Population Censuses on the Internet

Author:

Dr. Werner Haug

Director, Population and Employment Division

Swiss Federal Statistical Office

Espace de l'Europe 10

CH-2010 Neuchâtel

Switzerland

E-mail : Werner.Haug@bfs.admin.ch

Abstract

The paper discusses the problems and limitations of data collection through the Internet, dwelling on recent experiences. Online Population Censuses have several advantages compared with paper and face-to-face interviews. Questionnaires on the Internet can be developed in a very user-friendly manner. The costs of data entry and data editing are much lower, the data are rapidly available and are of better quality. The experience of Singapore and Switzerland clearly demonstrate the feasibility of Electronic Censuses.

But there are also limitations. A central database with names and addresses as well as a centrally managed mail and password system are preconditions for the successful use of multi-modal data collection strategies (Singapore and Switzerland). Scaleable and flexible solutions, secure and powerful data connections as well as the permanent monitoring of online transactions are expensive. Major problems include also the lack public confidence with regard to confidentiality and the digital divide (USA, Switzerland). Population Censuses of the future (at least in highly developed countries) will probably follow a multi-modal strategy, combining register information, interviews by Internet, telephone and face-to-face.

Population Censuses on the Internet

Werner Haug, Swiss Federal Statistical Office

1   Censuses in Transition

Population and Housing Censuses are still conducted in most countries of the world and represent a major source of demographic and socio-economic statistics. Their main advantage is the richness of geographically detailed data and their capacity to inform about small social, cultural and demographic groups. The possibilities to analyze and present population census data have increased rapidly in recent years thanks to the development of data ware-housing and powerful geographic information systems, driven by an increased demand for small area and neighborhood statistics.

At the same time, traditional Population Censuses are increasingly challenged - at least in the developed countries - because of their high costs and the labor-intensive fieldwork, the need to reduce the response burden and increase the timeliness of the data. This has created strong incentives to develop alternative ways of census data collection and more effective methods of data processing. The efforts are going into different directions and Censuses are becoming increasingly complex combinations of different data sources, data collection strategies and statistical estimates. A clear sign of this rapidly changing landscape are the European Recommendations for the Population and Housing Censuses 2000, which - for the first time in the history of official statistics - define Censuses not by a common method of data collection but by a common output of statistical results, independently of the way in which they are produced (UN/EUROSTAT, 1997).

A first major trend in the search for new ways of data collection is the increased use and systematic linkage of administrative registers. The Nordic countries are far ahead on this way. Denmark, Finland, Iceland, Norway and Sweden are in a position to conduct entirely or largely register based Censuses, without questionnaires and interviews. Several other small countries are in transition from a traditional enumerator based Census to a register based Census, e.g. Austria, Belgium, Switzerland and Luxembourg. They are already conducting partially register based Censuses and are building up their capacities for a fully or largely register based Census after 2000/2001 (for an overview see EUROSTAT, 2001).

A second trend is the increased use of sample surveys and statistical estimates in order to replace or complement the full count. France, the Netherlands, the US - among others - illustrate this trend. France will conduct a continuous partial Census after 2000, involving a geographically rotating total count for rural areas, sampling for urban areas and estimates to fill the gaps (Desplanques, 2000). The Netherlands combine data from administrative sources with estimates based on sample surveys and they apply a complex imputation and weighting scheme. The US introduced sample surveys to correct for the partial undercount, especially in urban areas, and they plan to replace the long Census form in the 2010 Census by regular data from a large ongoing intercensal sample survey, called the American Community Survey, which is planned to start in 2003 (Alexander, 2000).

A third trend is the use of modern mailing and ICT technologies for data collection. This includes the building up of databases with names and addresses for mail-out and mail-back, the use of call centers to support field work and conduct interviews by phone and finally the use of the Internet as a complementary way of data collection. Switzerland, Singapore and the US were the first countries to use the Internet in Population Censuses. This paper discusses the possibilities and limitations of data collection through the Internet, dwelling mainly on the evaluation of the Swiss experience.

2   The Electronic Census: recent experiences

The year 2000 marked a breakthrough in the use of the Internet for big population surveys. In March 2000 the US Bureau of the Census offered the option to fill in the short Census form on the Internet, although only in a limited test version and without featuring it prominently. 300'000 persons filled in their questionnaire on the Internet. The tests continue for the American Community Survey, where a test panel has the option to respond by Internet. The results are compared with the regular survey as a control group (Alexander, 2001).

Singapore conducted a register based Census which combined the use of administrative data from the Household Registration Database and the National Database on Dwellings with an in-depth survey for a 20% sample population. For this survey, a tri-modal data collection strategy was adopted, which comprised Internet e-filing as a first stage, offered to everybody, phone interviews as a second stage for those who did not file in by Internet and field work with paper forms as a third stage for households which remained uncontactable. From the 215'000 households in the sample, 15% filled in by Internet, 60% were contacted by phone, 10% were interviewed face-to-face and 15% submitted their information in a combination of ways (Eng Chuan Koh, 2001).

Switzerland offered in December 2000 the option to fill in the main Census form on the Internet (E-Census). The option was open to those cantons and communes of the Swiss Confederation that were linked to the central Census database (with names, addresses and basic demographic data), covering 90% of the Swiss population. The paper forms were preprinted with the data of the Central database and mailed out to all the households. Internet E-filing was offered as an alternative to mailing back the paper forms (Haug, Buscher, 2000). For 10% of the population (especially in small communes and in rural areas), the Census was conducted in a traditional way with enumerators who distributed and collected the forms personally around Census day.

121'000 households with 280'000 persons or 4.2% of the population completed the Census forms successfully via the Web. 90% completed them during the two weeks before and after Census day. The remaining 10% completed them during the four months which followed the Census, in relation with the automated reminder and inquiry system of the Census Office (SFSO, 2001). Approximately the same proportions (90% return around Census day, 10% during the following months) applied to people who sent back the paper forms by post.

The good return rate of 90% was largely unexpected for the Swiss Statistical Office. Several arguments could to explain it: the discipline of the Swiss population, the

highly visible and well targeted public relations campaign and the modern image of the Census, thanks to the use of up-to-date ICT.

3 Why online surveys?

Several goals can be achieved via the option of the Web:

1. A modern image for the Census

In the early days of Censuses and almost until the end of the 20th century, there was no other possibility to collect the necessary information on persons and households than with the help of paper forms and enumerators. But this traditional survey method began to take an increasingly antiquated image as the information society developed at a breathtaking pace. Modern ICT tools enhance the image of the Census and can reduce some of the resistance against Census taking.

2. Better user-friendliness

The Web has no limits to the size of the questionnaire, the support and help functions. Respondents can choose the language in which they want to respond, they are assisted by the good legibility of the questions and a large number of electronic aids and notes about completing the questionnaire. The questionnaire can be customized (personalized questions, suppression of non applicable questions) depending on the answers to preceding questions and to the characteristics of a person according to age, sex, economic activity etc. In addition, it is possible to print a personal copy of the completed questionnaire.

3. Better data quality

The Web offers the opportunity to build in consistency checks and show immediate warnings for missing or implausible data. Each field can be linked to attributes, which have to apply before the questionnaire can be completed. Selection menus reduce the range of wrong or unclear answers and support the coding process. Data entry is completed directly, without intermediate stages of writing, posting, scanning or copying. Progress can be monitored immediately and any problems arising with the data can be recognized early on.

4. Cost reduction

The costs of data acquisition decrease thanks to the direct data entry in a central database, the skipping of intermediate stages (scanning, manual data entry) and the lower proportion of missing, wrong or implausible data thanks to better coding and on-line checks.

However, in order to reach these goals, a number of technical and political challenges have to be met. And if they are not well handled, they can easily offset the possible gains in efficiency or can even have reverse effects on the image and the quality of the Census.

4   Technical challenges

The more technical challenges concern four main fields:

1. The identification of users

A central database with identifiers is a necessary precondition for an Internet survey, where data entry is independent of the precise location of the person and where the right to access the database has to be checked online. In the context of a multi-modal Census, the dependable identification of the user is of utmost importance for security reasons and in order to avoid double counts and/or under-enumeration. In the Swiss case, a unique User-ID and a password for E-census were printed on the household questionnaire and mailed out together with the paper forms. During the login procedure, people had to identify themselves with their User-ID and password. The number of successful logins (166'000) was one third lower than the number of visitors to the E-census homepage (240'000).

After the login procedure the visitor found the household questionnaire and the questionnaires of the household members on the screen, with names, addresses and basic demographic data already inserted. He/she had to correct missing or wrong data and answer the questions for which no information was available from the registers (90% of all questions). It was also possible to add new persons to the household or delete persons that were no longer living in the household. All household members had to complete the questions via the Web. A multi-modal strategy was impossible on the level of the individual household.

2. Security and the protection of privacy

The rules for the protection of privacy do not depend on technology - the same rights apply in the digital area. However, technology should provide the tools to protect privacy in an efficient and trustworthy way. But as opinion polls show, large segments of the population in Europe and the US don't trust the security of the Internet and the protection of privacy on the Web. This limits the range of potential users not only for E-commerce but also for E-statistics. In any case, official online surveys have to use the most up to date security technologies, including certified encryption, secure connections and fully protected computer environments as well as the permanent monitoring of on-line transactions.

The recent Population Censuses on the Web did not encounter security problems. The solutions in Switzerland and in Singapore were safe. The Swiss solution was four months on-line without privacy problems and without suffering from uncontrolled cyber attacks. But the initial investment and the monitoring of the transactions in the open and vulnerable space of the Web are quite costly. This clearly underlines the need to go beyond ad hoc solutions for individual surveys and develop permanent platforms for online transactions that are equipped with the necessary infrastructure in terms of hardware, software and manpower.

3. The dimensioning of the technical solution

The programming of the questionnaire, the ergonomics of the site, the possibility to interrupt and resume a session – in short: forms processing and session

management do not pose major problems any more. Standardized solutions are increasingly available.

However, Population Censuses in particular pose problems of logistics to the Web. Offering an Internet option to large parts of the population during a very limited time frame, combined with a targeted public relations campaign, puts enormous strains on the technical solution. There is an immediate rush on the servers and call centers, which have to handle a huge number of contacts and calls. The agility and flexibility of the technical solution as well as of the staff involved are essential success factors. In the Swiss case, it was only thanks to the close co-operation with the private sector that the most serious constraints could be overcome.

4. Incentives for Internet users

The immediate advantage of the Internet option is on the side of the Statistical Office - not of the respondents. In the Swiss case, the average E-census session lasted 18 minutes and the telephone costs were charged to the respondents. The paper form was filled in at least as rapidly, it had less constraints (password, online checks etc.) and mail-back was free of charge.

The public relations campaign stressed the modernity of the tool and its user-friendliness. But the full potential of the Internet (with over 30% regular Internet users at the time of the Census) was not exploited. The co-existence of alternative options, which are less burdensome for the individual (but more expensive for the community), is clearly not to the advantage of the Web. Even regular Internet users weigh the pros and cons as long as they have alternative solutions, which seem easier to use, safer, faster and less costly.

Therefore, multi modal data collection strategies have to carefully consider the incentives linked to the different options and promote at least equal opportunities for the different approaches. This is also illustrated by the experience of Singapore where the Internet had a hard stand against the telephone interview.

5   The digital divide: who are the Internet users?

The digital divide is of utmost importance for Census methodology. If we compare the profile of the E-census users in Switzerland with the profile of the total population we find a clear bias, which is stronger than the usual bias of sample surveys (Stamm, 2001). We found fewer differences than expected according to cultural criteria: between the language regions, foreigners and Swiss nationals, urban and rural areas (4.4% in urban areas,: 3.8 % in rural areas, with variations from 0 to 16% between the communes). However, there were marked differences according to demographic and socio-economic criteria: sex, age, household size, educational background and economic status.

The typical e-census user is a young, educated, single, male professional. Family households, women, elderly persons and persons with a low educational background are largely underrepresented (see table). The gender and age bias is even stronger if we take into account the characteristics of the person who actually e-filed the questionnaires of the household.

Table: Profile of E-Census users, Switzerland, Population Census 2000*

	E-Census, %	Total population, %	Difference
Sex			
Female	45.5	51.2	- 5.7
Male	54.5	48.8	+ 5.7
Age group			
0-19	25.7	23.3	+ 2.4
20-39	40.7	29.6	+ 11.1
40-64	29.4	32.0	- 2.6
65-79	3,2	11.2	- 8.6
80+	0,9	4.0	- 3.1
Citizenship			
Swiss	85.1	80.6	+ 4.5
Foreign	14.9	19.4	- 4.5
Main Language			
German	63.2	63.6	- 0.4
French	19.9	19.2	+ 0.7
Italian	5.8	7.6	- 1.8
Romansh	0.3	0.6	- 0.3
Others	10.8	8.9	+ 1.9
Marital Status			
Single	49.4	42.0	+ 7.4
Married	44.3	46.6	- 2.3
Divorced	4.3	5.5	- 1.2
Widowed	2.0	5.9	- 3.9
Educational attainment (Age 25-64)			
Compulsory	12	19	- 7
Secondary (vocational)	43	51	- 8
Secondary (general)	10	7	+ 3
Higher (vocational)	21	13	+ 8
Higher (university)	14	10	+ 4
Economic status (Age 15+)			
Economically active	74.9	62.4	+ 12.5
Homemakers	8.6	7.6	+ 1.0
Pensioners	6.4	18.9	-12 .5
Students	6.9	7.8	- 0.9
Unemployed	3.3	3.3	-
Status in employment			
Self-employed	10.8	17.8	- 7.0
Chef executives	12.7	14.7	- 2.0
Middle management	19.6	15.1	+ 4.5
White/blue collar workers	48.3	47.7	+ 0.6
Apprentices	2.6	4.7	- 2.1
Without indication	5.9	-	+ 5.9

*Provisional data for total population

Censuses are ground truth. They tell us about the real situation in the homes of the citizens. And we have to acknowledge that the Internet is still out of range for a lot of persons, especially low-income households, big families and many elderly persons. Access to the Internet and the ability to use it (not only for surfing but for complex transactions) are limitation factors for the successful promotion of online surveys, not only in developing but also in developed countries.

6   Conclusions and outlook

The global experiences of the year 2000 proved that the Electronic Census is feasible from a technical and operational point of view and that safe solutions exist. In Switzerland, the general acceptance and the image of the Internet option was very positive. The data were available much more rapidly and in better quality than the paper-based data, due to online data entry and online checks, thus saving several months of processing time. However, scaleable and flexible solutions, secure and powerful data connections as well as the permanent monitoring of online transactions are quite expensive. And as long as the access to the Internet and the incentives to use it are not radically improved, official statistics will continue to use a mix of data collection methods.

Harmonized and up to date population and housing registers are the backbone of the central database and mailing systems, which are essential for online surveys. Therefore, improving electronic registers and their linkage is also an investment in online surveys. Future trends are promising and will lead to safer and less costly solutions. Today, mailings are still largely paper based. But they could rapidly become e-mail based, which is already the case with many B2B (business to business) contacts. Permanent platforms for online surveys will combine e-mail, Internet and telephone. In Switzerland, they will first be developed with government authorities on the local and cantonal level and second with businesses for regular business surveys. Household surveys will follow at a later stage.

Will enumerators and paper questionnaires entirely disappear in future Population Censuses? This is not just a question of content and technology, but also of public trust, education and equity. It is essential in this context, to place online surveys in the more general perspective of E-Government, which will promote a new administrative culture, building on transparency and the standardization of the available information and fostering the regular electronic exchange between administrations and the empowered citizen.

References

Alexander Charles H., The American Community Survey and the 2010 U.S. Census, Paper presented at the INSEE/EUROSTAT Workshop "Censuses beyond 2001", Paris, November 2000

Alexander Charles H., Internet test panel for the American Community Survey, Personal communication to the author, Washington D.C., May 2001

Desplanques Guy, Le recensement rénové en France, Paper presented at the INSEE/EUROSTAT Workshop « Censuses beyond 2001 », Paris, November 2000

EUROSTAT: Censuses, Item 6, Working Party on Demographic Statistics and Population and Housing Censuses, DOC.DEM/CEN/e4/3/01, Luxembourg, February 2001

Haug Werner: The Electronic Census: A ground breaking achievement, in: Delivering on the E-Government Vision, Microsoft Government Leaders Conference, Seattle, March 2001

Haug Werner, Buscher Marco: E-Census: The Swiss Census 2000 on the Internet, Paper presented at the INSEE/EUROSTAT Workshop "Censuses beyond 2001", Paris, November 2000

Koh Eng Chuan: Data Collection Strategy for Census 2000, Singapore, Personal communication to the author, Singapore, Mai 2001

UN Economic Commission for Europe, Statistical Office of the European Communities (EUROSTAT): Recommendations for the 2000 Censuses of Population and Housing in the ECE Region, Statistical Standards and Studies, No 49, Geneva, 1998

Stamm Hanspeter: E-Census – Users and user profiles, an analysis of the server statistics and the database, Report to the Swiss Federal Statistical Office, Zurich, June 2001

Swiss Federal Statistical Office: The Population Census 2000 – a positive balance, Press Release, Neuchâtel, June 2001

Neuchâtel, July 11, 2001