E-Census as a New Approach to the Population and Housing Census

1. Introduction

The Population and Housing Census of Korea has been carried out on a quinquennial basis since 1925. The results have played an important role in national and sub-national policies and plans for socioeconomic development, research, and business purposes. Statistically, the census provides a "base population" for the population projections for the whole country as well as for the provinces. It provides a "bench mark" and also functions as a sampling framework for sample surveys.

Despite its success, the census in Korea has been confronted with criticism including cost, huge amount of resources, burden on the public, increase in the workload of local government, and untimely data release. Among them, ever increasing costs in census-taking and untimely data release are the major issues. It is far from being feasible to claim that the census should be carried out merely for historical tradition. Without significant improvements in terms of cost and benefits, the criticism against the census and questions for the necessity of census-taking will persist.

To overcome the financial shortages of the 2005 Population and Housing Census, the Korea National Statistical Office (KNSO) recognized high level IT infrastructure in Korea and introduced the e-Census system to minimize the cost of census-taking and maximize efficiency of available resources. The e-Census system consists of functions such as survey management, human resource management, cyber education, Internet survey, housing registers information, post-enumeration survey, and web-based field data input and processing.

This paper aims at evaluating the e-Census system in the 2005 census and provides an insight into a cost-effective approach to the 2010 census by sharing the KNSO's e-Census experience derived from the 2005 census.

2. History of the e-Census system

The census is the largest survey consuming a huge amount of resources, and the census cost will continue to grow rapidly if ways of saving cost are not applied. The cost of the 1990 and 1995 census-taking was 2.6 times greater than the previous census-taking cost. The ever increasing cost of census-taking broadens financial resource gaps between ideal and actual census budgets. These funding constraints force the KNSO to seek cost-saving methods in every stage of census process ranging from planning to data dissemination. The census-taking cost mainly depends on the cost related to hiring enumerators because the cost for enumerators accounts for almost three-fourths of the total costs of the Korean census. Therefore, census planners have explored ways of reducing the number of enumerators to be paid. For the 2000 census, the possibility of introducing a self-enumeration method had been intensively researched to replace the traditional canvasser method, which requires a large number of paid enumerators. The self-enumeration method was partly adopted for an efficient 2000 census-taking.

Tables 1. Costs of the Korean census-taking

	Census year				
	1985	1990	1995	2000	2005
Total cost Million Won	8,174 (-)	20,956 (156%)	53,920 (157%)	83,387 (54.6%)	128,951 (54.6%)
Per person Won	202 (-)	483 (139%)	1,209 (150%)	1,807 (49.5%)	2,727 (50.9%)
Per household Won	854 (-)	1,846 (116%)	4,160 (125%)	5,794 (39.3%)	8,065 (39.2%)
Data collection Method	Interview	Interview	Interview	Interview + self-enumeration	Interview + self-enumeration + Internet

Note : ( ) indicates an increase rate

On the other hand, the deteriorating environment of census-taking put census planners in an awkward situation. With the growing awareness of privacy, more and more people had become reluctant to cooperate with surveys. Also, the increasing number of hard-to-enumerate households, such as one-person households and day-time absent households contributed to the changing statistical environment. In addition, the weakened statistical function of local governments through the downsizing program of the Korean governments threatened the success of the 2005 census-taking.

Korea has a business-friendly environment for Internet service. At the time, more than 80 percent of the Korean population lived in urban areas and apartments accounted for 53 percent of Korea's housing stock. Therefore, Internet service can be provided at a low price. Based on desirable surroundings for Internet service, the Korean society moved towards an information-oriented society and Korean's Internet user population has grown dramatically. At the end of 2003 Korea was third in the world in Internet service penetration rate (60%) with 29.2 million users after Iceland (68%) and Sweden (62%)[1].

For a low cost-high efficiency 2005 census and to overcome a deteriorating census-taking environment, the KNSO recognized high level IT infrastructure in Korea and introduced the e-Census. Regarding data collection methods, the traditional data collection methods (enumerator's face-to-face interviewing method and self-enumeration methods) remained as major data collection methods of the 2005 census. The KNSO provided Internet census forms for hard-to-contact households and those who preferred the Internet survey to protect privacy. The feasibility of the Internet survey was studied through pilot surveys on the basis of the following criteria: the efficiency of data entry, response rate, accuracy, and network capacity.

[1] Source: Ubiquitous Network Societies: The case of the Republic of Korea(2005), International Telecommunication Union

3. e-Census system of the 2005 Population and Housing Census

The e-Census system was developed under objectives such as plans corresponding to changes in survey environment, economic census with low cost and high efficiency, improvement of data quality, and shortening data release time. To meet the objectives, the e-Census system was organized into seven components including Survey Management, Enumerators Management, Internet Survey, Web-based Input, Editing, Post-Enumeration Survey, and Data Compiling/Analysis.

In terms of system hardware, the e-Census system was made up of server, storage, backup equipment, network equipment, and security equipment. Internet network systems of local governments were used for the e-Census system to reduce cost and enhance data security. For the safety of the network between local governments and the KNSO, a duplex network and security system requiring two sets of networks and security equipment were applied.

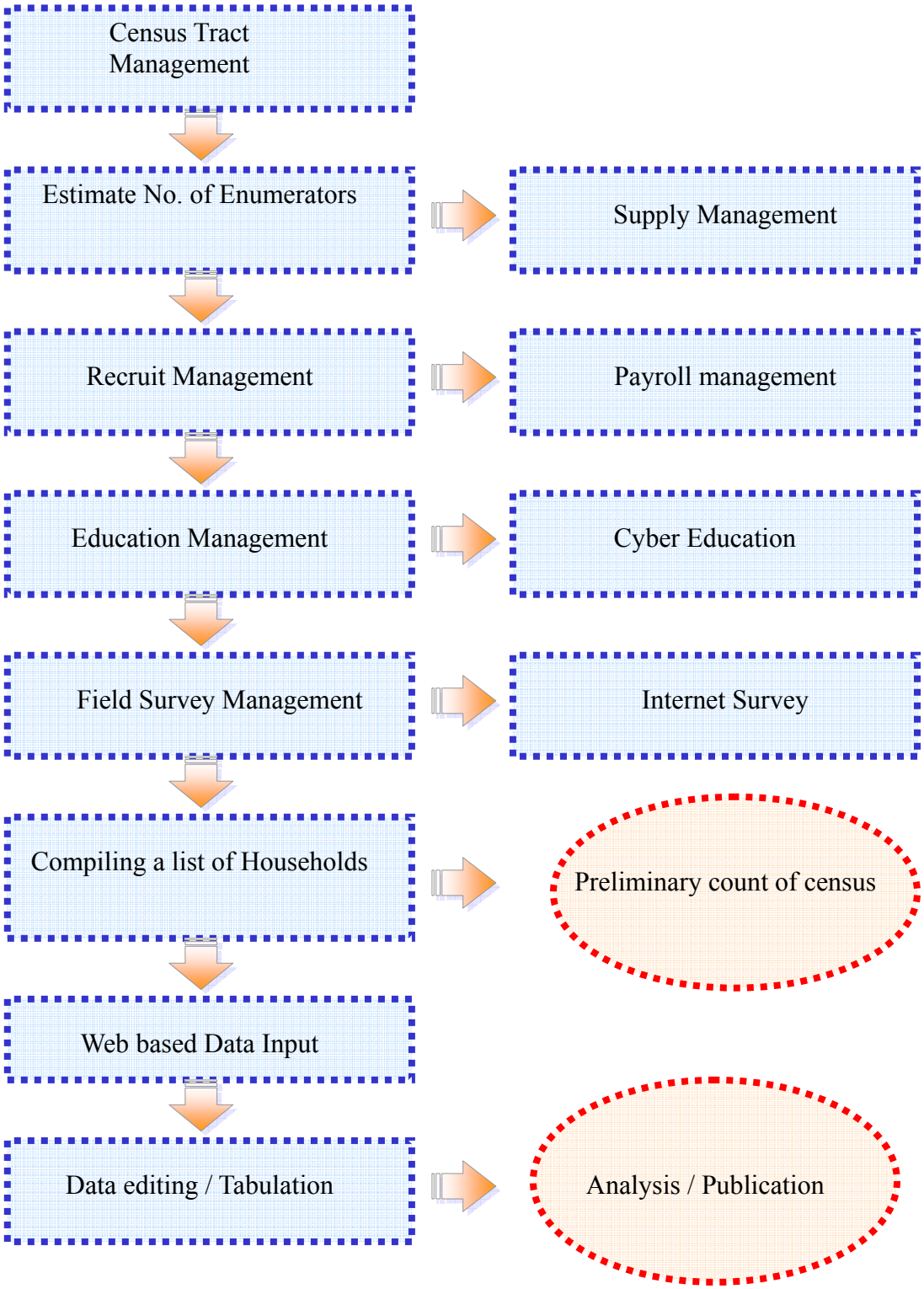
The Survey Management was comprised of sub-functions of census tract management, supply management, education management, cyber education, short messaging system to mobile phone, and survey result management (number of households surveyed/not-surveyed). Officers in local governments registered census tracts and reported modifications of a census tract through a census tract management function. Supply Management provides information about volume of survey supplies for local governments and transportation information from the KNSO to local governments. Request and supply for shortage of survey goods, eg. Census Questionnaire, enumerator's sheets, etc. was conducted through this site. Information for enumerator education such as location, participants, and instructors was provided in Education Management. Cyber education provided education programs for enumerators and 82.6% of enumerators participated in the cyber education prior to their off-line education.

Enumerator Management included sub-functions of recruitment, assignment of census tract to enumerators, payroll, and information of local officers. People who

wanted to participate in the 2005 census as an enumerator or supervisor could apply through this site. The number of enumerators by local governments and information whether employed or not could be obtained at the recruiting site. Assignment of census tracts to enumerators and calculation of enumerator's pay was also conducted.

The Web-based Input was adapted as a data input method for the 2005 census which carried out data capture "on the spot". This method enabled us to input enumerated data through an on-line process using the same format in all regions of the nation. Inputted data could be monitored by the KNSO at real time. This method contributed to shorten the data input period and improve data quality compared with the previous census by decentralizing data input and checking the relationship between answered items. In the traditional data processing systems, enumerators complete the census questionnaires in the survey fields, and questionnaires are sent to the headquarters or the local offices of the KNSO. The questionnaires are inputted using the Off-line Batch processing method.

Chart 1 : Flow Chart of the e-Census System

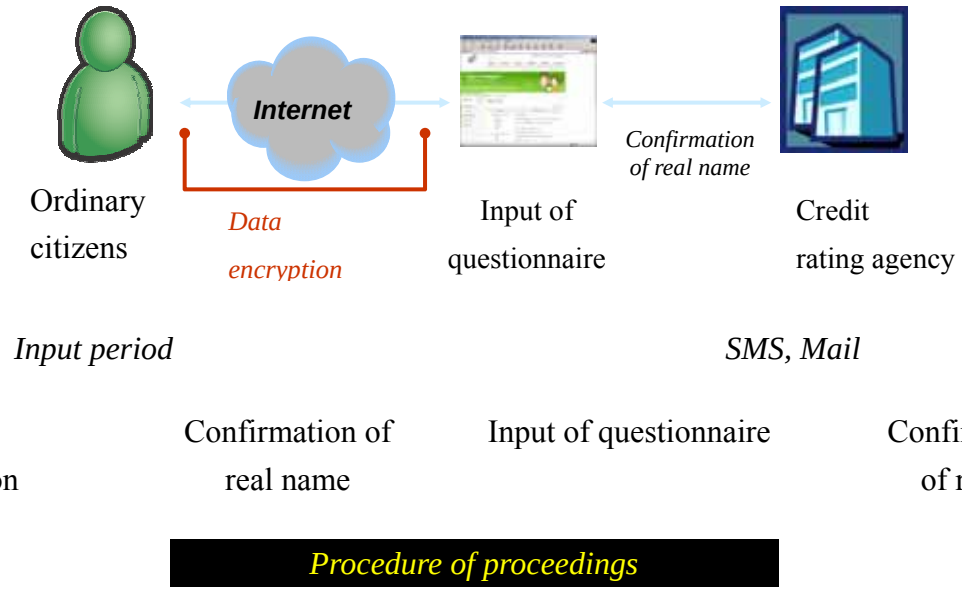


4. Internet Survey

The operation of the Internet survey is a four-step process. The first process is application. Respondents who want to complete the census questionnaire through the Internet must make an application. Second, the KNSO certifies his/her real name and personal identification number through a Credit Assessment Authority. Third, the applicant has to input the rest of the necessary information such as ID, password, telephone number, e-mail, address, etc. Finally, a short or long form questionnaire pops on the screen after matching an applicant's dwelling place with the address D/B. After that, the applicant can complete the questionnaire through the Internet directly.

The advantages of the Internet survey includes reduced coverage error compared with the previous census by providing a way for hard-to-enumerate households to participate, improved data quality by interactive user guidance and automatic filtering of irrelevant survey items, and cost reduction. While the Internet survey had several advantages, it also raised new challenges and problems that must be resolved in the next census. First, automatic address matching rates between address DB in the system and address respondents answered were 62.1%. The automatic matching rates of 62.1% were lower than the target rate of 80%. The 37.9% respondents who did not receive automatic address matching information had to wait until they received the number of census tract by the local government officer. Second, there were difficulties in estimating the appropriate capacity of the server and network system according to expected responses. The capacity of the system should be designed to accommodate peak time user frequencies over the survey period.

Chart 2: Flow Chart of Internet Survey



Households that completed the questionnaire using the Internet survey system totaled one hundred and forty thousand which accounted for 44.0% out of a target population of three hundred and twenty thousand. The participation rates were 0.9% of the Korean population. On the second day of the survey, 14.6% of Internet respondents completed the census questionnaire and 12.2% Internet respondents answered on the first day of the survey. The peak time was between 10-12 a.m. and 8-10 p.m. The average time for filling out the Internet survey form was approximately 14 minutes for the short-form questionnaire and 23 minutes for the long-form questionnaire. Respondents in their 30s accounted for a majority of the respondents, and persons who had received a higher education tended to respond more frequently through the Internet survey.

The KNSO conducted a follow-up survey to determine experiences and opinions among respondents concerning the Internet survey system. About 70% of respondents

replied that the Internet survey was convenient in terms of guideline, screen interface, and access. In terms of inconveniences, the number of survey items exceeded the respondent's expectations, waiting time to get the number of census tract, and unsatisfactory explanations for survey items were answered. Only 3% of respondents marked the speed of the Internet as an inconvenience.

5. Census Items

The application of sampling in the census is one way to save on the cost of the census. The long-form, which ask a full set of questions is addressed to sample households and the short-form, with a restricted set of questions, is addressed to the remaining households. If the full set of questions were asked to all households, the cost of the census would be enormous since more survey time is needed. The Korean census has adapted this sampling technique since the 1966 round and the sample size amounted to 10 percent of the total population.

In the Internet survey of the 2005 census, the percentages of sample household respondents were 11.3% of the total respondents, which was a little higher than those of the census sample size. It suggests that the number of survey items did not affect the Internet survey response rates. Therefore, in order to save costs and also lighten the burden on respondents, this sampling method will be adapted for future Korean census, even if the Internet survey is adapted as well.

A census conducted in the years ending in 'zero' is a regular census that surveys more items than a simplified census conducted in the years ending in 'five'. In the 2005 census, 21 items were asked on the short-form and 23 items were asked on the long-form. To find out the needs of the 2005 census items, Sample surveys directed toward users, such as policy makers and researchers, were conducted in 2003 and 2004.

Tables 2. Number of Items in the Korean Census

	1970	1975	1980	1985	1990	1995	2000	2005
Total	31	31	45	30	44	28	50	44
Short-form	15	11	26	30	32	17	20	21
Long-form	16	20	19		12	11	30	23
Population	17	22	25	16	21	16	29	24
Household & Housing	14	9	20	14	23	12	21	17
Items for city and province								3

Among the requested census items by users, questions that were not appropriate were excluded including items that were sensitive to the public, placed a great demand on a person's memory, or could be seen as an intrusion on privacy. To increase the cooperation and also to correspond the need of new census items from local governments, three long-form items were selected by province and metro-city and those three items were surveyed only within their boundary of local governments.

Compared to the 2000 Census, several items and response categories were added to the 2005 census. New items and response categories are as follows:

- Items related to low-fertility: year and date of marriage, number of children desired
- Items related to quality of living conditions: floor number on which household is located, total number of building floors, heating facilities, housing ownership
- Items related to social welfare and others: disability, place of work, dispersed families between North and South Korea, religion.

6. General Directions for the 2010 e-census of Korea

The cost of the 2010 census will be close to 200 billion won according to an estimation based on increase rates of 54. 6% between 2000 and 2005, but it is uncertain that the KNSO will be able to secure the necessary funding. As a consequence, the KNSO must look for ways of cutting costs. When compared to the Census 2005, the field survey environment of the census has further deteriorated with the growing awareness of privacy and hard-to-enumerate households.

The IT infrastructure of Korea will get better and better. Therefore, expansion of the e-Census as a cost-effective approach to the Census will be a major alternative to correspond with the changes of census environment of the 2010 census. Through the expansion of e-Census system, the goal of a cost-effective 2010 census can be achieved because the e-Census covers all processes of the census and contains opportunities for cost savings in all of its operations. The possibility of using more register information such as population register and health insurance data will be studied through pilot surveys scheduled from 2007.

Appendix: Outline of the 2005 Census of Korea

1. Historical Background

○ The Population Census has been conducted on a five-year basis since 1925. In addition the Housing Census has been carried out every five years since 1960.

○ In 1949, the Population Census was conducted for the first time by the Korean Government following the establishment of the Korean Government in 1948.

○ The 2005 Population Census is in its 17th round, and the 2005 Housing Census is in its 9th round.

2. Legal Basis

○ The 2005 Census was designated as approved statistics by the Article 4 (1) and Article 8 of the Statistical Act.

- The Population Census : Designated Statistics No. 10101
- The Housing Census : Designated Statistics No. 10102

○ The regulations of the Population and Housing Census were established on July 1, 2000 (Decree of the Ministry of Finance and Economy No. 143)

- The first revision occurred in 2005 providing a legal basis for an Internet survey

3. Census day : November 1, 2005

4. Coverage

○ The Census covers all Koreans and foreigners, and their housing units within the scope of the administrative jurisdiction of the Republic of Korea as of November 1, 2005.

5. Census Period : November 1 ~ November 15 (15 days)

6. Organization of the Census

○ Supervising agency : KNSO

○ Fieldwork organization : local governments

7. Enumeration Methods : Mixed mode

○ Direct interviews by enumerators, self-enumerations, and Internet survey

8. Census Items : 44 items

○ Complete enumeration : 21 items

○ Sample enumeration : 23 items (3 special items designed for Metropolitan cities & provinces are included.)

9. Human Resources : 110,000 persons

10. 2005 Budget : 129 billion won

11. Main Goals

○ Develop enumeration methods to meet the current Census environment

○ Utilize IT infrastructure in the form of e-Census

○ Intensify field management to reduce coverage errors

○ Select survey items to meet the demand of policy makers

○ Reinforce public relations to the public

○ Establish a systematic field organization

○ Operate a 080 toll-free center for counseling by specialists

○ Reduce data processing time by using Web-based direct input and editing

○ Completion of data release by the end of 2006

12. Release of the Results

○ Preliminary results : December, 2005

○ Complete enumeration results

- Population : May 2006

- Households and Housing : July 2006

○ Sample enumeration results : September ~ December 2006

○ Special analysis by characteristics : September 2006 ~ December 2007