ESA/STAT/AC. 223/S5.7

Third International Seminar on Early Warning and Business Cycle Indicators

17 – 19 November 2010
Moscow, Russian Federation

Development of National Central Data Hub: The Case of Korea

Chanil Seo
Statistics Korea

3rd International Seminar on Early Warning and Business Cycle Monitoring
(Moscow Russia, 17-19 November 2010)

Topic (ii): Developing Data Hubs

Development of National Central Data Hub: the case of Korea

Paper Prepared by Chanil SEO (Mr),
Director, Statistical Information Portal Div, Statistics Korea

I. Introduction

1. The Republic of Korea has been adopting decentralized statistical system in which the KOSTAT, a central statistical organization, and many other statistical agencies produce statistics. Currently, approximately 370 designated organizations produce about 840 different kinds of statistics. Therefore, it was hard to share information among agencies, and data were incompatible due to their non-standardized information systems and database systems. Statistical users experienced inconvenience and confusion because they should know which system each organization operates and where the information is stored. To ease these inconveniences and improve data sharing through standardization, the KOSTAT promoted 'the Integrated DB Project for National Statistics' from 2006 to 2009. The KOSTAT is providing one-stop statistical services to users by standardizing and integrating data produced by all the statistical organizations.

2. Currently, the KOSTAT provides the integrated data from 113 organizations. This paper presents the background, contents, achievements and development process of the Integrated DB and Portal Service Project for Korea National Statistics.

II. Portal service system for National Statistics

A. Status and Problems in 2005

3. In Korea more than 840 official statistics are produced by about 370 statistical agencies. Each organization compiles and manages statistical data in their own way. Some organizations maintain their statistical DB systematically. Meanwhile, some provide data in simple forms such as Microsoft Excel and Words files. Others even don't provide information to the public. Therefore, statistical users feel great difficulties in finding what they need.

4. In the 2005 survey, it was revealed that human resources and budget in statistical organizations were quite insufficient. Only 53 organizations had exclusive teams and personnel for statistical work. The average number of personnel in charge of statistics in these 53 organizations was 1.8. In the other organizations only one person was in charge of not only statistics but also other work. Developing a statistical DB and standardization process was sluggish. 40 percent of designated statistical organizations did not build their statistics DB those days. Data sharing was difficult due to non-standardization in terms of DB structure, data management and application systems. Also, data were incompatible on account of various types of DB systems.

5. To take proper measures against these circumstances, the KOSTAT decided to make one-stop portal service for national statistics by integrating major designated statistics into the DB.

B. Objectives and strategies

6. The integrated DB for national statistics aims at increasing efficiency under the vision of "providing user-oriented, high-quality statistical services" by integrating major statistical information from selected organizations and providing one-stop service. So as to accomplish these aims, the KOSTAT has developed the integrated DB for national

statistics and one-stop portal service for national statistics, sharing national statistics and revising relevant laws and regulations.

7. The KOSTAT integrates and links the statistics DB from the organizations, provides one-stop service on the statistical portal site, provides various statistical contents such as advanced analysis, provides customized statistical service to various users, shares the integrated DB with organizations, develops and distributes common service modules with cutting-edge web service technology, revises the legal system such as the Statistics Act, sets relevant regulations like a guideline to DB management and lays the foundation for the cooperation among the organizations.

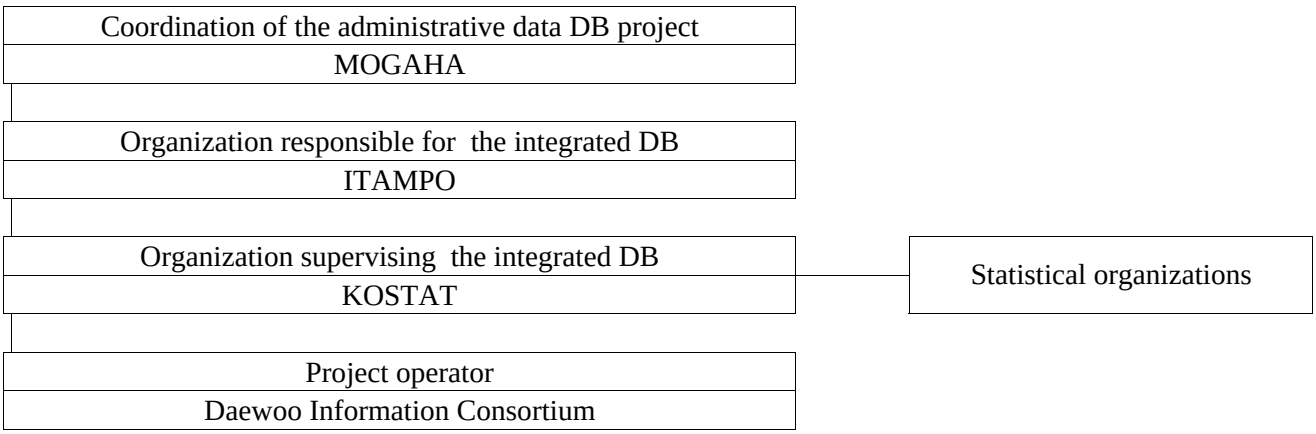
8. Project strategies include building the statistics DB for designated statistical organizations step-by-step as well as developing the statistics DB system for organizations that don't have their own DB system. Workshops for statistical personnel were held to strengthen ties among statistical organizations. In the workshops, the project plans were introduced. The KOSTAT asked statistical organizations to cooperate for the integrated DB and accepted their opinions on the improvement of the system.

9. Since this project required close relationships among organizations, the KOSTAT composed a working-level committee for the Integrated DB, which monitored the structure of the integrated DB and the portal service system, and shared the experience of system development among organizations. In 2007 the KNSO staff visited 45 organizations to grasp their current status; whether organizations built the DB, whether they could provide data to the public, and which type they managed data (DB, printed publications, files, etc.).

10. In the meantime, "the Statistics Act" was revised to add an article "Building the statistical DB". According to the Act, statistical organizations should build the statistical DB and provide relevant data for building, linking and integrating the statistical DB if necessary.

11. The KOSTAT presented a guideline for standard codes. Currently 225 kinds of standard codes are being used for KOSIS database

12. The 4-year Integrated DB Project for National Statistics started in 2006. The KOSTAT implemented the '1st stage of the Integrated DB for National Statistics' from April 6, 2006 to November 15, 2006 as part of '2006 Administrative Data DB Project' promoted by the Ministry of Government Administration and Home Affairs ('MOGAHA'). Designated statistics from 40 organizations were integrated into the DB and a portal service system was developed. This portal service has been provided to the public since July 2007. After 2007, the KOSTAT implemented the '2nd stage of the Integrated DB for National Statistics' at which statistical information from 73 designated organizations was added to the integrated DB. Organizations and their roles for the Integrated DB Project are shown in [Figure 1] and <Table 1>:



[Figure 1] Organization of the Integrated DB Project for National Statistics

Organization	Role
MOGAHA	- Plan and supervise the administrative data DB project
ITAMPO	- Sign a contract (in cooperation with the KOSTAT) - Support a project operator - Manage the project and provide technical support /select and support an audit organization - Execute the budgeting and accounting
KOSTAT	- Supervise the system development and DB building - Select a contractor and sign a contract (in cooperation with the ITAMPO) - Manage, supervise and monitor the project - Support auditory activities and inspect the project
Statistical Organizations	- Provide opinions and requirements for system development - Cooperate for the Integrated DB Project - Provide statistical data and participate in building the integrated DB

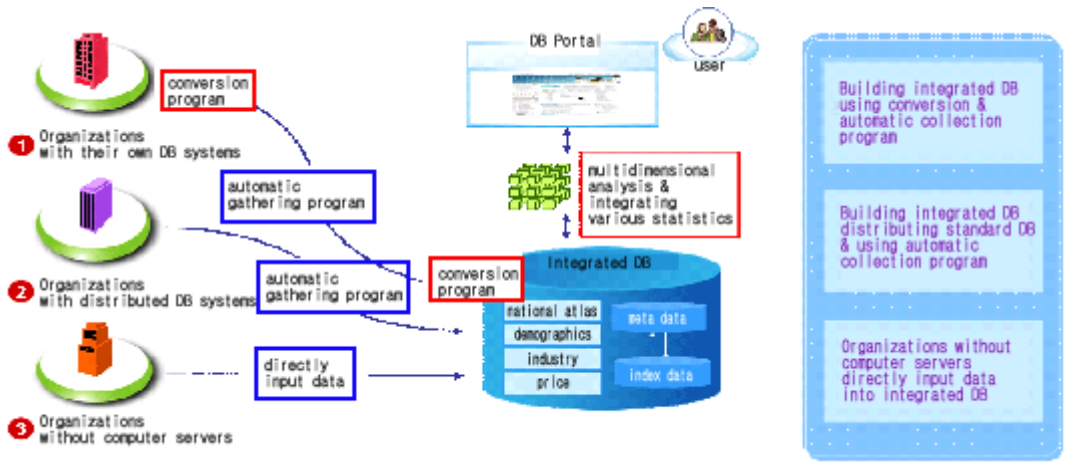
<Table 1> Roles of Organizations

C. Developing the National Statistics Portal Service

13. The objective is to design the most suitable model for effectively integrating all statistical data from organizations. So the Integrated DB is designed to effectively store and operate standardized classification and item codes and makes it easy to access to the DB through meta data such as information about statistical organizations, statistical classification and users. Also, by accepting the standardization trend in the international society, the Integrated DB effectively supports data retrieval and analysis.

14. The KOSTAT interviewed statistical organizations in person for 5 weeks (August to September 2005) to grasp their status such as statistics production and statistical DB, and to obtain opinions on the method to integrate and link the statistical DB. They said that it was hard for organizations to exchange and share information due to non-standardization of DB and systems. More than 90 percent of respondents hoped to build the Integrated DB.

15. Figure 2 is a schematic diagram which shows how to automatically integrate statistical data from organizations into the Integrated DB. By considering the status of the organizations, they were classified into 3 groups: (i) organizations with their own statistical DB, (ii) organizations with the statistical DB provided by the KNSO, and (iii) organizations without computer servers for statistical DB services.



[Figure 2] Schematic Diagram for the DB Integration Project

16. Two kinds of modules were designed to transmit statistical data to the Integrated DB – one for data extraction and the other for data transmission. For data security, data should be encrypted and then transmitted through exclusive transmission ports by means of sockets.

Data transmission methods vary among statistical agencies

O Agencies using their own server and own database
- Extracted data are converted to XML file formats.
- These files are compressed and encrypted to be transmitted by means of sockets(SOAP).

O Agencies using their own server and KOSTAT database system
- statistical data to be transmitted are converted to XML files thru the KOSTAT data management system.
- XML files are compressed and encrypted to be transmitted by means of sockets (SOAP).

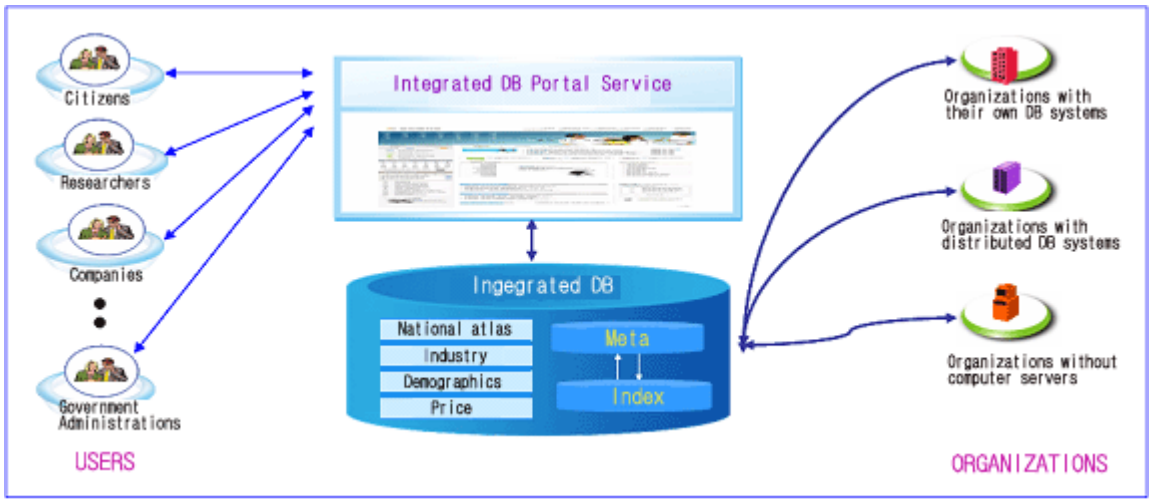
O Agencies using the KOSTAT server and KOSTAT database system
- The KOSTAT provides a hosting server.
- Therefore, without data conversion and socket use, data are transmitted by means of DB links.

17. For organizations with their own DB and those with the DB provided by the KOSTAT, their DB was converted to the standardized DB. And a program for the automatic data collection was developed and applied. For organizations without a statistical DB, the KOSTAT provided the standardized system for building the DB and automatically collecting statistical data free of charge.

18. In 2006, 40 organizations such as the KOSTAT, Bank of Korea, the Ministry of Labour and the Ministry of Health & Welfare were selected for the integration of statistical data. Their 246 different kinds of statistics were converted and inputted into the Integrated DB.

19. In 2007 approximately 90 kinds of statistics from 45 organizations were integrated into the DB for national statistics. In 2008~2009, all of the designated statistics from the other organizations was integrated into the DB. In particular, newly approved national statistics will be integrated into the DB after interviewing organizations in person.

20. The KOSTAT developed a portal service system so that users may obtain statistical data they need just by accessing to a portal site. And the integrated DB project will be completed in 2009. Currently, data from 113 organizations are being provided, which allows users to obtain data from all statistical organizations.



[Figure 3] Portal Service System

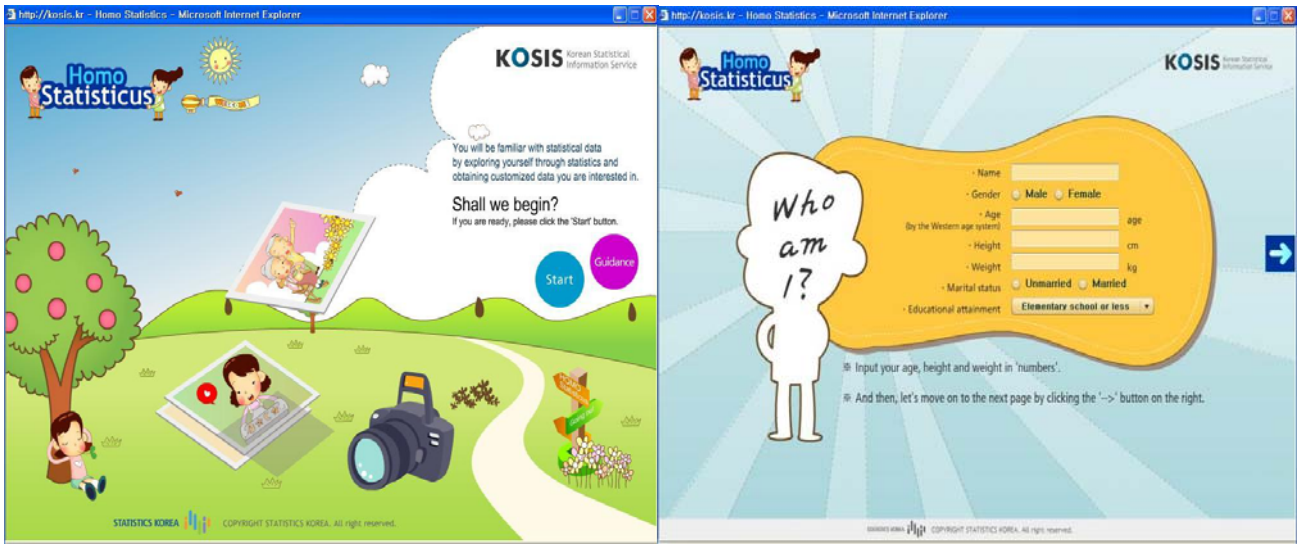
21. Figure 3 shows a portal service system which enables users to retrieve data easily by utilizing specialized tools such as X-Internet and OLAP. This portal service system for national statistics was developed in 2006. [Figure 4] represents the homepage of the Korea Statistical Information Service (KOSIS: www.kosis.kr). Statistical users may retrieve data in various ways e.g. data search by topic, by survey and by organization, integrated index search, and quick search. And 100 major statistical indicators are also presented.

[Figure 4] Homepage of KOSIS

22. This Web-based portal service system automatically converts and collects existing data into the Integrated DB. Some organizations are applying common web service modules for data provision. The number of organizations applying common modules will increase.

23. The KOSTAT proposed a standard for web-based linking, a standardized structure of the integrated DB and a standard for meta data. To stabilize cooperation among organizations, the KOSTAT composed a special committee. The committee discuss these standards, decide the level of data provision, discuss data service policies, and discuss the method to publicize the Integrated DB system. The KOSTAT endeavoured to take the current status of each organization into consideration.

24. In the context of new trend of transforming information into Knowledge, the KOSTAT worked out KOSIS development plans in which statistical information will be customized, visualizes and transformed into knowledge. As a result the KOSTAT devised 'Homo Statisticus (Statistical Self-portrait'(figure 5)), which is a visualized story-telling program that help users to explore 'their statistical self' through statistical data in 2008. Also in 2009, stated the BCC(Business Cycle Clock) to visualize business cycles and in 2010, launched 'G20 Statistitics Dashboard'(figure 6) displaying visualized statistical data for easy and convenient comparison of financial and economic indicators of G20 countries



[Figure 5] Visualization Program - Homo Statisticus

[Figure 6] Visualization Program - G20 Statistics Dashboard

D. Difficulties encountered and measures taken

25. In the course of the project a lot of difficulties concerning data input, cooperation among organizations and standardization occurred. To minimize data input error during database integration stage, following quality checking procedures have been applied.

- Input Management Program for statistical table information (classification, item, etc)
- Intelligent checking like sums and ratio during data input
- Inputting the same data twice by different person and cross-checking
- Conducting inspection on the input data 2 times
- Checking errors by program while loading onto DB
- Conducting inspection on the loaded statistics data

26. The statistical organization has not been so cooperative regarding the implementation of database due to the lack of understanding in this project. Following countermeasures and strategies below have been on the way to secure cooperation.

- holding workshops for the statistical agencies to better understand the project
- forming and managing working-level committee of integrating database
- adding article to Statistics Law to mandate establishing statistical database and submitting statistical data

27. In the near future, SDMX will be adopted for international compatibility. Non-official statistics that the public want will be provided, too. And interactive communication with statistical users will be implemented by using ICT.