*International Journal of Engineering & Technology*, 7 (2.12) (2018) 281-285

International Journal of Engineering & Technology

*Website: www.sciencepubco.com/index.php/IJET*

***Research Paper***

# Visualization and the development of software connection technique to enhance the usability of science data

**Eunji Yu [1] *, Wonkyun Joo [1], Kiseok Choi [1]**

*[1] NTIS Center, KISTI, 245, Daehak-ro, Yuseong-gu, Daejeon, 34141, Korea*
**Corresponding author E-mail: Eunjiyu08@kisti.re.kr*

**Abstract**

It would be necessary in the future to provide service on not only national R&D information but also science data, but it would be necessary to effectively provide the meta information of currently owned science data before providing primitive data on science data. In this study, using visualization technique, this study will propose a science data map service model to visualize the information of science data into a transnational science data map, as well as to easily acquire the meta information of desired science data. Also, in order to enhance the usability of acquired information, this study will also suggest a way to immediately use acquired information by interworking with various simulation software. Through the service developed in this study, researchers have been able to acquire science data and usable simulation information without accessing different services.

***Keywords***: *Government R&D; Visualization; Knowledge Visualization; Liking; Science Data; Experimental Data; Simulation SW; Open Science*

## 1. Introduction

Open science has been accelerating faster since its projects were discussed as key assignments during the OECD Scientific Technology Ministry Summit in 2015. Open science is a digitalized form of researches performed with public fund and includes all efforts to provide them widely to the science community, business field, and society. Sharing research achievements among nations and international collaboration research transcending borders have led to the emergence of new issues solving global issues such as climate change and infectious diseases. Open science policy, which develops around open access, open data, and open cooperation claims to disclose and reveal the procedure, outputs, results of public research which is invested by the government's public funds[1,2]. In the past, disclosing science data that includes experimental data generated during research was difficult in a domestic research site due to institutional and cultural differences, and the status of generating and managing data in actual public institutions was not even identified either. In response, this study performed a huge current status survey on the science data of the government, research field, and academia of the past 3 years (2014~2016), and could figure out the current status on the science data based on the results. Now it is necessary to find a way to effectively provide these to users.

In Korea, National Science & Technology Information Service (NTIS)[3] was established in regards to providing national R&D information. Through these, the information of projects, tasks, and achievements performed with the government's R&D budget are collected from 17 departments and agencies to manage and disclose them. It would be necessary in the future to provide service on not only national R&D information but also science data, but it would be necessary to effectively provide the meta information of currently owned science data before providing primitive data on science data. The first access to increase the usability of science data starts from the visualization of data. Currently, with the growth of the big data technology market and open data policy, visualization technology has become an important element. While traditional visualization technology shows statistical information of data in graphs, the methodological element of visualization that helps summarize data to be looked at a glance is becoming more important in big data visualization[4]. Using such visualization technique, this study will propose a science data map service model to visualize the information of science data into a transnational science data map, as well as to easily acquire the meta information of desired science data. Also, in order to enhance the usability of acquired information, this study will also suggest a way to immediately use acquired information by interworking with various simulation software. In the end, this study will propose how to interwork with related software models to visualize, use, and provide science data through the NTIS in order to increase the usability of science data.

This paper consists as follows: Chapter 2 will examine the open data portal and data visualization service in NTIS, Korea, and abroad as related studies, and Chapter 3 will briefly describe the overview of science data map, which is the final output of this study. Next, Chapter 4 will explain the design and establishment of science data map. Finally, Chapter 5 will summarize the achievements and limitations of this study.

## 2. Related studies

### 2.1. National science & technology information service (NTIS)

NTIS comprehensively provides the national R&D information of government R&D projects from their planning, achievement, and usage in connection with 17 departments and agencies. It was established to increase the investment efficiency of research, production, and achievements through the joint use of information. By connecting with not only 5.1 million types of government

R&D information but also the National Discovery for Science Library (NDSL) and 16 policy services, NTIS services 85 million cases of national science technology information including papers from the private sector, patents, and policy/technology trend, as shown in Figure 1.

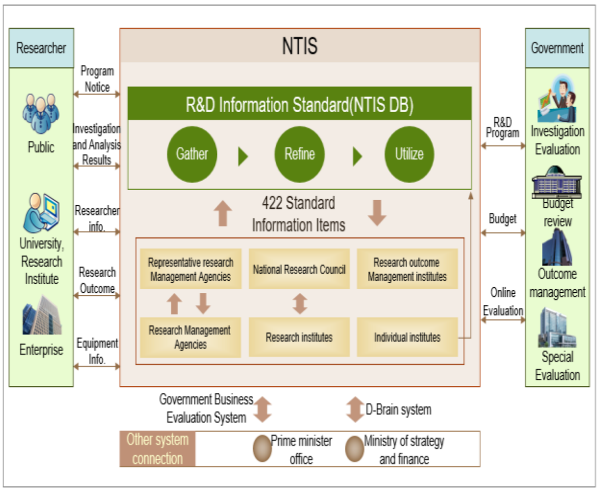


**Fig. 1:** Nits Conceptual Diagram.

### 2.2. Open data portal in Korea and abroad

With the development of big data technology, the range of data producers has been expanded largely into the generation and management of data by individuals and institutions, not only large organizations such as the National Weather Service, Ministry of Land, Infrastructure, and Transport, and the Ministry of Health and Welfare. With such change, the demand for revealing rapidly increasing public data has gradually increased and as a result, public data open policy has been actively carried out in Korea and abroad to meet such demand. For example, Korea provides 20,805 datasets in 16 categories such as the public administration through the public data portal (data.go.kr) for its citizens to more easily share and use various data owned by the government. The U.S., an advanced nation, provides 195,384 datasets of 14 categories through its public data portal (data.gov), and Mckinsey announced in 2013 that expected effects of the U.S.'s open data generates a value of 3 trillion dollars every year worldwide. The U.K. provides 52,167 datasets in 12 categories, while servicing information to apply the open data principle by disclosing evaluation results based on openness score, according to the 5-step development direction of open data proposed by Tim Berners-Lee, the founder of WWW (World Wide Web). In addition, the U.K. expected to earn economic effects of 1.8 billion pounds directly and 6.8 billion pounds indirectly through open data. Australia provides 28,204 datasets in 34 categories.

### 2.3. Data visualization

First, visualization research is divided into data visualization, information visualization, and intelligence visualization[5]. Information visualization means finding significance from information to effectively deliver information to users, by using graphical elements. Next, the purpose of intelligence visualization is to make smooth exchanges among users by using various visualization methods and delivering facts. In contrast to data which is a simple list, information is data with meanings and follows classification and summary methods. Intelligence is valuable information created by judgment, experiment, and rules and thus uses methods such as comparison and correlation[6,7]. According to several empirical research, the excellence of visualization has been proven in terms of relation expression rather than language and sequential expression, pattern identification, expression of overview and detailed information at the same time, and delivering different types of intelligenc[7-10]. The science data map proposed in this paper uses classification and summary of information visualization and the correlation of intelligence visualization so that users can summarize and use the science data drawn from national R&D projects. Details will be discussed in Chapter 4. Design.

## 3. Choosing the targets of science data management and use

### 3.1. Managing and sharing science data

(A) Distributed Type (B) Centralized Type.
Science data means a large volume of data which is difficult to manage and analyze with the existing data processing technology, among data generated through the process and results of science technology. From national R&D perspective, it means not only result-centered information of national R&D projects but also experiment data and intermediate outputs, which are produced through national R&D activities. Science data are various complicated forms of data such as particle accelerator, electron microscope, and dielectric, which require very highly specialized intelligence and complicated process for usage and analysis. In addition, with the 4th industrial revolution, science data enables new scientific discovery and can contribute to solving global issues and pending national agendas. Such science data is managed in 2 forms as in Figure 2. While distribution type allows immediate response to research sites through specialized data management by each field, it is not easy to have uniform data management and service as the subjects of management are dispersed. In contrast, centralized form is easier to build and more efficient to manage due to the unification of infrastructure, but data and users are concentrated into the center which can cause issues of excessive cost and data ownership.


**Fig. 2:** National Science Data Service Establishment Plan.

### 3.2. Targets of science data usage and service items

To effectively provide science data, accurate survey of current status and field selection must precede on the domestic science data. For 3 years from 2014 to 2016, a huge current status survey was conducted and 3 fields were selected as services as their data usage was judged to be significant. The fields were divided into government-affiliated institutions, institutions of large facilities/equipment, and institutions owning simulation software.
It was then identified that 17 professional representative departments that were classified as government-affiliated institutions, institutions in charge of 9 key results, government-funded research institutes, and universities owned 328 types of experiment data on climate and medical/ healthcare. Also, institutions owning large research facilities/equipment had 1,111 types of information.
Large research facilities/equipment mean large-scale research facilities that have excellent performance in the science technology field and are used in high-tech research field. As of 2017, there are currently 127 types of large research facilities built with more than 5 billion KRW of national R&D budget. In accordance with related laws, the Korea Basic Science Institute (KBSI) was designated as an exclusive institution for the management of research facility/equipment field[11]. According to the survey results of KBSI current status performed in 2016, there were 1,111 types of experiment data managed by 27 institutions that owned large research facilities/equipment and the number of data owned was 180,000, which was 90TB in total.
EDISO, which was classified as an institution owning simulation software, is an open-type computational science platform providing a simulation program usage environment for the education and research of natural science/engineering based on simulation software and education contents; it was developed by experts in 5 fields of computational science and engineering. It owns 391 types of simulation SW and 55 educational contents including 97 types of computer thermodynamics, 108 types of nanophysics, 43 types of structural kinetics, 27 types of computer design, and 22 types of computer medical science. The items for science data service based on the above data include the names of institutions, information system, information system overview, name, classification, and the number of data, fixed/non-fixed categorization, data management method, data ownership, disclosure/openness, form of provision, range of usage, and meta information on information system manager and operator.

## 4. Development of science data map design for using science data

### 4.1. Overview of science data for visualization

To enhance the usability of science data, it is necessary to build a service to comprehensively identify the current status such as the data list, producer, and location of science technology and easily search them in two ways. To fulfill these roles, a science data map closely integrates and provides the database of national science technology, data generated by large research facilities and equipment, and simulation SW information around visualization functions, as shown in Figure 3.

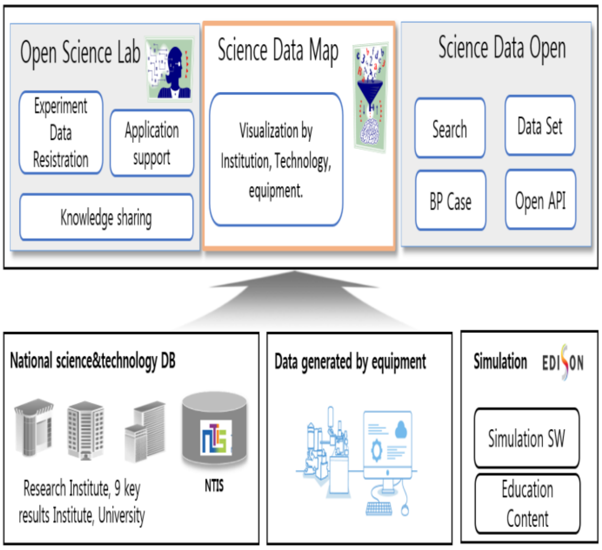


**Fig. 3:** Science Data Map Overview.

### 4.2. Design of science data map

The information for science data map defined in the survey results of the current status of science data in this study totaled 92,847,280 cases of 1,439 types. In order to enable users to understand and acquire the collected science data, this study applied the classification system for each institution, 6T, science technology standard, and large research facilities/equipment, and applied an information visualization method in a map form. For easier access, institutions were classified into departments/representative specialized institutions, result institutions, and government-funded research institutes. Classification for 6T was divided into IT, BT, NT, ST, ET, CT, and others to make more convenient access to each type. The classification of science technology standard consists of 33 large, 369 medium-sized, and 2,899 small classifications, and the design of science data map was mainly based on 33 large classifications for convenient accessibility. The classification system for large research facilities/equipment enabled the provision of information by owned institutions, and the overall interface enabled finding information by using navigation from a tree structure for easier access.
Next, the study proposed how to interconnect with various simulation softwares to immediately use the acquired information. First, it extracted the keyword information in science data and the explanation of EDISON simulation software, used the correlation of intelligence visualization based on the extracted keyword information and classification information, and drew the correlation between science data and EDISON simulation software. After, the study mapped the correlation into usable simulation software for each science data. Since the EDISON simulation SW is a professional field of natural science and engineering, experts in 5 categories chose and reviewed mapping information in which the correlation method of intelligence visualization was applied, in order to develop the final mapping table. hrough this, users who separately searched science data and simulation software before can now easily acquire desired science data and receive usable simulation SW at the same time.

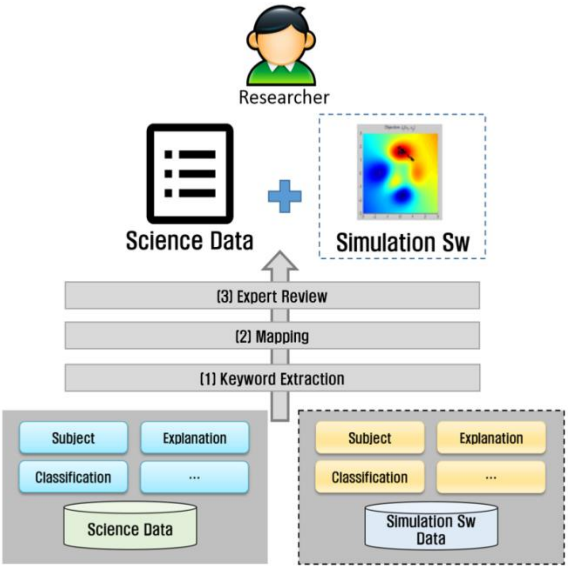


**Fig. 4:** Link Scientific Data with Simulation SW.

### 4.3. Development of Science data map

The main screen of science data map that provides a comprehensive current status of science data and the information of using EDISON simulation SW on science data is as in (Figure 4). Users can first check the information of the current total amount of science data amount, including experiment data, on the main screen. The main screen applies a system that can classify science technology data from various perspectives, by institutions, 6T, standard classification of science technology, and institutions owning large research facilities/equipment, and provide them in a visualized form as in Figure 5. Users can also acquire information for simulation from the science data map as it even interworks and provides usable simulation SW in each science data.

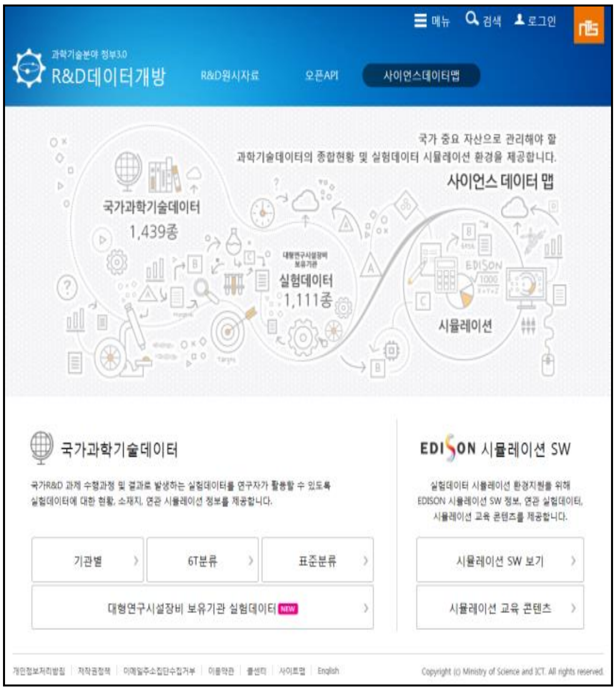



**Fig. 5**: Science Data Map Service Page.

## 5. Conclusion

The current status of which science data are being generated and managed by each public institution has not been identified for the data of scientific technology experiment; thus, researchers who wished to use science data had to separately access each institution and service. This study built a science data map by systematically organizing a process to visualize national R&D information in the NTIS as well as current information, based on the results of current status survey, and to provide usable EDSION simulation SW with science data. Through the service developed in this study, researchers have been able to acquire science data and usable simulation information without accessing different services. Future researches will devise a plan to provide large-capacity data storage and virtual collaboration environment, besides a simulation environment, at the same time, and perform a user satisfaction survey on developed services to measure service usability.

## Acknowledgment

This research was supported by Construction of NTIS funded by the Ministry of Science, ICT and Future Planning.

## References

[1] Eunjeong S, Open Science OECD Trends and Implications. Science and Technology Policy Institute, 2015, 22(4), pp.1-21. https://www.dbpia.co.kr/SearchResult/TopSearch?isFullText=0&searchAll=%EC%98%A4%ED%94%88+%EC%82%AC%EC%9D%B4%EC%96%B8%EC%8A%A4%28Open+Science%29%EC%97%90+%EA%B4%80%ED%95%9C+OECD+%EB%85%BC%EC%9D%98+%EB%8F%99%ED%96%A5%EA%B3%BC+%EC%8B%9C%EC%82%AC%EC%A0%90

[2] Eunjeong S, hyungjun A, Wonkyo J, Recent Trends and Challenges in Open Science Policy: Focusing on Control Mechanisms and Legal Rights, Korea Science and Technology, 2016, pp.114-128.

[3] National Science & Technology Information Service, http://www.ntis.go.kr

[4] Gwangseon C, Youngkyung H, Seonho L, Big data Visualization. Korean society of computer and information, 2013, 20(1), pp. 33–43.

[5] Zeiller M, A Case Study Based Approach to Knowledge Visualization. Proceedings of the Ninth International Conference on Information Visualization, 2005.

[6] Seokhyun J, Jooyup L, Kyungwon L, Study of Representation Methodology by Comparative Analysis between Information Visualization and Knowledge Visualization, 2009, 8(2), pp. 1242–1248.

[7] Kyungwon L, Kyungwon L, Jooyup L, A Study of Information Design Methodology by Comparative Analysis between Information

Visualization and Knowledge Visualization, Society of Design Convergence, 2009, 8(2), pp. 19–30.
[8] Glenberg A, Langston M. Comprehension of Illustrated Text: Pictures Help to Build Mental Models. In Journal of Memory and Language, 1992, 31(2), pp. 129-151.
[9] Bauer M, Johnson-Laird P, How diagrams can improve reasoning. In Psychological Science, 1993, 4(6), pp. 372-378.
[10] Burkhard R.A., Learning from architects: the difference between knowledge visualization and information visualization. Proceedings of Eighth International Conference on Information Visualization, 2004.
[11] WonKyun J, MyungSeok Y Eunji Y, Yunjeon K, Seonghee L, Kiseok C, Utilization of Research Data Derived from Large Research Facilities and Equipments for the Data Driven Open Science. The Institute of Electronic Englneers of Korea, 2016, 11, pp.776-777.