



**Economic and Social
Council**

Distr.
GENERAL

ECE/CES/GE.41/2009/18
19 August 2009

Original: ENGLISH

ECONOMIC COMMISSION FOR EUROPE

CONFERENCE OF EUROPEAN STATISTICIANS

Group of Experts on Population and Housing Censuses

Twelfth Meeting
Geneva, 28-30 October 2009
Item 6 of the provisional agenda

ON LINE DATA COLLECTION

Theoretical and Practical Analysis about Census by Internet

Note by the National Statistics Institute, Spain

I. INTRODUCTION

1. It has been some time since new technologies became part of everyday life. It is now difficult to imagine life without the existence of mobile phones, laptops or Internet connection.
2. Regarding the Internet, it can be used for very different purposes like sending an email (instead of sending a letter), reading the newspapers, looking up all kinds of technical questions or taking a look at personal banking information.
3. This paper explores the use of the Internet as a channel with a lot of possibilities (and an enormous potential of growth) for the next Census round of 2011.
4. Population and Housing Censuses are among the most important statistical operations, not only because of their magnitude but also because of tradition. For example, modern Censuses in Spain have been carried out since 1768 (Censo de Aranda) or 1787 (Censo de Floridablanca). The 2011 Census will be the 20th in Spanish history:
18th Century (3): 1768, 1787, 1797
19th Century (5): 1842, 1857, 1860, 1877, 1897

GE.09-

20th Century (10): 1900, 1910, 1920, 1930, 1940, 1950, 1960, 1970, 1981, 1991

21st Century (2): 2001, 2011.

5. Over the years, life habits, especially those related to technology and electronic devices, have changed significantly. With the passage of time, the speed of changes is accelerating.

6. Spain was one of the first countries in the world that made it possible for people to answer to 2001 Census questions via the Internet. However while less than 1 per cent of Spanish dwellings (13,768 of the 14,187,169 dwellings) answered the 2001 Census by Internet the experience is useful in planning the 2011 Census.

7. Several reasons make the situation for the 2011 Census much more promising than 2001. The numbers of people using the Internet everyday has increased significantly, bandwidth offered by IT companies is much higher than in 2001, Web applications are much more frequent and their possibilities are not so restrictive (for example, Internet security protocols have been significantly enhanced) and device storage size has grown exponentially in the last few years. Because of these reasons the scenario in the 2011 Census, where the Internet will play an important role, will be very different from that at the time of the 2001 Census, where the existence of the Internet was essentially symbolic.

II. THEORETICAL ANALYSIS

8. This section focuses on the theoretical features of answering the Census by Internet and specifically on its strong and weak points.

A. Building a Web application: worthwhile or not?

9. From an economical point of view, the development of a Web application that makes it possible to answer Census questions by Internet (including a Web page, security protocols and a server that can deal with an adequate number of connections) involves only fixed costs. If this prospect is compared to other methods of responding exhaustively to the Census (paper, CATI or CAPI with hand held devices), all alternatives have both fixed and variable costs.

10. For non Internet solutions, variable costs come about because the numbers of people that answer the Census directly affect the cost. For example, the amount of printed or scanned paper, of telephone interviewers (CATI) and of enumerators (CAPI) depends on the number of people that make use of the different channels, and this entails a series of variable costs.

11. Using the Internet approach all variable overheads disappear since the development of the Web application does not depend on the amount of people that make use of this channel¹.

12. Thus it can be concluded if fixed overheads of developing a Web application are not high

¹ Except for the capacity of the system that has to support it (the more people use this alternative, the more capacity and budget are needed). Nevertheless this does not involve outstanding extra costs.

and the Web rate of response is appreciable, then the Internet channel is worthwhile. In section 3 of this paper this statement will be numerically explained.

13. Another conclusion is that if it is assumed that the Web rate of response is initially unknown, the development of a Web application that enables Internet responses could sometimes not be worthwhile, because the fixed overheads involved in Census by Internet are higher than the sum of fixed and variable costs when responding Census by any other channel.

14. One crucial point which has been magnified by the current economic global situation is that it is important that all Censuses should be designed with the maxim of minimizing costs as much as possible.

B. Information: quality and possibilities

15. One of the most important advantages of online Censuses is that it improves the quality of the information received through this channel. Inconveniences like difficult to recognize (or scan) characters and impossible or blank values in some topics can be avoided. For example, a person's answer in the "country/place of birth" topic would be accepted only if its value is present in a drop down menu.

16. Online Censuses can also eliminate inconsistencies among different answers that belong to the same form or even between different forms.

17. Scanning and recognizing characters takes time and requires significant resources. In addition, the quality of the information in Census paper forms is not always guaranteed.

18. Another advantage of online Censuses is that, depending on the person's profile, the application could ask only those questions that are necessary.

19. For example, it would be nonsense to ask an 8 year-old child about his legal marital status. Defining what kind of questions should not be asked according to the person's profile is not an easy task.

20. An additional advantage of answering the Census online is the possibility of using pre-filled answers to some questions where there is a high probability of knowing the correct answers in advance. In Spain, those pre-filled answers could be obtained through the 2001 Census and also through administrative registers.

21. For instance, if one question was asked in the 2001 Census, a person's answer could be shown as a predefined answer on the 2011 form (or do not directly ask that question to the user and assume the information from the 2001 Census or the information from registers). If the person completing the census form does not agree with the answer, there will be the possibility of changing it, but if they agree, it will only be necessary to press a confirmation button and the next question will be provided.

22. To conclude, efficiency of the Internet Census is much higher than other methods such as paper because the questionnaire's flow will be personalized for each individual and only those

questions that require an answer will be asked to each person (for example, question about the total number of children born alive will only be asked of women 16 years or older).

C. Importance of push-methods

23. It has been demonstrated (there are studies in other countries like Canada or Sweden) that the possibility of answering at the same time by different channels (like paper, CATI or telephone) affects negatively to the Web percentage response.

24. If a high percentage of Web responses is the objective, then one or more of the following should be considered:

- (a) The letter that is sent to each dwelling to motivate them to answer the Census by Internet should be written very carefully. Outsourcing to experts in sending suitable letters to large numbers of people should be considered. In addition, cognitive techniques to evaluate how effective letters are in achieving their objectives;
- (b) The Internet option to answer should be opened to census respondents before any other channels;
- (c) Only if a dwelling expresses its preference for not answering by Internet, other channels like paper, CAPI agents or CATI will be available.

25. In Spain during a Pilot Census conducted from 13 April to 31 July 2009 the following two methods were used:

- (a) Method 1: Both a letter and Census paper forms were sent to each dwelling at the same time. Users could answer by paper, Internet or telephone;
- (b) - Method 2: Only a letter was sent to each dwelling. Firstly users could answer by Internet or by telephone. Secondly, information about all the dwellings that did not answer before a deadline was collected using CAPI-enumerators.

26. Because of low rates of response, after the first month for both methods, an additional letter was sent reminding people about the importance of taking part in the Pilot Census.

27. Section 3 of this paper comments on differences between the Web rate of response for both methods.

D. Other key points of Census by Internet

28. Answers provided by the internet to different topics could be disseminated much earlier than those that come from any other channels because they do not need to be converted into electronic format. Also inconsistent or not possible values will be avoided if the Web application is programmed with suitable constraints.

29. The design of the application is another essential point in order to guarantee the success of the Web Census operation. It should be designed in a comprehensive way and be easy for

people to understand. Factors such as speed, attractive visual features and the robustness of the application are also very important.

30. On the other hand, security and confidentiality of the application have to be taken into seriously. The application will have to deal with personal and private data that are sensitive. Computer attacks by hackers or other unauthorised users will need to be avoided. For this reason it is essential to have:

- (a) Passwords for authentication (perhaps including an electronic signature);
- (b) Firewalls that block unauthorized accesses;
- (c) Design of Web pages with http security protocol (https).

31. If a dwelling responds to the Census using different channels (for example Internet, paper, CATI or CAPI) then the Internet application has to be designed in such a way that responses can be synchronised. For instance, if a dwelling has answered the Census by Internet it would not make sense to send Census paper forms to that dwelling. As another example, if a dwelling has answered the Census questions by paper, it would be inappropriate for a CAPI-agent to visit that dwelling.

32. Implementation of a central database where answers from different channels are coordinated could assist in managing this issue.

33. Processing of inconsistent and incompatible answers in the Web responses of the Census questions need to be carefully studied.

34. An application with a large number of controls and messages could be difficult understand by census respondents while at the same time creating programming difficulties for the technical team. An application without controls is likely to cause inconsistent data and therefore a balance has to be reached and this is not an easy task.

35. The application that Spain designed for 2009 Pilot Census allowed for two different error types:

- (a) Critical errors (incompatibility): Users must change their responses. Otherwise, it is impossible to continue answering questions;
- (b) Other types of errors that are not critical: Users can both confirm or change their responses and then continue answering subsequent questions.

E. SWOT diagram

36. To conclude, a SWOT (Strengths, Weaknesses, Opportunities and Threats) diagram is shown below. This diagram attempts to summarize the characteristics about Census by Internet.

37. The SWOT analysis involves specifying the objective of the business project and identifying the internal and external factors that are favourable or unfavourable to achieving that objective.

<u>Strengths (S)</u>	<u>Weaknesses (W)</u>	
<ul style="list-style-type: none"> • Existence of only fixed overheads • Experience: it was possible to answer by Internet in 2001 Census and 2009 Pilot Census • Quality of information • Dissemination of data in a short period of time • Respondent burden reductions 	<ul style="list-style-type: none"> • Web application has to be safe, solid and with an attractive design • Integration among different channels • Security of the Web application • Simultaneousness of paper and Internet 	I N T E R N E T
<ul style="list-style-type: none"> • More and more people have an Internet connection • According to the economic global situation: minimizing costs • All the population will know a Census will be carried out in 2011 	<ul style="list-style-type: none"> • It is not possible to predict exactly the Web rate of response • People's knowledge of technology and familiarity with Web applications • Internet is not available in the 100% of dwellings 	E X T E R N A L
<u>Opportunities (O)</u>	<u>Threats (T)</u>	

III. NUMERICAL DATA ABOUT VIABILITY AND USEFULNESS OF CENSUS BY INTERNET IN SPAIN

38. In this section, numerical data that illustrates the veracity of the different statements explained in section 2 will be included.

A. Internet Use

39. According to the percentage of dwellings with an Internet connection (Table 1), it is reasonable to believe that by 2011; at least 60 per cent of Spanish dwellings will have their own Internet connection.

40. In addition many people are able to use the Internet at work or at a relative's or neighbour's dwelling and this increases the potential use of this channel in 2011 to a higher level than the proportion of dwellings with internet connection.

41. Differences between 2011, when Internet will be something familiar for almost everybody, and 2001, when Internet was not are clearly illustrated in the following table.

Table 1

Percentage of dwellings with an Internet connection

	2002	2003	2004	2005	2006	2007	2008
Spain	17	28	34	36	39	45	51
UE-average	39 ²	43 ²	42 ³	48 ³	51 ³	54 ⁴	60 ⁴

B. Economic view of Census by Internet

42. If we consider the Spanish experience in 2001, the costs of developing an Internet application for the Census (including the call centre) were two million Euros. If the same situation applied in 2011, then a three point five million 3.5M Euros budget would be required.

43. On the other hand, according to Spanish studies, the cost per capita of answering the 2011 Census questions by different channels will be approximately:

- (a) Paper, 5 Euros each questionnaire
- (b) CATI, 12 Euros each questionnaire
- (c) CAPI, 25 Euros each questionnaire.

44. If these two quantities are related (foreseen budget of 2011 Internet Census and cost per capita of answering the 2011 Census by different channels), the conclusions are:

- (a) With more than 700,000 answers via Internet (around 1.5 per cent of the population), Internet strategy would be worthwhile (versus paper strategy);
- (b) With more than 300,000 answers via Internet (around 0.7 per cent of the population), Internet strategy would be worthwhile (versus CATI strategy);
- (c) With more than 140,000 answers via Internet (around 0.3 per cent of the population), Internet strategy would be worthwhile (versus CAPI strategy).

45. Thus there is no doubt about viability and utility of Census by Internet.

C. 2009 Spanish Pilot Census

46. As commented in section II, the National Statistics Institute of Spain (INE) carried out a Census Pilot Survey which commenced on 13 April 2009. The purpose of the Pilot is to test different technology enterprises, new forms and methods of collecting data in advance of the next full Census in 2011.

47. The Pilot is very valuable. What is learnt from it will be used to help decide on questions, methods and other aspects of the 2011 Census.

² EU-15

³ EU-25

⁴ EU-27

48. Sixty Enumeration Areas (10 Enumeration Areas in 6 different Regions) throughout the country were selected for the Pilot. In total, about 30,000 of the approximately 15.6 million households in the State will be included in the Pilot.

49. As mentioned in section II, two different methods are being tested (both of them are mixed strategies) for collecting data. Each method is being tested in 30 Enumeration Areas. The details are presented in Table 2.

Table 2

Timetables and collecting data channels according to each method

	Paper	CATI	Internet	CAPI
Method 1	From 13 April until 31 July			
Method 2	No Paper	From 13 April until 31 July	From 13 April until 31 July	From mid June until 31 July

50. The percentage of respondents in the different approaches (methods and channels) varies within each method and channel.

51. As anticipated, paper is the channel with the highest response in method 1. Nevertheless it is surprising that almost 10 per cent of the dwellings (15 per cent adjusting data to 100 per cent of response) decided to answer the Census by Internet even though they had the Census forms at home.

52. Under method 2, the total percentage of respondents (40.0) is lower than for method 1 (62.5). However if the different channels of method 2 (CATI, Internet) are considered together with the fact that the Pilot Census was conducted without any publicity, then these response rates are not unexpected.

53. It also seems clear that the scenario to be faced in 2011 will be different from the one during 2009 Pilot test. If it is possible to motivate or convince people about Internet's advantages, the potential growth of this channel is high.

Table 3

Percentage of respondents⁵

	Paper	CATI	Internet	CAPI	TOTAL
Method 1	45.6	7.4	9.5	No data available	62.5
Method 2	No Paper	18.1	21.9 ⁶	No data available	40.0
TOTAL	20.7	13.3	16.3	No data available	50.3

⁵ Provisional data at 21 June.

⁶ In some Enumeration Areas, we obtained percentages higher than 42.5%.

Table 4

Percentage of respondents⁵ (adjusted to 100% of response in each method)

	Paper	CATI	Internet	TOTAL
Method 1	73.0	11.8	15.2	100
Method 2	No Paper	45.2	54.8	100
TOTAL	41.2	26.4	32.4	100

IV. CONCLUSIONS AND FUTURE STRATEGIES FOR IMPROVING WEB RATE OF RESPONSE

54. Although the percentage of respondents by Internet in the Pilot is higher than expected, different strategies could be applied to improve the Web rate of response for the upcoming 2011 Census.

55. In order to increase the percentage of respondents by Internet, one or more of the following ideas could be considered:

- (a) Advertising campaign in “classic media” such as television, radio or newspaper;
- (b) Advertisements in Web pages. Outsourcing of experts in viral publicity⁷ could also be useful;
- (c) Advertising in social networks like Facebook (www.facebook.com), Myspace (www.myspace.com) or Twitter (www.twitter.com);
- (d) Careful composition of the letter sent to each dwelling. This letter is the only direct contact INE makes with a high number of people. As stated earlier, outsourcing of experts could be considered;
- (e) Making it easier for people that do not have an Internet connection at home and are willing to answer the Census by Internet to have access to an internet enabled computer. Collaboration of municipalities and regions may be useful in this regard.

56. Even though a 100 per cent responses by Internet is not foreseen, INE is convinced that if all the ideas to improve the use of the Internet are effectively implemented a high percentage of responses will be obtained. The amount of money that could be potentially saved together with improved quality of the data received via the internet will be two key drivers to maximise response from the Internet.

⁷ "Viral Publicity" can be defined as a strategy to attract attention to one's company, whilst causing fans to spread one's message by themselves, thereby creating new uses of one's product, new slangy slogans, new communities of shared interest and new ways of being who one thinks.