

Design and Pilot Test of Software Prototype for Linking National R&D Information with Result Materials

Inje Kang¹, Byeong-Hee Lee^{2*}, Seulki Kim¹, Chulsu Lim² and Ki-Seok Choi²

¹Science and Technology Management Policy, University of Science and Technology, 217, Gajeong-ro, Yuseong-gu, Daejeon, 34113, Korea; ijkang@ust.ac.kr, seulkim1@ust.ac.kr,

²National Science and Technology Information Service Center, Korea Institute of Science and Technology Information, 245, Daehak-ro, Yuseong-gu, Daejeon, 34141, Korea; bhlee@kisti.re.kr, cslim@kisti.re.kr, choi@kisti.re.kr

Abstract

Objectives: With open public data trend, Korean NTIS is considering national R&D information system including result materials. This research designs information integration software for linking the information with result materials.

Methods/Statistical Analysis: To provide information integration in Korean NTIS including result materials along with 422 metadata items, this study designs and pilot tests a software prototype. The software prototype is designed in Agent Mode, which is based on previously existing standard platform. 120,000 cases of data and 2.5GB of result materials were used to test the software prototype. **Findings:** According to the test result, the entire result materials were transferred in 250 seconds, while the renewal ones were sent in 50 seconds. It implies that the result materials can be transferred time-efficiently through the software. Preceding researchers studied on developing information integration system based on standard platform; however, there was lack of studies on developing the system including result materials. This study suggests a way to provide information service by linking national R&D information with result materials included as well as 422 metadata items. **Improvements/Applications:** To apply the proposed target model to the existing NTIS environment, further studies need to verify the stability or system quality even when transferring large amount of result materials.

Keywords: Information Integration, Korean NTIS, Metadata, National R&D Information, Pilot Test, Result Material, Software Prototyping

1. Introduction

In the past, public information was only accessible to limited people. However, these days, people's right-to-know started to be protected and has become global trend¹. Korea also has been endeavoring to satisfy the public's demand of publicly accessible information by several revisions of law². It is one of the government's policy to establish advanced systems in order to improve the transparency of government administration and enhance the efficiency of R&D productivity^{2,3}.

Another problem of public information is that it is relatively more diffused to the past because there is no centralized information management⁴. The lack of federated

information management or 'inter-agency collaboration' decreases policy effectiveness or prevents the public participation due to insufficient information⁵. However, the inter-agency collaboration is getting more imperative as public sectors tend to be relatively more dependent on information management⁶, and more active usage of information on R&D is required to increase the efficiency of science and technology³.

Korea is also aware of the problem of usage of the related information and absence of cooperation among agencies and research institutes, causing inefficiency of National R&D investment^{6,7}. To improve the utilization of information on national R&D, data needs to be gathered and provided to public more precisely and integratively^{2,3,8}.

*Author for correspondence

Hence, the National Science and Technology Information Service (NTIS) of Korea was introduced in 2008^{9,10}. NTIS has been compiling detailed information on national R&D programs or projects and has been providing the compiled information to the public, government institutes or universities¹¹. It integrates and manages information on the programs, based on the 'national R&D information standards', mentioning to include information of performance, facilities, or workforces. By centralizing the information, it can help the national R&D participants to prevent duplicate R&D investment and to increase efficiency of R&D projects¹². However, the request for more publicly accessible information is increasing, and the government is promoting agencies to share more information regarding national R&D projects.

Due to the request for the changes in information service environment, NTIS has been considering to provide improved information service by integrating information with result materials contained within. Therefore, the aim of this study is to develop software providing information integration including result materials as well as 422 metadata items to implement the policy paradigm of open public data trend. To achieve the goal, we focused on developing software for NTIS to integrate result materials produced from national R&D projects along with existing DB. We designed and pilot tested the software prototype to examine the feasibility of applying to the existing information integration environment among NTIS and other agencies.

2. Relevant Research

2.1 The Concept of ETL Process and Preceding Research

Commonly, during the process of constructing central DB, such as data warehouse, the quality of compiled information can deteriorate⁶. Hence, the process of 'cleansing data' is necessary to improve the quality and to confirm consistency, and it is called Extraction-Transformation-Loading (ETL)^{6,13,14}. An ETL tool is to take responsibility for the extracting, cleansing, and customizing data¹³.

The first phase of ETL is designed to extract the data periodically. Because extraction is conducted periodically and repetitively, it is important to establish principles or rules in the beginning. The second phase is transformation of data, which is also called as the process of

refinement. Because, each organization has different Database Management System (DBMS), data needs to be transformed or refined in order to improve the accuracy and consistency. The last phase is transportation, where the transformed data is transferred to central DB⁶. Because of its role of refining and centralizing sources, many studies were conducted to utilize the ETL tools to spur the usage of information.

In¹⁴ created a prototype model of the data repository called Business Intelligence (BI) to provide one composite view regarding radiology. They applied ETL to integrate sources, check for consistency, and convert the sources into the unified form. In¹¹ focused on conceptualizing the notion of ETL activities by offering demonstration.

2.2 Information Integration System in NTIS

At present, NTIS is performing information integration with 17 ministries or agencies (representative research management institutes). It manages information consisting of projects, outputs, or facilities and categorizes them into 422 standard information items (metadata items) according to the national R&D information standards.

In¹⁵ studied on standard framework for information integration between two heterogeneous systems, digital Brain (dBrain) and NTIS to objectively examine national R&D programs. In¹⁶ designed NTIS information integration platform to standardize information integration for increasing the efficiency of R&D management. By examining the state of Information Integration System (IIS) between NTIS and 15 ministries or agencies, they constructed the integration platform. Then, they compared several methods of information integration models and analyzed expected effect.

According to¹⁶, to integrate information based on standardization, variation area and common area need to be identified. Variation area is where data regarding national R&D transfers from institutional DB to NTIS integration DB, and common area, where standardization of DB is possible, is where the data is transferred from NTIS integration DB to NTIS Central DB.

Based on the concept of abovementioned, NTIS has constructed national R&D II under the national R&D information standards. It is managing information standardization with 17 representative research management institutes, 10 research outcome institutes, and 2 joint research institutes. When the institutes transfer institutional DB, the DB is classified into three types,

organization DB, error DB, or monitor DB. When the consistency of the data has been confirmed, then it is categorized into organization DB, and it is transmitted to NTIS central by information integration software. Then the data is publicly accessible to the public, government agencies or universities.

As explained above, NTIS plays an important role in managing information on national R&D programs. To perform its role of gathering and managing the data of national R&D projects, NTIS needs to reflect characteristics of its function of integrating information and characteristics of its status of sustaining cooperative relationship. First, NTIS needs to retain several functional features such as confirming renewability, performing in standardized method, or monitoring information integration to guarantee the credibility of data provided. Secondly, NTIS needs to show its role of maintaining the cooperative relationship with other institutes or agencies by reflecting characteristics of each institute. Table 1 shows the summary of functional and relational aspects of NTIS IIS.

3. Architecture Design of NTIS IIS

NTIS has been compiling national R&D information and categorized it into metadata items informed by Ministry

of Science, ICT and Future Planning (MSIP). In 2016, MSIP announced to increase the number of metadata items from 389 to 422 “to expand the sharing of national R&D information and to promote cooperation among agencies or ministries to improve patency or utilization”¹⁷. To implement the announcement made from MSIP, NTIS IIS needs to establish an environment to compile the 422-metadata items as well as result materials.

Providing result materials to public is one of the three main strategies of NTIS 4.0 to prevent duplicate or enhance transparency of national R&D¹⁸. In addition, allowing the public to access to information can promote participation of national administration as well as protecting people’s right-to-know¹⁹. Thus, NTIS is considering ways to develop IIS including nine main result materials such as research plan or research report. Before designing a software prototype to achieve the goal, this paper first identified aspects of result materials to prepare fundamental preconditions Table 2.

3.1 Schematic Diagram of IIS with Result Materials

Currently, under the regulation of the management of national R&D, NTIS has IIS based on the national R&D information standard.

Table 1. Aspects of NTIS IIS

Category	Aspects
Functional Aspects	Providing service based on information integration platform Compiling data lively Monitoring information integration Maintaining capacity to response swiftly Checking business rule(BR) conformity, renewability, and data quality Strengthening security
Relational Aspects	Reflecting modification announced from NTIS administrative advisory group and NTIS expert advisory group Reflecting the institutes’ characteristics such as institutional DB system or integration schedule

Table 2. Preconditions to design software prototype

Preconditions	Details
Capacity to transfer full result materials	Continuous downloading is necessary for result materials. Result materials needs to be checked whether there is damage and retransfer if necessary.
Enhancing Security	Transferring result materials needs Secure Sockets Layer (SSL, security protocol) or encrypted application to strengthen security.
Monitoring Integration	The system needs to monitor such as transmission error or integration state and retransfer if transmission error occurs during transmitting result materials.
Interoperability	When transferring metadata and result materials, separate/integrate transfer should be easily possible because metadata and result materials can be separately managed according to the institutes’ characteristics.

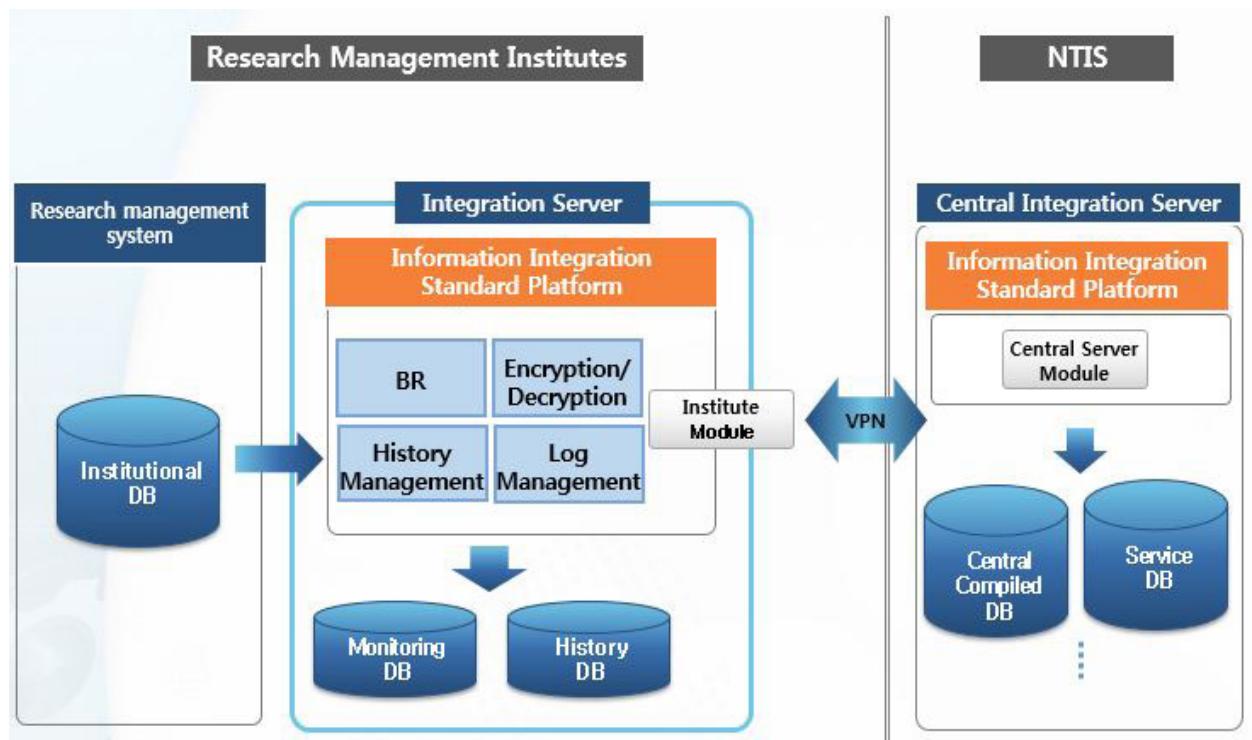


Figure 1. AS-IS model of NTISIIS.

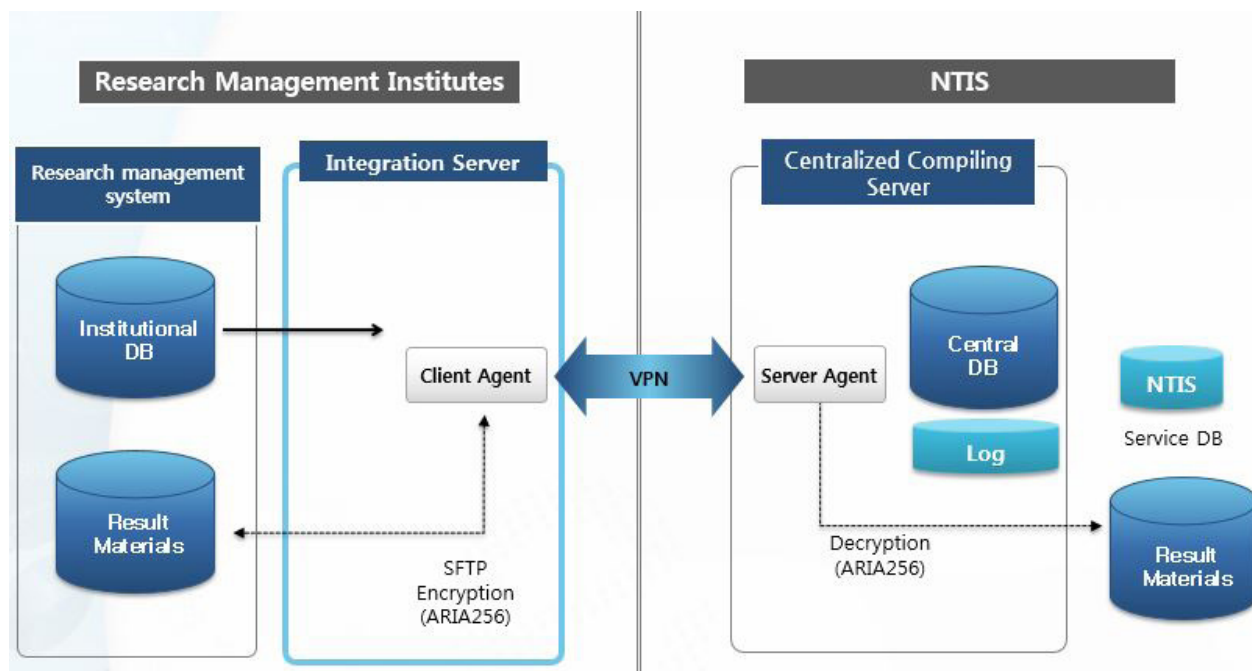


Figure 2. To-Be system configuration of agent mode including result materials.

Figure 1 demonstrates how NTIS is operating information integration. Integration server is installing into the research management institutes, and through infor-

mation integration standard platform, DB is transferring to NTIS. Virtual Private Network (VPN) is also service to secure the DB during transmission

Based on the currently existing information integration environment in NTIS, this study designed to-be model of NTIS information integration including result materials. Institutional DB and result materials will be transfer from integration software, which is installing into institutional systems, to NTIS integration server. Then, NTIS provides integration DB and result materials after going through processes of confirming credibility of the DB by checking compatibility or BR.

3.2 Design of NTIS IIS in Agent Mode

The structural aspects of the proposed model are that integration software is installing into the research management institutes, and the DB transferred from the software from to the NTIS integration servers. Figure 2 shows a system configuration of the study's proposed method for NTIS to perform information integration with result materials included by installing the agent. As shown in Figure 2, this paper designed software prototype of Agent Mode to apply it to current integration environment. The Agent Mode is based on existing standard platform and consisted of one information integration standard platform of NTIS central and integration Agent from various institutes.

The agent was installed into the institutional systems to integrate them with NTIS central. NTIS then can compile result materials through the existing integration server and VPN. As it is performing information integration based on the currently operating environment, the Agent Mode is considering to strengthening the platform by centralizing it and enhancing the operation efficiency.

Table 3. Summary of Agent Mode and Open API Mode

Category	Agent Mode	Open API Mode
Aspects	Information integration standard platform located at NTIS central Consists of one information integration standard platform at NTIS central and integration agent ATA numbers of the institutes (1:N)	NTIS first constructs Open API server(web service) and then provide the manuals to the institutes Institutes can transfer data to NTIS in xml form by applying their system according to the manuals directly.
Advantages	Confirmed security Swift reaction Versatility or Scalability No need of additional job to transfer result materials	Reassuring institutes due to direct system management
Disadvantages	Prior consultation with the institutes	Increase of expanse due to Open API installation or management Difficult to find responsibility Additional job is needed to confirm renewability or compatibility Additional development is needed to integrate result materials

The security has been confirmed by the standard platform, and versatility or scalability is improved by its simplification. In addition, as the Agent Mode is designing to have centralized function, it is possible to react swiftly to error or modification request. Another advantage of designing the Agent Mode is that, unlike Open API Mode, additional job is not needed for information integration to include result materials. Open API Mode is method of providing manuals or sample sources to the institutes after developing open API for NTIS information integration.

Table 3 summarizes the aspects, advantages and disadvantages of the modes. Open API Mode allows institutes to manage directly and reassures them; however, supplementary works are required to develop ways to transfer result materials. Moreover, because it is difficult to transmit high volumes of files through open API, this paper conducted research on designing software prototype in the Agent Mode.

4. Software Prototyping and Pilot Test

4.1 Software Prototyping

The goal of this study is to construct national R&D information integration including result materials in addition to 422 national R&D metadata items. Therefore, to achieve the goal efficiently, this paper first identified which system to apply the new design by comparing IIS including NTIS.

Table 4. The Comparison of NTIS with other IIS

Category		NTIS	A	B	C	D	E
Runtime performance	Error management	○	△	△	△	△	△
	Processing runtime performance	○	○	△	△	△	△
	Load Balancing	○	○	○	△	○	X
	Connection Cache	○	△	○	X	X	X
Encryption function	256Bit Encryption Module	○	△	△	△	△	△
	Interlocking electronic certification	△	○	○	○	○	○
Operation and Management	Centralized administration and providing web management tool	○	○	○	△	○	△
	Monitoring the usage of system resource and loading condition	○	△	△	△	△	△
	Monitoring workflow	△	○	○	△	△	○
	Checking version	X	X	X	X	X	X
Failure detection and recovery	Automatic restart when instance failure occurs	X	X	X	X	X	X
	Fault tolerance	○	△	○	X	○	X
	Alerting error	○	○	△	△	○	△
Standard support	TCP/IP, HTTP	○	○	○	○	○	○
	XML	○	○	○	○	○	○

Based on the result from comparing IIS, this study aims to design software prototype information integration to include result materials Table 4. The criteria for the pilot test were considered, and this research conducted pilot test to demonstrate how fast the prototype can compare renewability and how fast it can transfer data or result materials.

4.2 Pilot Testing

This study conducted pilot test of Agent Mode prototype. Table 5 shows the composition of the program, and Table 6 demonstrates the test environment. As shown in Table 5, the agent and the central server are the main compositions of the prototype to conduct the pilot test. The agent, who is installing into the institutional systems to integrate with NTIS, inquires DB, compare renewability, and apply encryption to the data and files. In addition, update data is transmitting to the Server from the agent, along with files of result materials to the central server. From the central server, data of the comparison of renewability is transferring to Client while receiving result materials. It also checks consistency of data such as type or length and receives result materials from the agent.

Table 5. Composition of the program

Category	Main Function
Agent	Inquiring DB and checking renewability Transferring renewal data Applying encryption to data and files(ARIA 256) Transferring files of result materials
Central Server	Transferring renewability comparison data Checking consistency(PK, FK, NULL, TYPE, LENGTH) Applying renewal data(IUD) Receiving result materials

The test environment was intentionally formed for improving objectivity or accuracy of the capacity of the software prototype. We did not want to get preferable result caused by well-equipped environments. This research conducted the test to demonstrate how fast 120,000 cases of data and 2.5GB of result materials can be transferred, and how fast update data can be compared even in poor condition.

Table 7 and Table 8 demonstrate the result of pilot test of Agent Mode prototype. It took 71 seconds to check renewability by comparing Server data and Client data,

and took 24 seconds to transfer and save the update data. When transferring the whole data, it took 95 seconds, meaning that about 1,260 cases were transmitted in a second Table 7.

Table 6. Test environment

Category	Details
Network	ipTime SG24000, 10Mbps
Client	Core i5 2.5GHZ
Server	Core i5 2.5GHZ
Data	120,000 cases
Result Materials	5Mbyte · 500 cases(total 2.5GByte)

When this research tested transferring result materials, the function of retransferring the files was applied Table 8. It took 1.2 seconds to compare Server data and Client data to identify update result materials. While it took 24 seconds to transfer update data, as mentioned above, transferring update result materials took 50 seconds.

Table 7. Test result of transferring data

Contents	Result
Comparing updatedata	71 seconds (Comparing Server data and Client data)
Transferring update data	24seconds
Transferring time	95 seconds
Transferring time per second	About 1,260 cases per second

Table 8. Test result of transferring result materials

Contents	Result
Comparing updateresult materials	1.2 seconds (Comparing Server data and Client data)
Transferring update result materials	50 seconds (5MB · 100 cases)
Transferring time	250 seconds
Transferring time per second	10 MB per second

5. Conclusion

Information service environment is changing. More information needs to be gathered and provided to public to spur the usage of information, and NTIS has been

considering providing information integration containing result materials along with 422 metadata items. Thus, in this paper, a software prototype was designed and pilot tested. The software prototype was designed in the Agent Mode on the base of the information integration standard platform that is currently operating in NTIS.

The test environment is intentionally established in a poor condition for the purpose of reducing the error caused by the performance of the computers. The result of the test shows that the software prototype has capacity to transfer data including result materials in a timely manner.

Because the ultimate goal of designing the software prototypes is to provide applicable systems to the actual integration environment of NTIS and other institutes, some further studies are needed. First, system stability needs to be verified. Second, system quality such as data transmission speed needs to be ensured even when huge amount of result materials are intensively transferring.

6. Acknowledgment

This research was supported by “Maximizing the Value of National Science and Technology by Strengthening the Sharing / Collaboration of National R&D Information” project, funded by Korea Institute of Science and Technology Information (KISTI).

7. References

1. Kim Y. A study on government 3.0 in the Era of the Right to Know. *Journal of Korean Society of Archives and Records Management*. Nov 2014; 14(4):37-62.
2. Jang B. A study on the government full-text information disclosure system through the survey on the government officials' perceptions. *Journal of Korean Library and Information Science and Society*. Mar 2016; 47(1):339-60.
3. Cho J. A Study on information integration for enhancing the usability of science and technology information. *Journal of Korean Institute of Information Technology*. Oct 2009; 7 (5):263-72.
4. Bayardo R, Bohrer W, Brice A, Cichock A, Fowler J, Helal A, Kashyap V, Ksiezzyk T, Martin G, Nodine M, Rashid M, Rusinkiewicz M, Shea R, Unnikrishnan C, Unruh A, Woelk D. *InfoSleuth: Agent-based semantic integration of information in open and dynamic environments*. *ACM SIGMOD Record*. Jun 1997; 26(2):195-206.

5. Gil-Garcia J. Towards a smart state? Inter-agency collaboration, information integration, and beyond. *Information Polity*. 2012; 17(3):269-80.
6. Yim J, Cho E. A study on data adjustment and quality enhancement method for public administrative dataset records in the transfer process-based on the experiences of data warehouses' ETT. *The Korean Journal of Archival Studies*. Jul 2010; (25):91-129.
7. Shon K, Lim J. Research quality improvement by factors analysis data quality problem: Focus on national R&D information linking structure. *The Journal of Korea Contents Association*. 2009; 9(1):1-14.
8. Verma A, Kaur I, Singh I. Comparative analysis of data mining tools and techniques for information retrieval. *Indian Journal of Science and Technology*. Mar 2016; 9(11):1-16.
9. Kang N, Park M, Choi K, Kim T, Joo W, Kwon O. A development of service model for mapping the ecology of scientific research using national science & technology information service. *Indian Journal of Science and Technology*. 2015; 8(S1):121-30.
10. Kim M, Shin D, Kim J, Lee B. An empirical analysis of technology transfer of national R&D projects in South Korea. *Advances in Multimedia*. Jan 2015; 8(S1):121-30.
11. National Science and Technology Knowledge and Information Service. Available from: <http://www.ntis.go.kr>. Date Accessed: 01/08/2016.
12. Lee S, Lee B, Jeong H. A study on the problem analysis and improvement plan of the data quality management system of national R&D Data. *Indian Journal of Science and Technology*. 2015; 8(23):1-8.
13. Vassiliadis P, Skiadopoulos S, Simitsis A. Conceptual modeling for ETL Processes. *Proceedings of the 5th ACM international workshop on Data Warehousing and OLAP*. 2002. p. 14-21.
14. Prevedello L, Andriole K, Hanson R, Kelly P, Khorasani R. Business intelligence tools for radiology: Creating a prototype model using open-source tools. *Journal of Digital Imaging*. 2010; 23(2):133-41.
15. Kim D, Choi J, Jung H. A study on the information integration model based on standard integration framework. *Journal of the Korea Institute of Information and Communication Engineering*. Apr 2014; 18 (4):861-6.
16. Choi H, Kim J. Design of a NTIS information integration model based on a standard integration platform. *Journal of KISSE: Computing Practices and Letters*. Jun 2012; 18(6):484-8.
17. Ministry of Science, ICT and Future Planning. Available from: <http://english.msip.go.kr/english/main/main.do>. Date Accessed: 2016 Feb.
18. Ministry of Science, ICT and Future Planning.[Government 3.0], MSIP Disclosure of Information on National R&D and Construction of Integrated Monitoring System regarding Research Funds. Press Release; Jan 2016.
19. Information Disclosure Portal. Available from: <http://www.open.go.kr>. Date Accessed: 08/08/2016.