The data collection method using the Internet

1. Background

1. Until recently, it was usual for surveys to be completely carried out with an interviewer to visit houses or establishments with a questionnaire and pencil. This traditional method requires a lot of time to complete the process of visiting respondents, filling out the questionnaires and recording the data. Moreover, it burdens respondents in the matters of time and privacy since they have to meet with the interviewer revealing some amount of personal information.

2. However, in accordance with the great developments in Information Technology, there has been a continuous effort to apply IT to statistical surveys to obtain a higher efficiency. In the late 90s, the computer assisted self-interviewing (CASI) method of using a palmtop or laptop was introduced in Korea, partly replacing the old method of paper questionnaires in the Consumer Price Survey (CPS) and Economically Active Population Survey (EAPS). In addition, the Current Mining and Manufacturing Survey (CMMS) conducted through PC communication has been put into practice as well.

3. After 2000, the Internet became widespread throughout the nation and it rose as an important factor in finding efficient methods in statistical surveys since the 90s. The idea of using the Internet for surveys has been tried in private sectors such as public opinion poll organizations. However, NSO has been reluctant to use the Internet methods because of the lack of Internet environment and protection of privacy.

4. However, as there was an explosion in the number of Internet users, the Internet was finally applied to official statistical fields. As table 1 indicates, the rate of the Internet usage in Korea was over 51% in 2001 and further increased to 73% in 2006.

< Table 1 >

Year	01	02	03	04	05	06
rate of the internet usage	51.6%	58.0%	64.1%	68.2%	71.9%	73.5%

Source: National Internet Development Agent of Korea (http://www.nida.or.kr/)

Initially starting with application of the Internet survey for businesses and enterprises, the Internet survey has partially extended its reach to the 2005 Population and Housing Census and Household Income and Expenditures Survey (HIES).

5. The changes in the survey methods are expected to bring about a whole new innovation to the traditional paper survey. It will save time conducting the survey,

reduce the burden of respondents, increase the accuracy of the data and shorten the data processing time.

6. The following will discuss how Internet surveys are conducted in Korea and suggest a future direction by assessing our experience with the Internet survey.

2. Applying the Internet in a statistical survey

1) On the business survey

7. In Korea, the very first case of application of an electronic survey was the CMMS. In 1997, the NSO conducted this survey using PC communication for the first time. It started by applying an electronic survey to a few of about 8,000 establishments. In 2001, the extent of the application was widened to 31% of the total, which included 2,500 establishments.

8. The survey was practiced according to a specific process. Software for an electronic survey was installed in the computer(s) of a respondent within an establishment. Respondents received education concerning the use the program. They were able to fill out the electronic questionnaire using the software at their convenience and then transfer the file to the NSO. The NSO examined the submitted files and [sums them up with the result conducted in a paper survey] to produce total statistics.

< Figure 1- Basic concept >
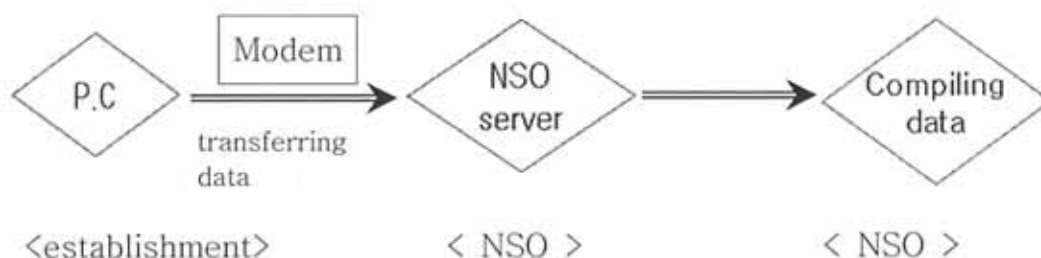
9. Data collecting methods using PC communication encountered various difficulties when operated for the first time. Firstly, due to its time saving qualities, there was some opposition from local offices because they were aware of the possibility of a reduction in staff due to this new method. Therefore, it was hard to receive cooperation from field survey staffs. Secondly, while installing the survey program to a respondents' computer and attempting to operate the program, some technological errors occurred. The only options were reinstallation and re-operation. Thirdly, there were internal problems related with the capacity of computers, especially in the cases of small-scale business. It was difficult to install the surveying program primarily because the PCs were out-of-date. Moreover, if respondents were inexperienced with computers it was difficult to get assistance. Additional problems included the possible dangers of hacking and weaknesses in information protection.

10. For these reasons, the number of establishments who cooperated in the new data collecting system was very small in the early stages. In 1998, the participation rate was below 10%. However as the efficiency of survey has risen, increases in the

rate occurred in the following years of 1999 and 2000 to 15% and 20%, respectively.

11. After the year 2000, the Internet became widespread and PC communication through a modem became out-of-date as a majority of establishments started to use the Internet instead of PC communication. In order to keep pace with this trend, the data-collecting system had also veered its direction towards the Internet. Instead of manually installing software for a questionnaire in a respondent's computer, the software for the Internet data-collecting system was available for download automatically through the Internet. It visualized far better function than before. As a result, there was a sense of elevated satisfaction levels among surveyors and respondents.

12. Respondents of the establishment can access the server of NSO with their IDs. When they filled out the electronic questionnaire, the data is automatically transferred to the NSO. The electronic questionnaire program offers help functions, explanations for survey terms and an editing function so that respondents can check input errors for increased accuracy.

13. For these reasons, the NSO could collect data from 3,862 establishments, which accounted for 50.4% of all entities (7,860) of the monthly industry survey and 1,242 establishments, which were 52.9% of the total entities (2,347) of the Manufacturing Production Capacity Survey (MPCS).

< Figure 2- Basic concept >

2) Introduction of an electronic housekeeping book over HIES

14. The NSO conducts a Household Income and Expenditures Survey (HIES) that collects approximately 9,000 households' daily housekeeping records. This survey is burdensome for both the interviewer and respondents. Recently, conducting a survey has become harder due to an increase of dual income households and recognition of privacy issues.

15. In this difficult situation, surveying methods was developed from a paper housekeeping record to an electronic one. Respondents download an electronic housekeeping record from the Internet website and input income and expenses.

16. The respondent can check their electronic housekeeping record at any time and make good use of it by using it as a reference when making a living plan for their daily life. Also, there is a helpful function for the respondents, which provides real-time

information of bank account and credit cards. This function is expected to raise the accuracy of the statistics. The E-housekeeping recording method is especially accepted by single and dual member households, since the data can be inputted at home or work.

< Figure 3- The flow of the electronic housekeeping book system >
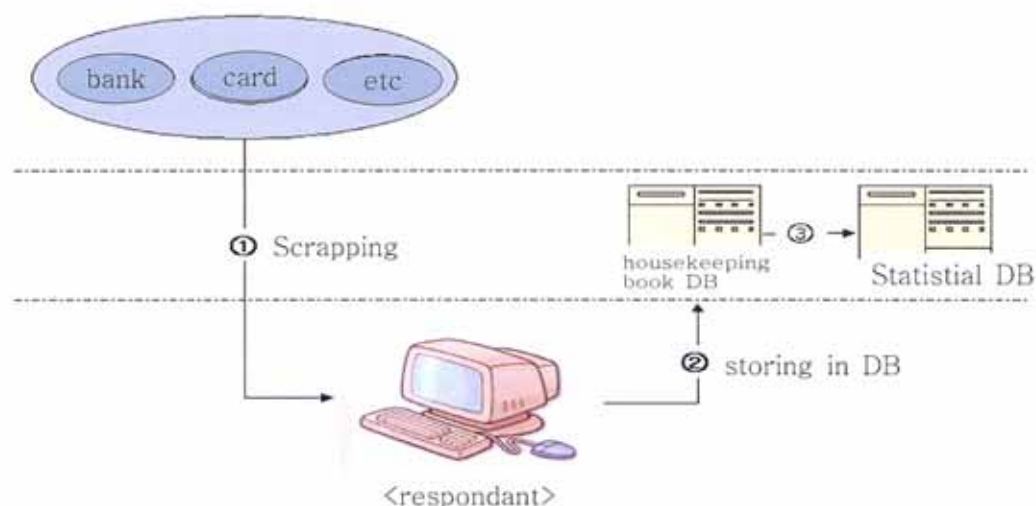
17. The Electronic housekeeping record survey contains beneficial characteristics. First, it is a housekeeping record survey that uses computer and the Internet. [A household transfers categorized income and expenditure divided by various.] It is convenient with its auto registering features, which inputs the data automatically according to the records of a bank account and credit card. The NSO operates a cyber surveying help center so that respondents can solve technological problems and ask questions such as how items were categorized or how to properly use the

program. The transferred records are stored in a housekeeping DB and data required in compiling statistics are stored in statistical DB after converting them into a special format.

18. Respondents can also use the outcome on administration of income and expenditures. For this, the electronic housekeeping record provides additional functions such as an analysis tool in household statistics and graphs. It will especially benefit single and dual income households since the electronic housekeeping record is a highly convenient time savor.

19. The electronic housekeeping record, which was introduced this year, is expected to bring about a huge innovation in the field. Firstly, it is thought to raise the timeliness of statistical data. As the simultaneous distribution and withdrawal of the paper version of the housekeeping record gradually decreases, it will save approximately 10 days during the data editing periods. Currently, it has only been introduced to volunteer households with on-line access so the remaining respondents are still using the paper questionnaire. The speed at which the number of households using the electronic housekeeping record increases will be the main factor in determining the total timeliness of the statistics. Secondly, improvement in the accuracy of statistical data is anticipated. Respondents can see the records of bank account and credit card. Moreover, this new method provides the function of automatically classifying items and convenience in electric entry. Also, due to this new method, the possibility of response from groups including dual income and single households have a higher probability of responding to the household survey improves. <As a result, the degree of representative also gets higher.> Thirdly, the

work load of the surveyor decreases. Therefore, due to the introduction of the Internet Survey and subsequent efficient use of time, workers can focus on other projects.

20. Among the samples of the HIES, the households able to access the Internet were 62% (4,409 households) and the households with members over fifteen-years-old who were capable of using the Internet recorded 57% (4,116 households). The main respondents using the Internet were 43% (3,105 households). Therefore, these factors suggest that there is a high possibility of the proliferation of the electronic housekeeping record. In order to encourage the momentum, the NSO issued 50,000won per month to each household as compensation.

3) Introduction of the e-census

21. Last year, the 2005 Population and Housing Census was carried out. Recently, the category of single and dual income family have increased in Korea, which are difficult to survey. As one of the solutions, the Internet survey was selected to aid in this problematic area. In fact, the e-census means the Internet survey contains a broader extent because it applies to Internet for the purpose of collecting data as well as examining survey data, managing surveyor and materials for the surveys. However, this paper only deals with the data collection through the Internet.

< Figure 4, Basic concept of the e-census >
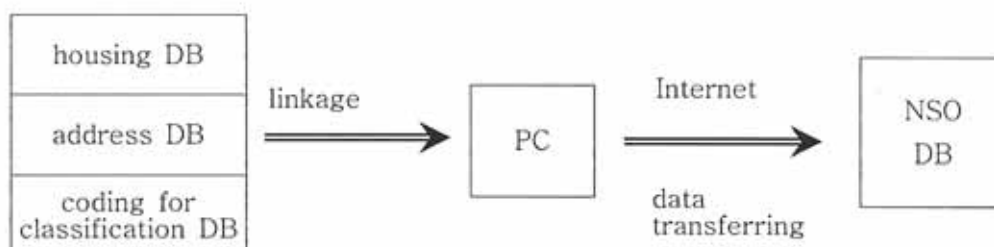
22. In order to carry out the e-census, extensive advertisement on implementation is required. Then, the NSO takes applications from households that want to respond through the Internet. The system of e-census consists of four steps. Beginning with the confirmation of the applications and one's regal name, respondents fill out the on-line questionnaire. The data of the electronic questionnaire is automatically transferred to the DB of the NSO. A Census staff of the NSO examines the data and approves the form once it is determined to be errorless. For convenience in the process, various help functions related to the survey are provided for ease of response.

23. However, since this was the first time to make use of the Internet in a census, it was applied to only a small portion of the population. In the beginning, 2%(280,000 households) of all households were expected to participate but only 1%(140,000 households) actually participated in the e-Census. Nevertheless, the first e-census provided a very valuable experience for the next Census because it laid a foundation for more a developed form of e-Census. Especially, It could be effective in encouraging the participation of groups such single and dual

households in the e-census. It is currently at the evaluation stage but this experience will have a huge influence on the improvement of conducting the future population census.

4) Cost-benefit analysis of Internet survey

24. The calculation of the B/C analysis is focused on direct-cost and direct-benefit. If we could add indirect benefits such as a decrease in the mount of interviews and commute times, the benefit would be larger than that of the table below.

< Table 2 >

(100 million won)

	Benefit	Cost	B/C ratio
CMMS	7.6	0.5	1,520%
HIES	9.5	16.1	59%
Population Census	4.6	2	230%

25. In the case of the CMMS, the e-survey annually produces a 760 million won benefit because of its reduction in labor cost. Expenditures do exist in the form of developing cost for software but amounts to approximately 250 million won. Hence, it costs 50 million won in direct-cost each year if the durability of the software is viable for 5 years. As a result, the B/C ratio reaches 1,520%. However, if we consider the time spent visiting establishments, commute times and other social cost, the B/C ratio would be much larger than the figure above.

< Table 3 : The participation ratio of the e-CMMS >

2003	2004	2005	2006.9
37%	46.6%	47.1%	53.7%

26. For these reasons, the NSO made an effort to persuade sample establishments to participate in the e-survey. Even though the ratio of participation in the e-survey was only 37% in 2003, the ratio steadily rose to 53.7% in 2006.

27. In the case of the HIES, it took 1.6 billion won of direct-cost per year and the benefit amounted to only 950 million won. The B/C average rate from 1996 to 1999 was only 59%. What was the reason for the low benefit ratio in the case of the household survey? Firstly, it issued compensation of 50,000 won per household to e-survey respondents in order to encourage participation. Therefore, the more households that participate, the more direct-cost that is needed. This point distorts the application of the economic of scale in the e-household survey. However, it is difficult to consider it as net cost for introduction of e-survey methods because the compensation contains a reward for the recording of monthly housekeeping records. Also it is an incentive to encourage people to use the electronic housekeeping record. Secondly, it is difficult to calculate the indirect benefits based on conveniences. The e-survey increased in recollection rate of the housekeeping record, reduced the possibility of exposing privacy, and decreased errors that might occur in the computation process. However, most of those social

costs were excluded. If these conveniences were included, the B/C ratio would increase to a higher percentage.

	2006.1	2006.3	2006.6	2006.9
The ratio of e-survey	4.8%	11.6%	23.0%	24.8%
# of household	341	825	1,639	1,767

28. The Internet housekeeping survey was hailed with good response.

Nine months offer its inception, 1,767 households are using this pace, it is predicted that 50% of the total HIES will be conducted in this manner by late 2008.

29. In the case of the e-census, 2-3% of the total households were expected to reply using the Internet survey, but the actual figure was 0.9%. Therefore, cost-benefit was 460 million won which was lower than anticipated. Despite the outcome, the B/C ratio was 230% because the Internet questionnaire development cost was only around 200 million won. It proves that the level of efficiency can greatly improve even though the rate of households participating in the e-census shows only a slight increase. This suggests a direction the census might in the future by implying an increase in cost-benefit by geometric progression as participation in the survey grows. Therefore, the underlying foundation of the e-Census is observed in the protection of an individual's privacy and attracting participation of problematic social strata such as single households and dual income families who would not otherwise cooperate in a statistical survey.

3. Prospects and limits

30. As shown above in the case analysis, there is a strong probability that Korea will expand the e-survey through the Internet. According to B/C analysis, the Internet survey has a very high efficiency rate compared to the traditional paper survey, as well as other various merits. Firstly, it cuts down the cost of the statistical survey. It can save on high personnel expenses and the surplus of labor could be put to use in other areas. Secondly, it can reduce the processing time of statistical data.

The monthly industry e-survey could save 1.5 days of processing and the household e-survey could reduce its response time by 10 days. Thirdly, it can increase the accuracy of data. The editing program can assure a high accuracy of responses and strengthen administration. After starting the Internet household survey, the response rate improved from 81% to 83%. It proves that the survey, which would otherwise burden respondents, such as the household survey, is effective using this method.

31. There are several problems in utilizing the Internet. An area deprived of Internet access cannot apply this method. Approximately 25% of the total populations in Korea, especially rural areas, do not have access to the Internet. Therefore, it is necessary to continue using the traditional survey in rural and/or elderly communities, who might be unfamiliar with computers.

Secondly, only those who want to respond through the Internet can be targeted in this method. Therefore, the e-survey is simply a supplemental method to the traditional paper survey. Thirdly, countermeasures for computation failure are

required. Finally, a system is needed in order to prevent privacy leakage against hacking. For this, data should be managed with the clearest extent of responsibility.

32. Despite its weaknesses, the Internet survey is proven to be a very efficient method, which has the ability to suit the changing environment. Hence, it seems that it will eventually replace the traditional way. It will develop as an applicable and valid method in all areas including large scaled Census, business and household surveys.

1