governance strategy as an organizational asset, and are experiencing similar issues related to their ability to successfully implement the management and control of data. Khatri & Brown (Khatri & Brown 2010) not only propose a data governance framework using five interrelated data decision domains, but also tested on a case study for a large insurance company. Nick Martijn developed a framework to make it possible to specify the effects of data governance interventions in enterprise, the cases were collected with the help of improving their data quality through a data governance framework, in particular in the financial, logistics and retail sector (Nick Martijn, Joris Hulstijn & Mark De 2015).

In addition to the large cooperation, some scholars recently pay attention to the application in the SMEs (small enterprises), SMEs often lacked full awareness and understanding of how to implement data lifecycle and governance issues, as well as perception of the value of data inside and outside the organizational boundary. Carolyn Begg explored the awareness of and attitude of data governance within each enterprise and given

the differing data focus between the enterprises, intended to find the potential benefits that could be provided by the implementation of data governance (Carolyn Begg 2012).

With a large number of data generated in the medical industry and the increased availability of accessing to large health datasets, medical data need to manage and control, data quality improvement research have emerged, someone explored data governance in distributed clinical research networks, stressing the privacy protect and provide a trustworthy platform for research (Katherine et al 2014).

We are pleased to find the type of organization is expanding, in order to assess electronic data resources, library used the data asset framework and data governance methodology, Andi Ogier share their experiences using the Data Asset Framework (DAF) methodology as an interview protocol to audit and assess electronic resources data management and associated library reports, discussesing benefits of using data audit protocols as a tool for assessing library data (Andi Ogier et al 2014). Studies on governance data or information stressed the importance of cultural factors in the governance of research data, in order to achieve higher education goals, scholars took Indonesia University as case study, assessing the value of data governance , organizational culture and leadership (Budi Laksono Putro 2015).

CONCLUSION

As organizations begin to realize that data are an important asset, the need for research data governance strategies has become prevalent, ttherefore, it is a challenging task to provide a comprehensive overview of data governance within a short article. We find that data governance are directly related to developing appropriate stakeholders and decision-making managers that are focused on managing privacy and compliance. There are several factors to be considered when performing data governance, such as data quality, metadata management, privacy and security and so on. The primary question that each organization needs to answer is: Who owns the data? Therefore, it is important to identify the responsibility of each sector, now there are various framework developed to guide the management and control of data, and data governance is applied in many areas except for just enterprises, such as university, library, and health sector as well as many small companies.

Based on the above revelation, combined with the development trend of the research, we believe that the current DG (Data Governance) research should from the following to expand.

(1) Broaden the sight of DG research, focusing on the research under the environment of Big Data, exploring the application in the not standardized management of the small company.

(2) Now data governance in China's corporate governance is still rare, actively explore our different groups, organizations, particularly administration agencies, educational institution and other organizations DG problems.

(3) Carry out empirical research, to solve the more practical issues. Focus on expanding and innovation research methods, conducting empirical research, using data governance framework or conceptual model, to verify the results in the industry's application of DG.

REFERENCES

Khatri, V. & C.V. Brown (2010). Designing data governance. Communications of the ACM, 53(1): 148-152.

Levitin, A.V. & Redman. (1998). Data as resource: Properties, implications, and prescriptions. Sloan Management Review, 89-101.

M. Schroeck , R. Shockley, J. Smart, D. Romero-Morales & P. Tufano (2012). Analytics: The real-world use of big data, IBV and Saiid School of Business, University of Oxford. [Online].Available: http://public.dhe.ibm.com/common/ssi/ecm/en/gbe03519usen/GBE03519USEN .PDF

E. Pierce, C. L. Yonke, P. Malik, & C. K. Nagaraj B (2012). The state of information and data quality industry report 2012,[IAIDQ, Baltimore, MD. [Online]. Available: http://iaidq.org/publications/

Economist Intelligence Unit. (2008). The future of enterprise information governance. London: The Economist Intelligence Unit Limited.

IBM Global CEO Study (2012). IBM Corp., New York, [Online]. Available: http://ibm.com/services/us/en/c-suite/ceostudy2012.

Kooper, M.N., R. Maes & E.E.O.R. Lindgreen (2011). On the governance of information: Introducing a new concept of governance to support the management of information. International Journal of Information Management. 31(3), 195-200.

M. Godinez, E. Hechler, K. Koenig, S. Lockwood, M. Oberhofer, & M. Schroeck, (2010). The Art of Enterprise Information Architecture: A Systems-Based Approach for Unlocking Business Insight. Boston, MA: IBM Press.

Donaldson, A., & Walker, P. (2004). Information governance—A view from the NHS. International Journal of Medical Informatics, 73, 281–284.

Thomas, G (2006). Alpha Males and Data Disasters: the Case for Data Governance. Brass Cannon Press

Budi Laksono Putro, Kridanto Surendro &Herbert. (2015). Leadership and culture of data governance for the achievement of higher education goals (Case study: Indonesia University of Education). Proceedings of International Seminar on Mathematics, Science, and Computer Science Education.

Fernandes L, O'Connor M (2009). Data governance and data stewardship. Critical issues in the move toward EHRs and HIE. J Am Health Info Manage Assoc 80(5):36–39

Rosenbaum, S. (2010). Data Governance and Stewardship: Designing Data Stewardship Entities and Advancing Data Access. Health Services Research, 45(5p2), 1442-1455.

Ladley J. (2012). Data Governance: How to Design, Deploy, and Sustain an Effective Data Governance Program. Newnes.

Powe, R.L.T.A. (2005). Lessons in Data Governance: A Survey of Legal Developments in Data Management, Privacy and Security. Business Lawyer, 61.

S. Soares, T. Deutsch, S. Hanna & P. Malik (2010). BBig data governance: A framework to assess maturity, IBM Corp., New York. [Online]. Available:

http://ibmdatamag.com/2012/04/big-data-governance-a-framework-to-assess-maturity/

P, Malik (2013). Governing Big Data: Principles and practices. IBM Journal of Research and Development, 57(3/4).

IBM (2007). IBM Data Governance Council Maturity Model: Building a road map for effective data governance. [Online]. Available: http://www.infogovcommunity.com/blog/maturity-model-white-paper/

Data Governance Institute Framework (2011) [Online]. Available: http://www.datagovernance.com

Weill, P. & Ross J. (2004). IT governance: How top performers manage IT decision rights for superior results. Harvard Business School Press, Boston.

Khatri, V. & Brown, C.V.(2010). Designing data governance. Communications of the ACM, 53(1), 148–152.

Weber, K., Otto, B. & Oesterle, H. (2009). One Size Does Not Fit All—A Contingency Approach to Data Governance. Journal of Data and Information Quality (JDIQ), 1(1), 4-8.

Carolyn Begg (2012). Exploring the SME Quandary: Data Governance in Practice in the Small to Medium-Sized Enterprise Sector. Electronic Journal Information Systems Evaluation 15(1).

Nick Martijn, Joris Hulstijn Mark De. (2015), Determining the Effects of Data Governance on the Performance and Compliance of Enterprises in the Logistics and Retail Sector.

Katherine K Kim, Dennis K Browe, Holly C Logan, Roberta Holm, Lori Hack, & Lucila Ohno-Machado (2014). Data governance requirements for distributed clinical research networks: triangulating perspectives of diverse stakeholders. Journal of the American Medical Informatics Association, 21(4), 714-719.

Andi Ogier, Monena Hall, Annette Bailey, & Connie Stovall. (2014). Data Management Inside the Library: Assessing Electronic Resources Data Using the Data Asset Framework Methodology. Journal of Electronic Resources Librarianship, 101-113.

Friedman T., Smith, M. (2011). Measuring the Business Value of Data Quality. Gartner, Stamford.

Eppler, M., Helfert, M. (2004). A classification and analysis of data quality costs. In: Paper presented at the International Conference on Information Quality.

Fisher, C.W. & Kingma (2011). B.R.: Criticality of data quality as exemplified in two disasters. Inf. Manage. 39(2), 109–116.

Wang, R.Y., Mostapha, Z., Yang, W.L. (2001). Data Quality. Kluwer Academic Publishers, Dordrecht.

Friedman T., Logan D., Newman D. (2008). Governance in an essential building block for enterprise information management. Gartner Research, USA.

Wende, K. (2007), A Model for Data Governance – Organizing Accountabilities for Data Quality Management, in 18th Australasian Conference on Information Systems, Toowoomba.

Newman, D. & Logan, D. (2006). Governance Is an Essential Building Block for Enterprise Information Management, Gartner Research, Stamford.

Abrar Haider (2012). Asset Lifecycle Data Governance Framework. Proceedings of the 7th world Congress on Engineering Asset Management, 287-296.

Neera Bhansali (2013). Data Governance: Creating Value from Information Assets. CRC Press.

Wang, R.Y. (1998). A Product Perspective on Total Data Quality Management. Communications of the ACM, 41(2), 58-65.

Aiken, P. (2016). Experience：Succeeding at Data Management—BigCo Attempts to Leverage Data. Journal of Data and Information Quality, 7(1-2), 1-35.

James A.J. Wilson & Paul W. Jeffreys (2013). Towards a Unified University Infrastructure: The Data Management Roll-Out at the University of Oxford. The International Journal of Digital Curation, 2(8), 235-287.

Singh, G., Bharathi, S., Chervenak, A., Deelman, E.,Kesselman, C., Manohar, M., Patil, S., & Pearlman, L. A metadata catalog service for data intensive applications. In Proceedings of the ACM/IEEE SC2003 Conference on High Performance Networking and Computing.

Marco, D. & Smith, A.M. (2006). Metadata Management & Enterprise Architecture: Understanding Data Governance and Stewardship, DM Review.

Eling M., Schmeiser H. & Schmit J.T. (2007). The solvency II process: overview and critical analysis. Risk Manage. 10(1), 69–85.

Jiang, K. & Cao, X. (2011). Design and implementation of an audit trail in compliance with US regulations. 8(5), 624–633.

10