Data Stewardship and Flow Management for Data Quality Improvement

Thien-Vu Tran[#1], Sunho Kim [2#], Ming-Hsiung Hsiao[3*]

[#]*Department of Industrial and Management Engineering, Myongji University*
135 Seosomun-Dong, Jung-Gu, Seoul, Korea
[1]`vutrandn@gmail.com`

[#]*Department of Information Management, Shu-Te University*
59 HengShan Rd., Yanchao Dict., Kaohsiung, Taiwan
[3]`msshiaw@stu.edu.tw`

Abstract— **In the information age, data is a vital asset for enterprises and quality of data impacts on enterprises' wealthy and future success. Therefore, they must take a great effort to manage and improve quality of data for achieving competitive advantage. In general, there are various methods to enhance quality of data; however, in this paper, we propose a framework to improve quality of data through data stewardship and flow management. Indeed, data quality is improved not only by managing, controlling, adjusting, and maintaining enterprise-wide data flow but also by organizing data stewards in an effective way to assign right roles to right decision areas with right accountability. The proposed framework is verified in a case study of Korea Credit Bureau.**

Keywords— **data quality, data quality management, data stewardship, data flow.**

1. INTRODUCTION

In the information age, data is a vital asset for enterprises. To conduct business, enterprises take an effort to collect, apply, and exploit data every day. Based on data captured, they keep records, produce reports, deliver information, monitor performance, make decisions, and much more. Therefore, data play an important role in most enterprises and contribute to enterprises' wealthy and future success. However, due to the rapid development of information technology, data volume is exponentially increasing and a variety of data is dramatically generated. Hence, data gradually become unreliable and polluted resources for enterprises. Consequently, managing and improving quality of data is a critical issue for any enterprise in today's business environment.

Poor quality of data may lead to enormous cost. According to The Data Warehousing Institute (TDWI) (2002), data quality problems cost U.S. business more than $600 billion a year [4]. Poor quality of data causes enterprises to cover much more expense to organize and manage data such as data diagnosis, data cleaning, and data improvement. Furthermore, poor quality of data also makes an effective decision frustrated. In fact, quality of data rapidly becomes worse over time. Experts say 2% of records in a customer file become obsolete in one month because customers die, divorce, marry, and move. In addition, data entry errors, system migrations, and changes of source systems generate bucket loads of error [4]. Based on such volatile data, data manager are hard to make effective decisions. Finally poor quality of data may result in both tangible and intangible damage of consumers' confidence [6]. Poor quality of data like inaccurate or inadequate data may cause incredible information for both enterprises and customers that leads to lose confidence of each other.

Consequently, managing and improving data quality is an essential issue. Data quality initiatives ensure that enterprises would reduce cost, make decision effectively, and raise customers' reliability. These are key factors leading to survive and success in an intense competitive environment today.

This paper proposes a framework of data stewardship and flow management that combines data stewardship management and data flow management that ever presents before. This paper is categorized into 6 sections. The section 1 introduces the data quality issues and research objective. Then, the section 2 reviews data quality methodology, models and framework. The background of data quality, data stewardship and data flow is presented in the section 3. Next, the section 4 proposes a framework of data stewardship and flow management. Then, a case study of Korea Credit Bureau is illustrated in the section 5. The final section describes conclusion.

2. LITERATURE REVIEW

2.1. Data Quality Management Approach

In general, there are a number of concepts and approaches in the domain of data quality management, in which Total Data Quality Management (TDQM), Total Quality data Management (TQdM), Total Information Quality Management (TIQM), and the framework for information quality management are the mainstream research of data quality management today.

The first approach proposed by a data quality group at the Massachusetts Institute of Technology (MIT) in 1991 was Total Data Quality Management (TDQM) [1, 19, 20]. In this approach, managing data quality is relied on the philosophy of quality management named Total Quality Management (TQM). Data is improved through the TDQM cycle, which outlines the definition, measurement, analysis and improvement of information quality. Then, Wang et al. (1998) proposed adoption of the Information Product (IP) approach and provided a framework. The key content of the IP approach is to manage information as a product, model many information systems as information manufacturing systems [3].Based on the IP approach, Shankaranarayan (2003) presented a framework for managing data quality in a dynamic decision environment preferred to an information product map or IPMAP [16], and next, Scannnapieco (2005) proposed a methodology for quality improvement based on the IPMAP framework and UML (Unified Modeling Language), named IP-UML [15].

The second approach is Total Quality data Management (TQdM) [5] which is made up of five processes for measuring and improving data quality. The goal of TQdM is to improve business performance and customer satisfaction through information quality improvements.

The third approach is a framework for Information Quality Management (IQM) composing a matrix with four views and four phrases in the information life cycle proposed by Eppler (2006) [6]. It is argued that quality of information can be improved by knowledge-intensive business processes such as on-line communication, strategy, product development, or consulting.

Finally, data quality for the information age developed by Redman includes a data quality program to detect error, control process, and improve data quality, design process and policy [13].

2.2. Models and Frameworks for Data Quality Management (DQM)

1. Data Governance Model

Previously, data quality issues are mostly assigned to department of IT in companies without attention to solve these problems in term of business-driven perspectives. Furthermore, companies often neglect to the organizational issues that are essential to the success of DQM. A Data Governance Model presented by Wende (2007) enabled company to implement corporate wide accountabilities for DQM that encompass processionals from business and IT [21, 22]. The Data Governance Model is comprised of three components that are data quality roles, decision areas, and responsibilities.

These three components are built up a matrix which each column is data quality roles and each row is decision areas and main activities. Next, each cell of the matrix is assigned with Responsible, Accountable, Consulted, and Informed that is an acronym of RACI [21]. Responsible is assigned to those who do the work to achieve the tasks. Accountable person is the individual who is ultimately answerable for activity or decision. Consulted person is those whose opinions are sought and with whom there is two-way communication. Finally, informed person is individual who are kept up-to-date progress and with whom there is just one-way communication [17].

2. Data Quality Management Maturity Model

Ryu et al. (2006) introduced a Data Quality Management Maturity Model to improve data structure quality that is a main root of both poor data value quality and poor data service quality. In this model, four maturity levels of initial, defined, managed, and optimized. At the level 1, the initial data management stage, data structure quality is managed through rules defined in the database system catalogue. At the level 2, defined data management stage is the data management stage through the logical data model and physical data model. At the level 3, managed data management is data management through data standardization. At the level 4, optimized data management is data management through data architecture management [14].

3. IBM Data Governance Council Maturity Model

The IBM (2007) developed a Data Governance Maturity Model based on the Capability Maturity Model (CMM) methodology which was used to develop and refine an organization's software development process. The CMM is composed of five maturity level from low level to high level that are initial, managed, defined, quantitatively managed, and optimizing progression. The Data Governance Maturity Model describes 11 essential elements of effective Data Governance and each element is developed through five maturity level [8].

4. Corporate Data Quality Framework

Otto (2007) presented a framework for Data Quality Management for the alignment technical aspects of corporate data quality with business-related issues. The framework is made up of the three layers of business engineering including strategy, organization and information systems and viewed by the two perspectives of data management including governance and execution [12].

5. CIHI's Data Quality Framework

Managing data and information in health-care information system is such an extremely important task that the Canadian Institute for Health Information (CIHI) (2005) built a framework providing an object approach to apply consistent data flow processes, assess the data quality of a data holding, and then produce standard data-holding documentation with continuous improvement [2]. The framework composes of three main components including a data quality work cycle, a data quality assessment tool, and document about data quality.

6. Federal DAS Data Quality Framework

The Federal Data Architecture Subcommittee (DAS) Data Quality Framework describes a series of disciplines and procedures to ensure that data meet ends of the quality characteristics required for usage. The DAS Data Quality Framework defines approaches for people, processes, and technology that are based on proven methods, industry, standards, and past achievements [18]. It also establishes guidelines for Federal Agencies to maintain and increase the quality, objectivity, utility, and integrity of information. Data quality principles and initiatives are analysed and scrutinized inside five Federal Enterprise Architecture (FEA) reference models that are Performance Reference Model (PRM), Business Reference Model (BRM), Service Component Reference Model (SRM), Data Reference Model (DRM), and Technical Reference Model (TRM). As a result, the thirteen Data Quality Initiatives process steps are built and used in best practices that enable a number of federal agencies to implement this framework successfully.

7. Process-centric Data Quality Management Framework

Kim et al. (2011) propose a framework of data quality management based on process. The framework illustrates three top-level processes including data quality monitoring, data quality improvement, and data operations and each level of these processes is segmented by three roles of person performing the processes including data manager, data administrator, and data technician. A 3x3 process matrix, consequently, is built up and describes 9 data quality processes. Inside of the data operations process, data architecture management, data design, data processing processes are clarified. Similarly, data quality monitoring process consists of data quality planning, data quality criteria setup, and data quality measurement processes. Finally, data quality improvement is made up of data stewardship and flow management, data error cause analysis, and data error correction [9]. In each process, necessity, activities, responsibilities are identified that corresponds to role and relationships among them.

3. DATA QUALITY, DATA STEWARDSHIP, DATA FLOW

3.1. Data Quality

Definitions of quality vary from organizations to gurus. The International Organization Standard (ISO) defines quality as degree to which a set of inherent characteristics fulfil requirements. According to Genichi Taguchi, quality is closeness to target and deviations means loss to society. However, J. Juran and American Society of Quality Control referred quality as “fitness for use” and fitness is defined by customers. In the thesis context, we study quality of data based on the definition of “fitness for use” and fitness is customers’ satisfaction for their purposes. So, data quality refers to fitness of data for use.

Management is the process of getting activities completed efficiently and effectively with and through other people. Management includes planning, organizing, staffing, directing, coordinating, reporting, budgeting, controlling activities. In summary, data quality management is the act of planning, implementing, and controlling the quality of data that apply quality management techniques to measure, assess, improve, and ensure the fitness of data for use.

3.2. Data Stewardship

Data stewardship plays a critical role in data quality improvement. According to the DAMA, data stewardship is the formal accountability for business responsibilities ensuring effective control and use of data assets [11]. The heart of data stewardship is to define and maintain business meta-data across organization in order to ensure that business meta-data is created and maintained effectively and it is used to make data stewardship and governance decisions.

Data stewards are often business leaders and/ or recognized subject matter experts who are classified into a hierarchical team including executive sponsor, chief steward, and business and technical data stewards. They are designated as accountable for business responsibilities in order to ensure quality of enterprise data assets.

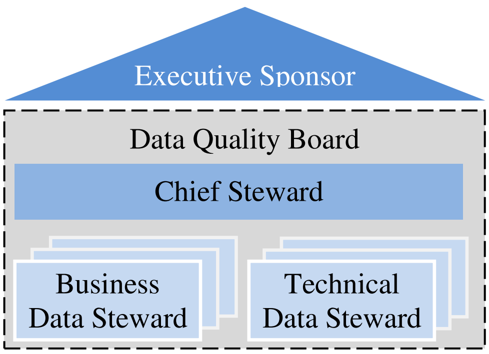

Fig. 1 Data stewardship management

3.3 Data Flow

As establishing an information system, we are confronted with challenges not only to build individual entities but also to connect these entities together and transfer data from entity to other entity. Consequently, a concept of data flow was set up. So far, there are many ways to present dataflow in both academic and practice such as Data Flow Diagram, Integration Definition for Function Modelling. These approaches are made in details following.

The Data Flow Diagram (DFD) is a graphical representation of the flow of data through an information system [3]. It is a visual tool to depict logic models and expresses data transformation in a system. In detail, it illustrates data items from an external data source or an internal data store to an internal data store or an external data sink through an internal process. Due to giving an entire system’s data flow and processing with a single process, a DFD is called a context diagram. DFD is a major part of the structured design and analysis method widely used in industries. It help users to visualize how the system operate, what the system accomplish, and how the system be implemented. However, a DFD does not mention the timing of processes, or whether processes will operate in sequence or in parallel.

Integration Definition for Function Modelling (IDEF0) is a common modelling technique for the analysis, development, re-engineering, and integration of information systems, business processes, or software engineering analysis [3]. IDEF0 is built on SADT (Structured Analysis and Design Technique), developed by Douglas T. Ross and SofTech Inc. [10]. It includes both a definition of a graphical modelling language (syntax and semantics) and a methodology for developing models [7]. IDEF0 describes a business process as a series of linked activities in which each activity is specified by four elements: inputs, outputs, controls, and mechanisms

Different from other modelling technique such as Block Diagrams or Flowcharts, IDEF0 provides a formal method of describing processes or systems used to show data flow, system control and the functional flow of life cycle processes. IDEF0 is capable of graphically representing a wide variety of business, manufacturing and other types of enterprise operations to nay level of detail. It provides rigorous and precise description, and promotes consistency of usage and interpretation [3]. IDEF0 provides a system engineering approach to support systems analysis and design at all levels, communicate among analysts, designers, users, managers and so forth, provide reference scheme for enterprise analysis, information engineering and resource management.

4. DATA STEWARDSHIP AND FLOW FRAMEWORK

In order to improve the quality of data, it is very important to identify the data flow and their relations in the information system. Data rarely exist in silos, but is always integrated and utilized in inter-organizational ways. Linking data among applications is actually necessary to identify roots of errors in the information system and predict the unexpected results due to the errors in any stage. For this reason, modeling the data flow plays an essential role in the information system.

Improving the quality of data is not only related to technical tasks such as modeling and controlling the data flow but also related to organizational tasks. As a matter of fact, assigning people who are responsible and accountable for their works in right decision area is also critical for data quality initiatives. Because IT staff hardly understands business rules while business staff often lacks of IT skills and IT knowledge. So it is important to set up a group contributing to data quality initiatives that combines sponsorship from the leaders like CEO, CIO and the contribution from business and IT staff.

Some thesis proposed the data stewardship framework how to manage data stewards in organizations efficiently while some thesis presents data flow management from point of view of technical. However, data flow is managed under the control of data stewards and a framework of combining the data stewardship and flow has never proposed before. Hence, in this thesis, we propose a framework of managing data stewardship and flow for data quality management. In this framework, data flow is managed in accordance with the horizontal side management whereas data stewardship is managed through vertical side management in any enterprise.

Since the entrance of data from outside, data is operated through several stages. Next, data is supplied to customers. At each stage, data is processed by business data stewards and technical data stewards at the operational level. Both business data stewards and technical data stewards belong to a data quality board which is managed by a chief data steward at the tactical level. Finally, an executive sponsor supports and controls all of data quality initiatives in organizations.

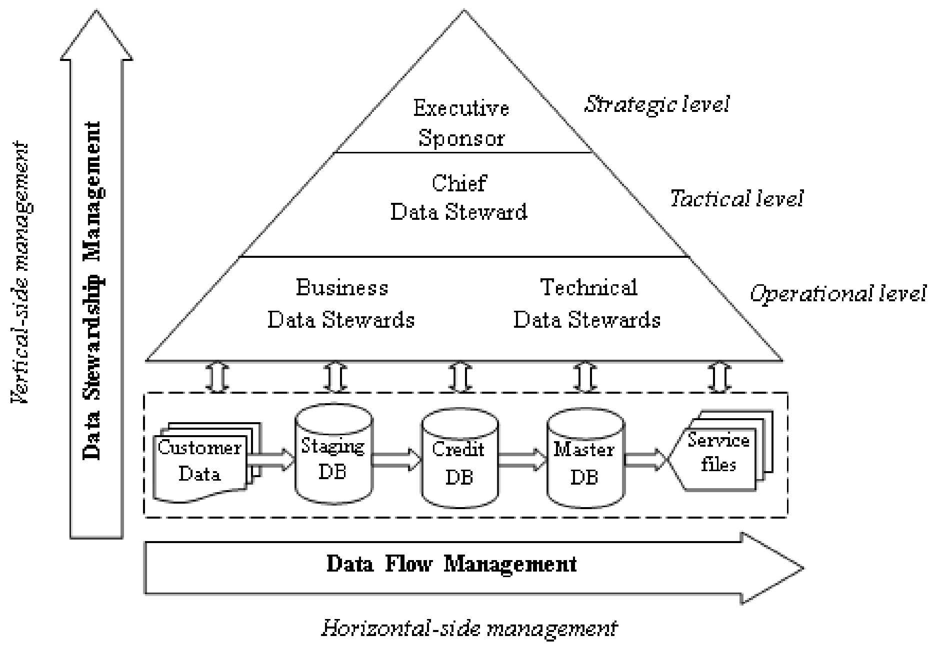

Fig. 2 Data stewardship and flow framework

5. CASE STUDY: KOREA CREDIT BUREAU

Korea Credit Bureau (KCB) is the service-providing company which was established through the supports of some Korea financial institutions including bank, card company and insurance company. The roles of KCB are to evaluate scientific and objective risk credit management and offer tailored customer management and value-added services. The KCB provides information and overall solutions necessary for all customer management from identifying valuable customers, credit screening, customer loyalty, preventing attrition, and recovering debt from customers with delayed payments.

The KCB's customers become prominent membership companies (MC) and supply their customer data daily and monthly. Then, the KCB has responsibilities for collecting data from various sources and providing them with service to evaluate credit scores and create credit profiles. The input of the KCB is mostly financial data like loan, credit card data daily transferred and open, use, payment of accounts monthly transferred. Once data are processed and evaluated by the KCB, the outputs of these processes are credit score and credit profile that are supplied to its member companies for its differential purposes. These data is continuously flowed in a non-stop cycle.

5.1. Data Flow Management

1. Input

Customer data is data about customer's profile like name, home address, company address, home phone and data about customer's credit history like bill-paying history, the number and type of accounts, late payments, collection of actions, outstanding debts, and the age of customer accounts. Customer data is also data about customer's status of opening or closing his or her account. All of member information are standardized and written in K-format according to KCB's rules before getting into information system of the KCB.

(1) Validate data

When the financial data is being entered into the KCB information system, the quality of these data is verified by KCB's software named integrated Data Quality Management System (iDQMS). The iDQMS checks and validates data elements of data files by rules to identify incorrect data, incomplete data, inconsistency data, missing data, duplicated data, misfielded value, non-standard representation and sent them back to member company. Otherwise, data is evaluated as high-quality, it is flowed into the next stage, encryption.

(2) Encrypt data

To load data which one can identify individuals such as SSN, Name, home address, banking account number, e-mail address, etc, KCB system encrypts them with encryption algorithm using off-the-shelf product. Next, encrypted data is loaded into staging area in chronological order.

(3) Standardize data

Standardizing data is accomplished through several steps. First, some of fields are divided into several parts. Then, each segment is encoded in a specific digital code according to the KCB standardization system. Finally, some of codes are assigned into KCB digital codes.

(4) Credit DB

Credit DB is a large-scale database in the KCB information system. It contains all of customers' data files which are ready for data mining process. Unlike data files in the staging DB, data files in the credit DB are divided into different tables that support making-decision such as evaluating credit profiles, credit scores, and warning overdue loaners for online service. The KCB uses Oracle software to run the credit DB under the control of database administrator.

(5) Master DB

Master DB is a data warehouse containing data with index. The master DB is established for batch service. Different from online service, batch service supplies information products for fixed time so that master DB is a place storing data for query and supplying them automatically to customers daily, weekly, monthly. The KCB uses the Nettiza software to manage information products in this stage.

2. Output

Report is generated for individual customers. This process serves individual customers online according to their requirement. First, he or she log-ins via his or her accounts in the kcb4u website and requests his or her information. Next, his or her data is processed within two or three minutes and, then, reported online.

Credit profile, credit score and early warning are generated for member's companies. Knowledge workers query data from the master DB by using Sequence Query Language (SQL) to create the final information products. Through querying, they evaluate member company's customer credit score and credit profile. The additional batch service is to produce early warning by analysing the customer profile and offer member company an early warning about their customers.

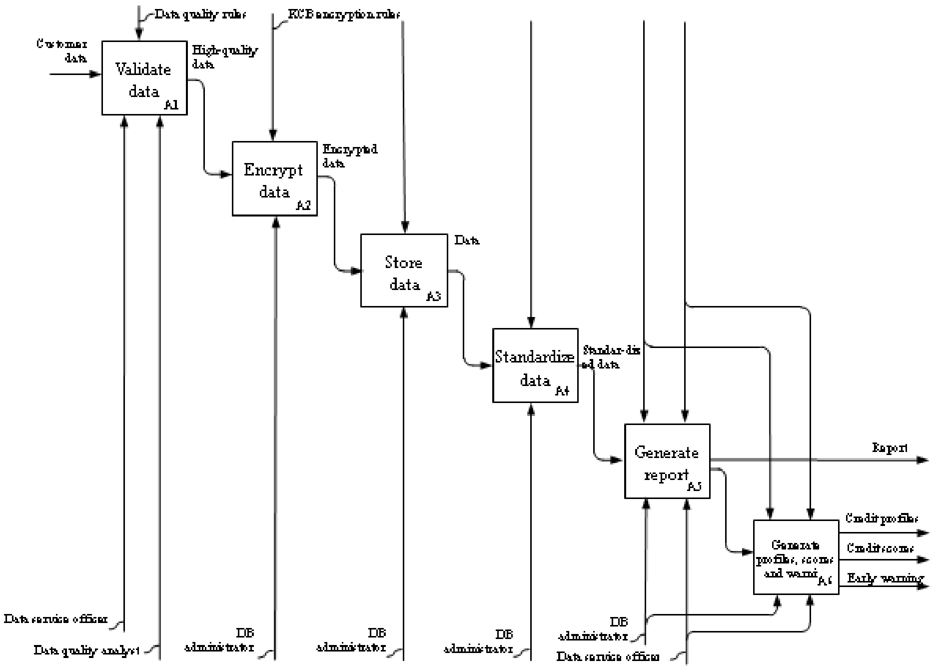

Fig. 3 KCB data flow

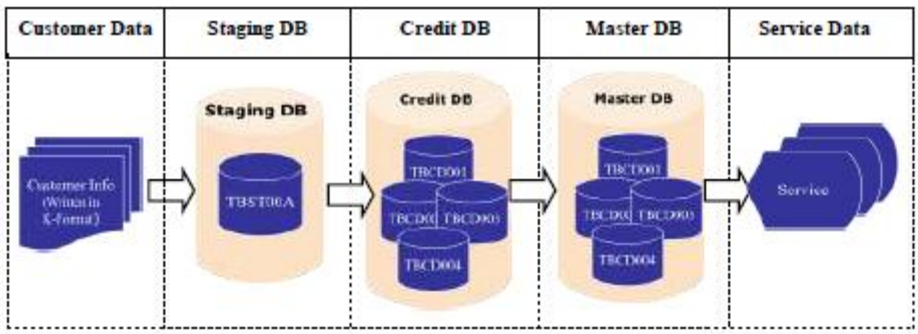

Fig. 4 Data mapping in the KCB information system

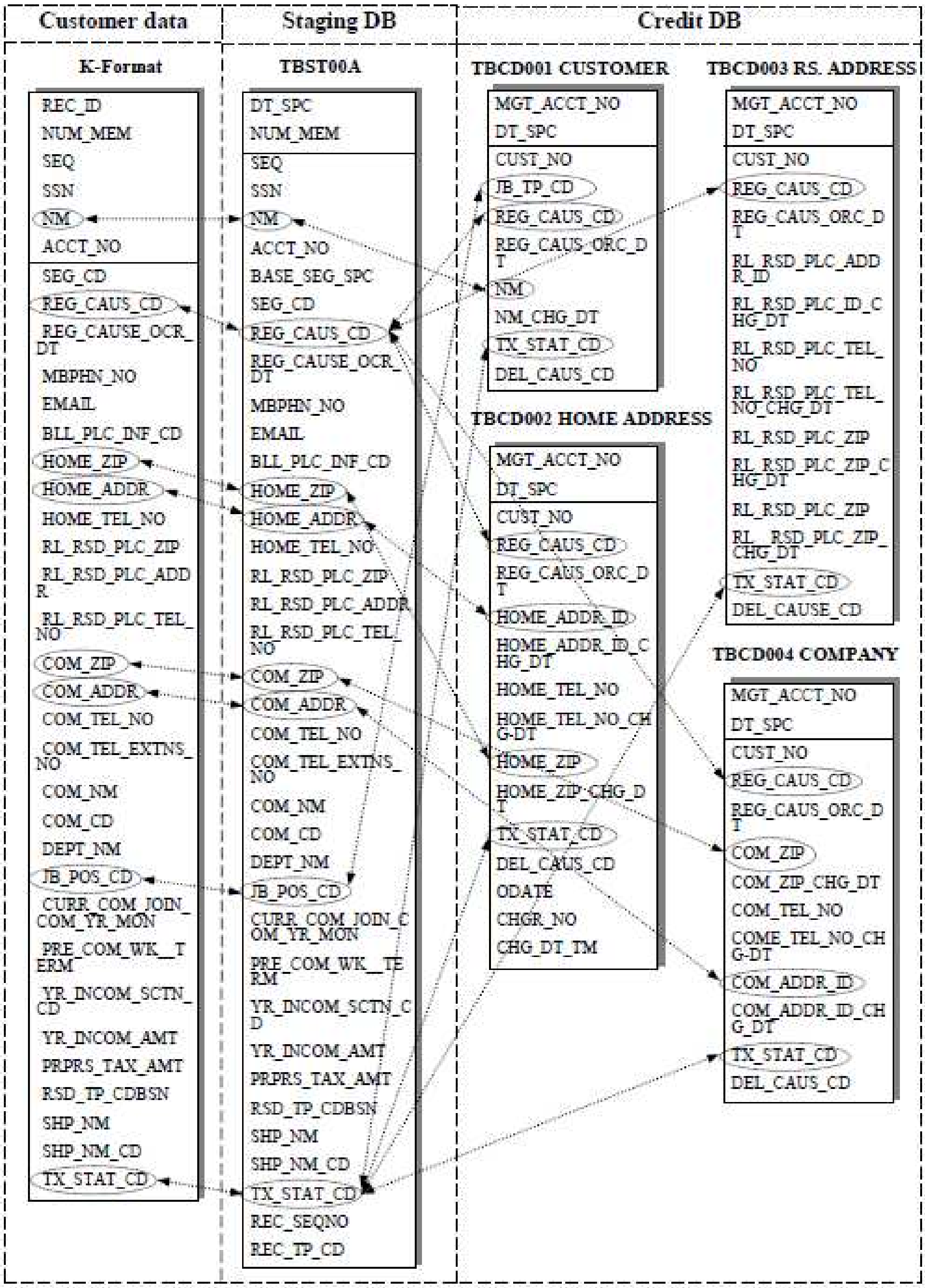

Fig. 5 Sample of data mapping in the KCB information system

The data managers in KCB apply the KCB4U software to check flow of data in batch or near real time. First, they select any attribute of entity in tables in the source; the KCB4U automatically identifies the target. The data managers' task is simple to tick off at the small squares at the first column. Next, they click the button of register to check data flow. Once checking data flow, the KCB4U shows the incorrect data flow. Finally, they clean incorrect data by ticking small squares at the first column and delete.

Table 1 and 2 show a sample of field-to-field mapping from K-format to staging DB and from staging DB to credit DB. For example, the KCB4U illustrates the connection of 'HOME_ZIP' attribute in K-format, staging DB and credit DB. In other word, the 'HOME_ZIP' attribute is in the state of matching field-to-field. However, there usually exists some of mismatching field-to-field attributes in the KCB information system. The table 14 shows some of sample mismatching attributes.

TABLE 1
SAMPLE OF DATA MAPPING FROM K-FORMAT TO STAGING DB

Source (K-format)			Target (Staging DB)		
Table	**Column**	**Type**	**Table**	**Column**	**Type**
K-format	NM	VARCHAR2 (88)	TBST00A	NM	VARCHAR2(88)
K-format	REG_CAUS_CD	VARCHAR2(2)	TBST00A	REG_CAUS_CD	VARCHAR2(2)
K-format	HOME_ZIP	VARCHAR2(6)	TBST00A	HOME_ZIP	VARCHAR2(6)
K-format	HOME_ADDR	VARCHAR2(216)	TBST00A	HOME_ADDR	VARCHAR2(216)
K-format	COM_ZIP	VARCHAR2(6)	TBST00A	COM_ZIP	VARCHAR2(6)
K-format	COM_ADDR	VARCHAR2(150)	TBST00A	COM_ADDR	VARCHAR2(150)
K-format	JB_POS_CD	VARCHAR2(8)	TBST00A	JB_POS_CD	VARCHAR2(8)
K-format	TX_STAT_CD	VARCHAR2(2)	TBST00A	TX_STAT_CD	VARCHAR2(2)

TABLE 2
SAMPLE OF DATA MAPPING FROM STAGING DB TO CREDIT DB

Source (Staging DB)			Target (Credit DB)		
Table	**Column**	**Type**	**Table**	**Column**	**Type**
TBST00A	NM	VARCHAR2(88)	TCBD001	NM	VARCHAR2(88)
TBST00A	REG_CAUS_CD	VARCHAR2(2)	TCBD001	REG_CAUS_CD	VARCHAR2(2)
TBST00A	REG_CAUS_CD	VARCHAR2(2)	TCBD002	REG_CAUS_CD	VARCHAR2(2)
TBST00A	REG_CAUS_CD	VARCHAR2(2)	TCBD003	REG_CAUS_CD	VARCHAR2(2)
TBST00A	REG_CAUS_CD	VARCHAR2(2)	TCBD004	REG_CAUS_CD	VARCHAR2(2)
TBST00A	HOME_ZIP	VARCHAR2(6)	TCBD002	HOME_ZIP	VARCHAR2(6)
TBST00A	HOME_ADDR	VARCHAR2(216)	TCBD002	HOME_ADDR_ID	VARCHAR2(216)
TBST00A	COM_ZIP	VARCHAR2(6)	TCBD004	COM_ZIP	VARCHAR2(6)
TBST00A	COM_ADDR	VARCHAR2(150)	TCBD004	COM_ADDR_ID	VARCHAR2(150)
TBST00A	JB_POS_CD	VARCHAR2(8)	TCBD001	JB_POS_CD	VARCHAR2(8)
TBST00A	TX_STAT_CD	VARCHAR2(2)	TCBD001	TX_STAT_CD	VARCHAR2(2)
TBST00A	TX_STAT_CD	VARCHAR2(2)	TCBD002	TX_STAT_CD	VARCHAR2(2)
TBST00A	TX_STAT_CD	VARCHAR2(2)	TCBD003	TX_STAT_CD	VARCHAR2(2)
TBST00A	TX_STAT_CD	VARCHAR2(2)	TCBD004	TX_STAT_CD	VARCHAR2(2)

5.2. Data stewardship management

1. *Procedure for changing data structure and flow*

In the dynamic business environment, customer's requirements are always changeable so that enterprises must adapt to them, in which data assets, in particular data structure and flow, are needed to change due to volatile business requirements. As any member company has a request to change data structure and flow, a data quality group of KCB has to solve this problem. This group is composed o data service officer, data structure operator, data flow operator at the operational level, data quality board at the tactical level, data flow manager at strategic level. They are always supported and sponsored by KCB Chief Executive Officer (CEO).

Figure 6 describes a processing chart to deal with the problem of changing data structure and flow. When a member company requests to change its data structure, data service officer receives this request and fires it to data structure operator. Data structure operator has a responsibility to analyse the impact of data structure change over all existing information system and send it to data quality board for validation and to data flow operator for analysing the impact of data flow. Next, data quality board reviews the impact of data flow change and submit it to data flow manager who makes final decision whether approve or not approve to change data flow. If what data flow needs to change is approved, it is sent to data flow operator for adjusting and registering in the existing information system. Finally, data flow is run automatically in the KCB information system.

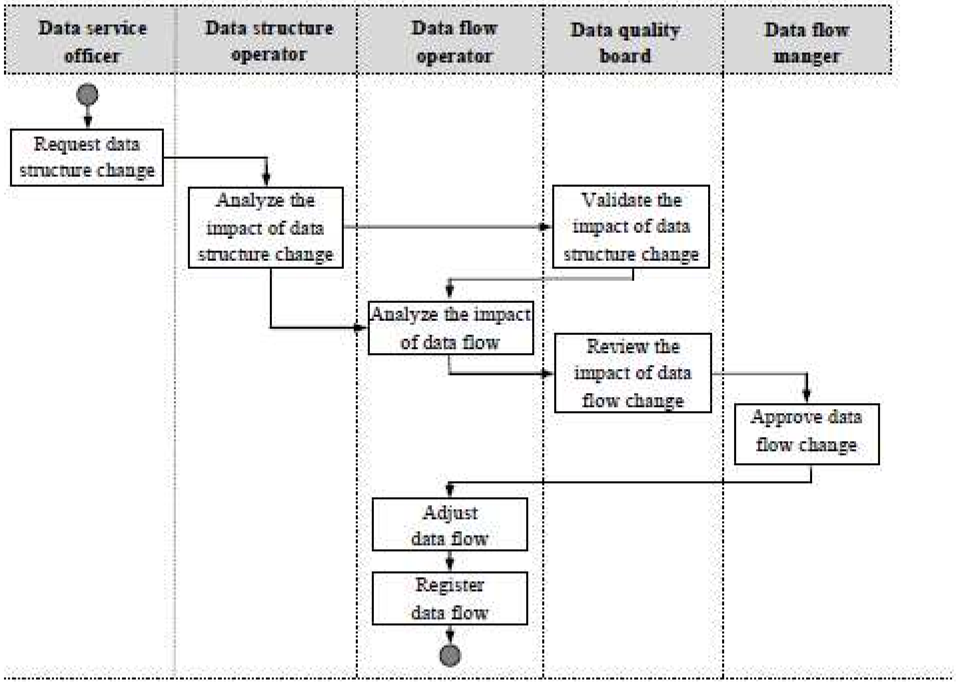

Fig. 6 Procedure for changing data structure and flow

TABLE 3
DATA STEWARDS IN KCB

Level	*Framework*	*Case of KCB*
Strategic level	Executive sponsor	Data flow manager
Tactical level	Data quality board	Data quality board
Operational level	Business data steward	Data service officer
	Technical data steward	Data structure operator
		Data flow operator

Roles and functions of individuals in data quality group are listed as follows.

(1) Data service officer: a coordinated person between member companies and technical data steward is responsible for accepting data structure change from member companies and supporting data structure and flow operator for analysing the data structure change. Data service officer plays a role as business data steward so they have to understand the business requirements from member companies and interpret their requests to technical data stewards.

(2) Data structure operator: a technical data steward who is responsible for analysing the impact of data structure change. The results of this stage are data structure changed and its report about impaction. The data structure changed is sent to data flow operator while its impact report is sent to data quality board for validation.

(3) Data flow operator: what data structure is changed influences on the data flow because of mismatching among fields and tables. Therefore, data flow operator has responsible to analysing the data flow change to ensure that data flow correctly. Other task of data flow operator is to adjust and register data flow after the data flow manager approval.

(4) Data quality board: a group that gathers cross-functional data stewards – data service officer, data structure and flow operator from operational level and data flow manager from strategic level. The data quality board, also called coordinating data steward, has responsible for validating the impact of data structure change and reviewing data flow change.

(5) Data flow manager: a senior manager serving on a data quality initiative has accountable to approving or not approving the data flow change on behalf of the CEO and shareholders. In general, data flow manager is considered as data manager because he or she is accountable for not only data flow approval but also other subject matters.

TABLE 4
DATA STEWARD ASSIGNMENT

Assignment	Data service officer	Data structure operator	Data flow operator	Data quality board	Data flow manager
Request for changing data structure	**I**	**C**			
Analyse the impact of changing data structure		**R**			
Validate the impact of changing data structure		**C**		**A**	
Analyse the impact of changing data flow		**I**	**R**	**I**	
Review the impact of changing flow			**C**	**A**	
Approve data flow		**I**		**C**	**A**
Adjust data flow			**R**		**C**
Register data flow			**R**		

R: Responsible, A: Accountable, C: Consulted, I: Informed.

Different from data flow management which is controlled and managed by data administrator, data stewardship management is organized and selected from cross-functional department of business and IT area. There are many people who are involved in the procedure for changing data flow and devoted to data quality initiatives. The important problem is how to assign right roles to right tasks or decision areas in the procedure. The thesis utilizes the matrix of assigning accountabilities in the case of changing data flow which is the main function to improve quality of data in KCB. There are 5 roles and 8 tasks in the Table 4.

As defined that responsible role performs tasks or specifies the way of performing tasks; Accountable role authorizes decisions or the results of tasks; consulted role has specific knowledge necessary for decision-making or performing tasks; informed role will be informed about decisions made or results of tasks.

(1) Requesting for changing data structure: When member companies have a request to change data structure, business data stewards receive their requests and send them to the technical data stewards. Business data stewards must inform the request to data structure operator. Additionally, data structure operator must consult the business data stewards about data structure in order to meet member companies' requirements.

(2) Analyse the impact of changing data structure: once receiving requests for changing, data structure operator has responsible for analysing the impact of data structure because of change.

(3) Validate the impact of changing structure: data quality board has accountable for validate the impact of changing data structure. In addition, data structure operator must supply the report of changing and consult data quality board about technical aspect.

(4) Analyse the impact of changing flow structure: similarity to the second task, data flow operator has responsible for analysing the impact of changing flow structure. Beside, data structure operator and data quality board must inform the data

structure changed and its impacts to data flow operator.

(5) Review the impact of data flow change: data quality board is accountable for impact of data flow change. In order to check, they are consulted by the data flow operator.

(6) Approve data flow change: on the top level of data stewardship, data flow manager has accountable to approving the data flow change under the consult of data quality board. Next, data flow manager must inform the approval to data flow operator in order to adjust and register data flow.

(7) Adjust and register data flow: the data flow operator has responsible to adjusting the data flow change and registering it to the KCB information system.

Once whole stages of this procedure are completed, a nouvelle data flow is load into to KCB information system and operated automatically. Briefly, thanks to data stewardship management, data flow is improved continuously and imperfection of data flow is decreased. That leads to quality of data better.

6. CONCLUSION

The data stewardship and flow framework helps organization improve quality of data from the perspective of both vertical and horizontal side management. In this framework, we model data flow through IDEF0 and, then, use software named kcb4u to map data from input to output. In addition, we can trace where data errors can occurs and reach where to impact. Thanks to data flow, we can control and prevent low-quality data at source. In the other aspect, data quality is improved through data stewardship management. A procedure is proposed to adjust data flow and structure more accurately and a RACI matrix is designed to organize data stewardship effectively. Assigning right roles to right decision area with right accountability contributes to data quality management from the organizational viewpoint.

REFERENCES

[1] Ballou, D., Wang, R., Pazer, H. and Tayi G.K. (1998). "Modeling Information Manufacturing Systems to Determine Information Product Quality", *Management Science*, Vol. 44, No.4, pp. 462-484.

[2] Canadian Institute for Health Information (2009). *The CIHI Data Quality Framework*, Ottawa.

[3] Department of Defense (2001). *Systems Engineering Fundamentals*, Supplementary text by the Defense Acquisition University Press, Virginia.

[4] Eckerson, W.W. (2002). *Data Quality and the Bottom Line: Achieving Business Success through a Commitment to High Quality Data*, the Data Warehousing Institute.

[5] English, L.P. (1999). *Improving Data Warehouse and Business Information Quality*, John Wiley & Sons, New York.

[6] Eppler, M.J. (2006), *Managing Information Quality*. 2nd ed., Springer, Berlin, Germany.

[7] FIPS Publication 183 released of IDEF0 December 1993 by the Computer Systems Laboratory of the National Institute of Standards and Technology (NIST).

[8] IBM Software Group (2007). *IBM Data Governance Council Maturity Model: The IBM Data Governance Council Maturity Model: Building a Roadmap for Effective Data Governance*.

[9] Kim, S. and Lee, C. (2011). *Process-centric Data Quality Management Framework*, working paper.

[10] Marca, D.A., McGowan, C.L. (1988). *SADT Structured Analysis and Design Technique*, McGraw-Hill Book Company.

[11] Mosley, M.., Brackett, M., Earley, S. and Henderson, D. (2009). *The DAMA Guide to the Data Management Body of Knowledge*, The Data Management Association.

[12] Otto, B., Wende, K., Schmidt, A. and Osl, P. (2007). "Towards a framework for corporate data management", *The 18th Australasian Conference on Information Systems*.

[13] Redman, T.C. (1996). *Data Quality for the Information Age*, Artech House, London, UK.

[14] Ryu, K.S., Park, J.S. and Park, J.H. (2006). "A data quality management maturity model", *ETRI Journal*, Vol.28, No.2, pp.191-20.

[15] Scannapieco, M., Pernici, B. and Pierce, E. (2005). "IP-UML: Towards a methodology for quality improvement based on the IM-MAP framework", *Proceedings of the Seventh International Conference on Information Quality*.

[16] Shankarananrayanan, G., Ziad, M. and Wang, R.Y. (2003). “Managing data quality in dynamic decision environments: an information product approach”, *Journal of Data Management*, Vol. 14, No. 4, 14-32.

[17] Shankaranarayanan, G. and Wang, R., (2000). “IPMAP: representing the manufacture of an information product”, *Proceedings of the 2000 Conference on Information Quality*.

[18] U.S. Federal Data Architecture Subcommittee (2008). *Data Quality Framework Applying Proven Quality Principles to Data Sources*.

[19] Wang, R.Y., Lee, Y.W., Pipino, L.L. and Strong, D.M. (1998). “Manage your information as a product”, *Sloan Management Review*, Vol.39, No.4, pp.95-105.

[20] Wang, R.Y. (1998). “A product perspective on total data quality management”, *Communication of the ACM*, Vol. 41, No.2, pp. 58-65.

[21] Wende, K. (2008) Data Governance Defining Accountabilities for Data Quality Management. In: Isari, D. (Hrsg.); D'Atri, Alessandro (Hrsg.); Winter, Robert (Hrsg.): Luiss University Press.

[22] Wende, K. (2007). “A Model for Data Governance- Organizing Accountabilities for Data Quality Management”, *18th Australasian Conference on Information Systems*.