The Analysis of Data Management, a Case Study of a Government Information System in Thailand

Wasinee Noonpakdee[1], Jirapon Sunkpho[1], Jirapon Tubtimhin[1], Taweesak Kijkanjanarat[2], Nawaporn Wisitpongphan[3]

[1]Collage of Innovation, Thammasat University, Thailand.
[2]Faculty of Engineering, Thammasat University Thailand.
[3]Faculty of Information Technology, King Mongkut's University of Technology North Bangkok, Thailand.
wasinee@citu.tu.ac.th

Abstract— **In this paper, we present our analysis of data management, a case study of a government information system in Thailand. The case study information system has been implemented and developed to collect data for monitoring and surveillance of social problems. The system has been employed by many government agencies. However, there have been some issues of data management. In this research, we analyzed the issues and requirements of the system by conducing in-depth interviews and focus groups which are divided into three groups: top management, user, and system administrative. According to the analysis, the problems of the system consist of six main factors: management and policy, software, infrastructure and hardware, process, people, and data. The solution and guideline of the system are proposed. Moreover, we suggest the system architecture which consists of three tiers: physical layer, service broker, and application layer. The results can be applied for improving the system, and can be a guideline for managing data in other information systems.**

Index Terms— **Data integration; Data management; Information system; System architecture.**

I. INTRODUCTION

For decades, the intelligent application of data has become one of the greatest competitive advantages in the organization [1]. The organization needs appropriate decision support tools which can facilitate decisions in order to cope with the environmental and technological changes to sustain competitiveness and remain in business [2]. Organizations are always required to gather data which will be used in developing a better opportunity [3]. In the system, data sharing between different database systems is one of the significant problems [4]. There are some challenges of Data Integration and Interoperability especially in Big Data [5, 6]. Data integration has gained a lot of research interests and has been studied in [7-9]. Furthermore, data management and policy is studied in [1, 6, 10, 11].

A case study government sector has developed an information system to collect data for monitoring and surveillance of social problems in Thailand. There are related agencies who involve in the importing data including some departments and ministries. The data is inputted and the system is accessed from many locations all over the country. From the time when the system has been developed, there have been some problems for managing data, such as lacking integrity, data redundancy, and data obsolescence. The collection, analysis, and monitoring could not be carried out efficiently. This results in the difficulty of data sharing and application. Therefore, this research aims to analyze and propose some solutions to data management issues for the case study's information system in order to ensure data integration, and sustainable application.

The rest of this paper is organized as follows. In Section II, theoretical foundation is explained. In Section III, the research methodology is provided. Section IV presents result and discussion. The problems and requirements are analyzed. The proposed solutions and system architecture are described. Finally, Section V concludes the paper.

II. THEORETICAL FOUNDATION

A. Information Systems

Information systems (IS) are combinations of hardware, software, and networks that people build and use to collect, create, process, and distribute useful data to support decision making and control in an organization. In addition to supporting decision making, coordination, and control, information systems may also help people analyze problems, visualize complicated subjects, and initiate new products [12, 13].

Hardware refers to physical computer equipment. Software refers to a program or set of programs that inform the computer to perform certain tasks. Networks refer to a group of two or more computer systems linked together with communications equipment [12].

An effective information system provides users with accurate, timely, and relevant information. Information is timely when it is available to decision makers when it is needed. Information is relevant when it is useful and appropriate for the types of work and decisions that require it. Therefore, data management is very necessary to support an effective information system [13].

B. Database Approach to Data Management

Database technology decrease some problems of traditional file organization, such as data redundancy and inconsistency, program-data dependence, inflexibility, poor data security, and an inability to share data among applications. A database is defined as a collection of data organized to serve many applications efficiently by centralizing the data and controlling

redundant data [13].

In a centralized database, all the components exist on a single computer or computer site. Access to the data stored in the centralized site has been provided by the use of intelligent workstations and remote access terminals that utilized directed communication links. A distributed database is a store of data that should be logically housed together, but it has been spread over a large geographic area [14].

III. METHODOLOGY

The research was conducted as a case study: an information system in a government sector. The research process consists of literature study and empirical study as presented in Figure 1.
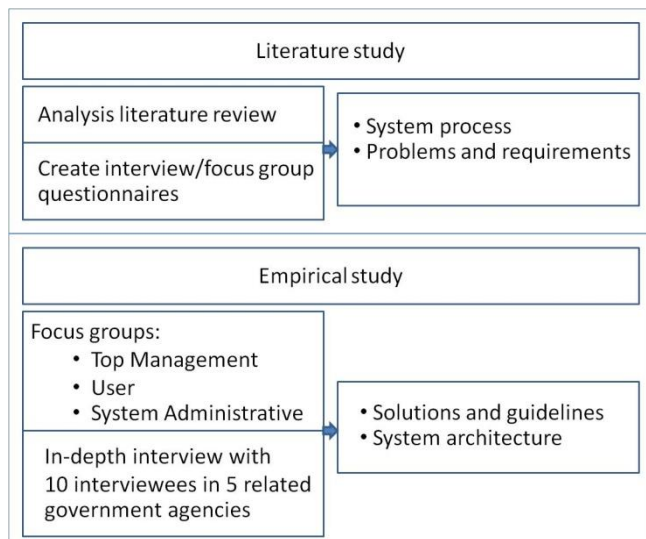

Figure 1: The research process

Literature study is applied to understand the system process and requirements. The problems and requirements are studied including data, user interface, process, and functions of the system. Regarding empirical study, focus groups are employed by dividing into three groups: (1) Top Management (2) User and (3) system administrative. The details of each group are described in Table 1.

Table 1
Focus groups

Group	Detail
Top management	The top management employs the system to help making a decision. The topic focuses on issues of implementation and policy direction.
User	User directly involves in the system by inputting and using data. The questions focus on issues of the operation, process, function, and data integration of the system.
System Administrative	This group is responsible for infrastructure and technical problem solving. The discussion is about system's problems, data management, and data integration.

Moreover, in-depth interviews with ten interviewees in five related government agencies are conducted to obtain deep comprehension about the issues, requirements, and system processes.

IV. RESULTS AND DISCUSSION

A. Problem and Requirement Analysis

According to the study, the issues of data management of the system include management and policy, software, infrastructure and hardware, process, people, and data.

i) Management and policy

This system has to access the database from civil registration; however, the civil registration allows the access on the specific time, for example only in office hours. Therefore, it is inconvenience for the employees who input data, and sometimes the employees decide not to input data in the system. Moreover, for the maintenance of this system, the budget is not adequate.

ii) Software

Many organizations have developed their own systems for their own use since this case study's information system is not exactly matched their requirements. Furthermore, the system that they developed cannot automatically connect and transfer data to the case study's system. Therefore, the organizations should provide their staff to input data to both their own developed system and the case study's system. Inputting data manually many times might result in some problems of data accuracy and redundancy.

iii) Infrastructure and hardware

There is only one server for the system located at a ministry with database that cannot support the large amount of data. The bottlenecks happen when the database is connected from many places. Moreover, a user has to connect to the database from civil registration and a server at the ministry. So if any network problems occur at one point, the system cannot be operated which causes stability issues.

iv) Process

The work flow of this system is unclear. There are a lot of agencies connecting to the system. However, these agencies still do not understand their role on the system. They do not know who should input data on the specific process.

v) People

The employee who is responsible for inputting data has a lot of workload to do, and he/she does not have time to input data. Moreover, each employee has different expertise and understanding, so data is inputted in the different way. Consequently, the data is not standardized.

vi) Data

Each agency has different dimensions of data and data cannot be used for comparison. Moreover, data is not updated. Database of agencies that use the system has different data field. Moreover, the system is not support data requirement of the agencies.

B. Proposed Solutions and Guidelines

According to the focus groups, in-depth interviews, and analysis, the solutions of data management of the system are proposed as the following.

i) Preparation of data standards and integrated database

All agencies involved with the system should be in charge of establishing data standards and integrated database. The integrated database should be compatible with all agencies' databases. All agencies should clearly define the main entity that is responsible for inputting data in each process. Moreover, data from each agency should be verified, for example checking with the civil registration database before linking to the integrated database.

ii) Guideline for accessing civil registration database

To solve the problem of accessing civil registration database which allow users to access only in a specific time, the system should allow user to submit the form anytime without connecting to civil registration database. Then, during the time for access for example the office hours, the system should automatically connect and update data.

iii) Guideline to handle bottleneck in the system

Instead of implementing a centralized database that is located, stored, and maintained in a single location, we proposed to employ a distributed database system that users can access the database at their location. In distributed database system, the transaction should be executed very fast in a flexible load balancing database environment [15, 16]. The mirror database server can be installed in some specific locations all over the country. The data is synchronized and backed up in a specific time. Moreover, the bandwidth at the center server should be expanded to handle large amount of traffic. Since the current system uses open-source and inexpensive database which has stability issues and poor performance scaling, the higher performance database can be considered in order to improve the system. However, the cost of implementation might be increased.

iv) Guideline for developing and improving the system

In order to develop and improve the system, the three main factors are taking into consideration: affordability, availability, and accessibility.

- Affordability: To improve the system, the budget should be prepared for each agency for connecting the integrated database, for maintenance the system in the future, and for improving the infrastructure such as server and network. In the future, Virtual Private Network (VPN) can be applied for improving network speed and security.
- Availability: The improvement or development in the next step should not cause the disruption of the system. The system that is undergoing maintenance has to be always available. The development of systems using a distributed database system or cloud system should be applied.

- Accessibility: The current database system cannot be scalable. Therefore, in the future, the higher performance database might be considered for accessibility of the system.

C. The Proposed System Architecture

According to problem and requirement analysis, the system architecture is proposed as illustrated in Figure 2. The architecture consists three tiers: physical layer, service broker, and application layer.
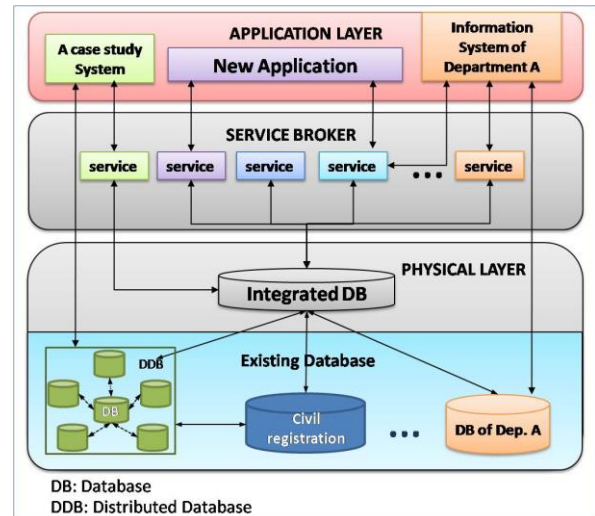

Figure 2: The proposed system architecture

i) Physical layer

This layer is composed of two main components: existing database and integrated database. The existing database is the database that all agencies such as some departments and some remote offices inputting data. The integrated database is a proposed database that will be newly developed. The integrated database will collect data from all the relevant agencies and storage to meet the data standards established information sharing. The original database that is already existed may be modified from a centralized database to distributed database which should be scattered in different regions in the mirror database.

This method will increase availability in case of network problems, and will increase speed and channel to access the system. The bottleneck problem is also solved. However, with distributed database system, information security would be one of major concerns. Moreover, the organization has to assign some staff who will responsible for managing the database in the regions.

ii) Service Broker

A service broker is studied in [17, 18]. In our proposed architecture, a service broker is acting as intermediaries between the information systems of each agency with integrated database. From Figure 2, users use their own information system in application layer. Then, the service broker provides the service to connect to the database. This will reduce data duplication problems.

iii) Application Layer

This layer is composed of information system of related agencies that employ the case study information system. All the system will connect to the services in service broker to store data in the integrated database. Users of existing systems do not need to modify their process because a service broker will manage and adjust the data dimensions to meet the standards.

In the future, if any agencies would like to develop their own information system that connects to this system, the agencies can applied service broker in the development of their system.

V. CONCLUSION

In this paper, we analyze data management problems of a case study of a government information system in Thailand. According to the analysis, the problems of the system consist of six main factors: management and policy, software, infrastructure and hardware, process, people, and data. The solution and guideline of the system are proposed such as the preparation of data standards and integrated database, the guideline to handle bottleneck in the system, and the guideline for developing and improving the system. We also propose a system architecture which consists of three tiers: physical layer, service broker, and application layer. The results can be applied for improving the system, and can be a guideline for implementing data management in other information systems. The future research includes the implementation plan, cloud computing, and security of the system.

REFERENCES

[1] A. H. Raymond, "Data management regulation: Your company needs an up-to-date data/information management policy," Business Horizons, vol. 56, pp. 513-520, 7/ 2013.

[2] B. S. Sahay and J. Ranjan, "Real time business intelligence in supply chain analytics," Information Management & Computer Security, vol. 16, pp. 28-48, 2008.

[3] R. T. Hans and E. Mnkandla, "Modeling software engineering projects as a business: A business intelligence perspective," in AFRICON, 2013, 2013, pp. 1-5.

[4] H. Xin and R. Chunming, "Multiple Data Integration Service," in 28th International Conference on Advanced Information Networking and Applications Workshops (WAINA), 2014, pp. 860-865.

[5] A. Kadadi, R. Agrawal, C. Nyamful, and R. Atiq, "Challenges of data integration and interoperability in big data," in IEEE International Conference on Big Data (Big Data), 2014, pp. 38-40.

[6] V. Kantere, "A Holistic Framework for Big Scientific Data Management," in IEEE International Congress on Big Data (BigData Congress), 2014, pp. 220-226.

[7] Z. Bo, "Data integration as a service for Applications Deployment on the SaaS Platform," in 6th International Conference on Biomedical Engineering and Informatics (BMEI), 2013, pp. 672-676.

[8] S. Kher, J. Dickerson, and N. Rawat, "Biological pathway data integration trends, techniques, issues and challenges: A survey," in Second World Congress on Nature and Biologically Inspired Computing (NaBIC), 2010, pp. 177-182.

[9] Z. Lingli and Z. Renliang, "Analysis and Design for Data Integration System of Urbanization Based on GIS Open Source Technology and Arc Engine," in International Conference on Environmental Science and Information Application Technology, 2009, pp. 146-149.

[10] T. Yongsheng, W. Yibing, and L. Hongyan, "A management strategy of monitor data in ICU based on data stream technology," in IDEAS Workshop on Medical Information Systems: The Digital Hospital (IDEAS-DH'04), 2004, pp. 163-169.

[11] S. M. Mesick, S. T. Gottfried, and E. Roby, "National Oceanic and Atmospheric Administration Office of Ocean Exploration end-to-end data management: a success story," in Proceedings of MTS/IEEE OCEANS, 2005, pp. 2070-2077 Vol. 3.

[12] Joseph A. Valacich and C. Schneider, *Information Systems Today: Managing in the Digital World*, 4 th ed.: Prentice Hall, 2010.

[13] Kenneth C. Laudon and J. P. Laudon, *Management Information Systems, Managing the Digital Firm*: Pearson Education Limited, 2014.

[14] C. D. Tupper, "22 - Distributed Databases," in *Data Architecture*, ed Boston: Morgan Kaufmann, 2011, pp. 385-400.

[15] I. O. Hababeh, "Intelligent Network Communications for Distributed Database Systems," in Second International Conference on Advances in Databases Knowledge and Data Applications (DBKDA), 2010, pp. 69-74.

[16] J. A. Hoffer, R. Venkataraman, and H. Topi, Modern Database Management: Prentice Hall, 2013.

[17] S. Colbran and M. Schulz, "An update to the software architecture of the iLab Service Broker," in 12th International Conference on Remote Engineering and Virtual Instrumentation (REV), 2015, pp. 90-93.

[18] N. M. DeVito, R. T. Emery, K. F. Kocan, W. D. Roome, and B. J. Williams, "Functionality and structure of the service broker in advanced service architectures," Bell Labs Technical Journal, vol. 10, pp. 17-30, 2005.