

---

## Estadística Oficial

---

### Theoretical and Practical Analysis about Census by Internet

Jorge L. Vega Valle

National Statistics Institute - Spain

✉ jorvegaval@ine.es

#### Abstract

It has been some time since new technologies became part of everyday life. It is now very difficult to imagine life without the existence of mobile phones, laptops or Internet connection.

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 our banking information.

This paper tries to justify the usage of the Internet as a channel with a lot of possibilities (and an enormous potential of growth) for the next Census round of 2011.

**Keywords:** Census, Web rate of response, Push-methods, SWOT analysis.

**AMS Subject classifications:** 62P25.

## 1. Introduction

Population and Housing Censuses (see [1]) are among the most important statistical operations, not only because of its magnitude but also because of its 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 20<sup>th</sup> in Spanish history:

18<sup>th</sup> Century (3): 1768, 1787, 1797

19<sup>th</sup> Century (5): 1842, 1857, 1860, 1877, 1897

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

21<sup>st</sup> Century (2): 2001, 2011

Importance of Population and Housing Censuses is absolutely unquestionable. Using Census information we can know the detailed structure of the

dwelling and features about the people that live in those living quarters. Results of questions about occupation, education, migration, household and family characteristics and living quarters, dwellings and housing arrangements are really useful for planning different kind of politics (demographic, educational, health-care or environment).

Apart from that, Census results are very useful both for assigning economical sources from the State or the European Union to different kind of regions and also for having a suitable dwelling frame that can be used in the different sampling surveys of the National Statistics Institute.

During all these years, life habits, especially those related to technology and electronic devices, have changed a lot. And with the passage of time, the speed of these changes goes faster and faster.

Trying to take advantage to people's new behaviour, Spain was one of the first countries in the world that made possible to answer to 2001 Census questions by Internet. People could only choose between paper or Internet for answering to the Census forms. Alternative methods for data collection, which were not foreseen during 2001 Census, are:

- CATI (Computer Assisted Telephone Interviewing): method based on a call-center system, in which the interviewer follows a script provided by a software application capable of customizing the flow of the questionnaire depending on the answers provided.
- CAPI (Computer Assisted Paper Interviewing): similar to CAPI except that the interview takes place in person instead of over the telephone, normally enumerators with hand held devices.

Although according to the rate of response it was not a success (only 13.768 of the 14.187.169 dwellings, which represents less than a 1‰ rate, answered to the Census by Internet), all the efforts made are and will be very useful to plan the 2011 Census.

Several reasons make the situation for the 2011 Census much more promising: the amount of people that use the Internet everyday has increased a lot, bandwidth offered by IT companies is much higher nowadays than in 2001, Web applications are much more frequent and their possibilities are not so limited (for example, concepts like Internet security protocols have been especially developed lately), device storage size has grown exponentially in the last few years. . . Because of all these reasons the scenario that we will face in the 2011 Census, where the Internet will play an important role, will be very different from that in which the 2001 Census was carried out, where the existence of the Internet was symbolic.

## 2. Theoretical analysis

This section will be focused on the theoretical features of answering the Census by Internet and specifically on its strong and weak points.

### 2.1. Building a Web application: worthwhile or not?

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 we compare this prospect to the one that we would face with any other way of responding exhaustively to the Census (paper, CATI or CAPI with hand held devices), we can see in all these alternatives the presence of both fixed and variable costs.

This difference can be explained because of the number of people that answer the Census using different channels, since it involves those variable costs. For example, the amounts of printed or scanned paper, of telephone interviewers (CATI) and of enumerators (CAPI) depend directly on the number of people that make use of the different channels, and this entails a series of variable costs.

Using the Internet approach all these variable overheads disappear since the development of the Web application does not depend on the amount of people that make use of this channel<sup>1</sup>.

We are ready to obtain a very simple and valuable conclusion (in section 3 of this paper this statement will be numerically clarified): if fixed overheads of developing a Web application are not very high and the Web rate of response is appreciable, Internet channel will be worthwhile.

Another conclusion is that if we assume 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.

One crucial point is that according to the current economic global situation it seems really important that all Censuses should be designed with the maxim of minimizing their own budgets as much as possible.

### 2.2. Information: quality and possibilities

One of the most important advantages of online Censuses is the quality of the information received through this channel. Inconveniences like difficult to

---

<sup>1</sup>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.

recognize (or scan) characters and impossible or blank values in some topics could be avoided providing a suitable online application. For example, we could require that 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.

Online Censuses can also guarantee the lack of inconsistencies among different answers that belong to the same form or even among different forms.

Scanning and recognizing characters operation takes time and requires a lot of resources. Besides, the quality of the information in Census paper forms is not always guaranteed.

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

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.

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 their right answers in advance. In Spain, those pre-filled answers could be obtained through the 2001 Census and also through administrative registers.

For instance, if one question was asked in the 2001 Census, we could use the person's answer and show that answer as the predefined one to the user (or do not directly ask that question to the user and assume the information from the 2001 Census or the information from our own registers as their answer). If the user does not agree with the answer, there will be the possibility of changing it, but if the user agrees, it will only be necessary to press the confirmation button and the next question will be provided.

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 to women 16 years or older).

### **2.3. Importance of push-methods**

Push-methods can be defined as those alternatives for responding Census forms in which the online channel is opened before any other option. 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, CAPI or telephone) affects negatively to the Web percentage response.

If a high percentage of Web responses wants to be obtained, it should be considered one or maybe more of these ideas:

- The letter that is sent to each dwelling to motivate them to answer the Census by Internet should be written very carefully. Outsourcing of experts in sending suitable letters to high amounts of people should be considered. In addition, cognitive techniques to evaluate how good these letters are should be used in order to check them.
- The Internet option to answer should be opened for users before any other channels.
- Only if a dwelling expresses its preference for not answering by Internet, other channels like paper, CAPI agents or CATI will be available.

In Spain during the Pilot Census Survey that we have carried out this year (since 13<sup>th</sup> April until the second fortnight of September 2009) we tested the following two methods:

- 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.
- Method 2: Only a letter was sent to each dwelling. Firstly, users could answer both by Internet or by telephone. Secondly, information about all the dwellings that did not answer before a deadline was collected using CAPI-enumerators.

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

In section 3 of this paper the differences of Web rate of response obtained in both methods will be commented.

#### **2.4. Another key points of Census by Internet**

Concerning Census by Internet, all the answers 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. Apart from that, a lack of odd (or not possible) values will be guaranteed if the Web application is programmed with the suitable constraints.

The design of the application is another essential point in order to guarantee the success of the Web Census operation. It should be designed on a comprehensive way and very easy for people to understand. Different factors like speed, attractive visual features and the robustness of the application are also very important.

On the other hand, security and confidentiality of the application have to be taken into consideration very seriously. The application will have to deal with personal and private data that are very sensitive. Computer attacks by hackers or other kinds of users will have to be avoided. For this reason, these aspects are essential:

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

If a dwelling has the possibility of responding by different channels (for example Internet, paper, CATI or CAPI) to the Census questions then the application has to be designed in order to synchronize the responses. For instance, if a dwelling has answered the Census by Internet it would be nonsense to send Census paper forms to that dwelling. As another example, if a dwelling has answered the Census questions by paper, it would be nonsense that a CAPI-agent visited that dwelling.

Implementation of a central database where answers from different channels are coordinated could be a good practice in order to deal with this problem of concurrency.

Processing of incongruities and incompatibilities in the Web responses of the Census questions should be studied very deeply.

An application with a lot of controls and messages could be difficult to tolerate by their final users and also difficult to program by the technical team, while an application without controls could cause inconsistent data. An equilibrium has to be reached and it is not an easy task.

The application that we designed for 2009 Pilot Census implemented two different errors:

- Critical errors (incompatibility): users have to change their responses. Otherwise, it is impossible to continue answering some questions.
- Other types of errors that are not critical: users can both confirm or change their responses and then continue answering different questions.

## 2.5. SWOT diagram

To sum up this explanation, a SWOT (Strengths, Weaknesses, Opportunities and Threats) diagram is included. It tries to clarify all the characteristics already explained about Census by Internet. 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.

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

Table 4: SWOT diagram

### 3. Numerical data about viability and usefulness of Census by Internet in Spain

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

#### 3.1. Internet Use

For example, according to the percentage of dwellings with an Internet connection (Table 5), it would be reasonable to believe that in 2011, at least 60% of Spanish dwellings will enjoy their own Internet connection.

If we consider all the people that can use Internet at work or at a relative's or neighbour's dwelling, the potential use of this channel in 2011 will be really

high.

Differences between 2011, when Internet will be something familiar for almost everybody, and 2001, when Internet was not consolidated, are very clear at first glance.

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

Table 5: Percentage of dwellings with an Internet connection (see [5])

### 3.2. Economical view of Census by Internet

If we consider our own experience in 2001, the costs of developing an Internet application for the Spanish Census (including the call centre) were 2M Euros. If we assume the same situation and adapt those prices to the ones that we will face in 2011, we could be talking about a 3M Euros budget.

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

- Paper: 5 Euros each questionnaire
- CATI: 12 Euros each questionnaire
- CAPI: 25 Euros each questionnaire

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

- With more than 700.000 answers via Internet (around 1.5% of the population), Internet strategy would be worthwhile (versus paper strategy).
- With more than 300.000 answers via Internet (around 0.7% of the population), Internet strategy would be worthwhile (versus CATI strategy).
- With more than 140.000 answers via Internet (around 0.3% of the population), Internet strategy would be worthwhile (versus CAPI strategy)

Taking into account all this data, there is no doubt about viability and utility of Census by Internet.

<sup>2</sup>UE-15, refers to the 15 countries that belonged to the European Union at that time

<sup>3</sup>UE-25, refers to the 25 countries that belonged to the European Union at that time

<sup>4</sup>UE-27, refers to the 27 countries that currently belong to the European Union



### 3.3. 2009 Spanish Pilot Census

As we commented in section 2, INE carried out a Census Pilot Survey since 13<sup>th</sup> April until the second fortnight of September 2009.

The purpose of the Pilot was to test different technology enterprises, new forms and methods of collecting data in advance of the next full Census in 2011.

The Pilot was very valuable - what we learn from it will be used to help us decide on questions, methods and other aspects of the 2011 Census.

60 Enumeration Areas (10 Enumeration Areas in 6 different Regions) throughout the country have been 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.

Just as we commented in section 2, we tested two different methods (both of them are mixed strategies) for collecting data. Each method was tested in 30 Enumeration Areas. The details are presented in the next table.

	<b>Paper</b>	<b>CATI</b>	<b>Internet</b>	<b>CAPI</b>
<b>Method 1</b>	April 13 <sup>th</sup> - mid Sept.	April 13 <sup>th</sup> - mid Sept.	April 13 <sup>th</sup> - mid Sept.	No CAPI
<b>Method 2</b>	No Paper	April 13 <sup>th</sup> - mid Sept.	April 13 <sup>th</sup> - mid Sept.	mid June - July 31 <sup>st</sup>

Table 6: Timetables and collecting data channels according to each method

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

As we expected, paper is by far the channel with the highest response in method 1. Nevertheless it is surprising that almost 10% of the dwellings (16% adjusting data to 100% of response) decided to answer the Census by Internet having the Census forms at home.

According to method 2, the total percentage of respondents (46.4) is lower than in method 1 (60.9), but if we consider the different channels of method 2 (CATI, Internet) and also that we carried out a Pilot Census without any publicity, these data could be described as something normal.

It seems clear that the scenario we will face in 2011 will be very different from the one we had during 2009. If we are able to motivate or convince people about Internet's advantages, the potential growth of this channel is really high.

	<b>Paper</b>	<b>CATI</b>	<b>Internet</b>	<b>CAPI</b>	<b>TOTAL</b>
--	--------------	-------------	-----------------	-------------	--------------

<b>Method 1</b>	44.3	7.1	9.5	No CAPI	60.9
<b>Method 2</b>	No Paper	17.9	23.1 <sup>5</sup>	5.4	46.4
<b>TOTAL</b>	20.1	12.8	17.1	3.0	53.0

Table 7: Percentage of respondents

	<b>Paper</b>	<b>CATI</b>	<b>Internet</b>	<b>CAPI</b>	<b>TOTAL</b>
<b>Method 1</b>	72.7	11.6	15.7	No CAPI	100
<b>Method 2</b>	No Paper	38.6	49.8	11.6	100
<b>TOTAL</b>	37.9	24.1	32.3	5.7	100

Table 8: Percentage of respondents (adjusted to 100% of response in each method)

#### 4. Conclusions and future strategies for improving Web rate of response

Although the percentage of respondents by Internet in the Pilot is higher than what we expected in the beginning, different strategies could be done for improving the Web rate of response in the upcoming 2011 Census.

In order to increase the percentage of respondents by Internet, one or maybe more of these ideas could be considered:

- Advertising campaign in “classic media” like television, radio or newspaper.
- Advertisements in Web pages. Outsourcing of experts in viral publicity<sup>6</sup> could be very useful.
- Advertising in social networks like Facebook (<http://www.facebook.com>), Myspace (<http://www.myspace.com>) or Twitter (<http://www.twitter.com>).
- Writing the letter that is sent to each dwelling very carefully. It is the only contact that INE makes with a high number of people. As before, outsourcing of experts could be considered.
- Making things easier for people that do not have an Internet connection at home and are willing to answer the Census by Internet. Collaboration of municipalities and regions would be very useful.

<sup>5</sup>In some Enumeration Areas, we obtained percentages higher than 42.5%

<sup>6</sup>“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.

Even though a 100% percentage of responses by Internet is not foreseen, we are convinced that if all these ideas are performed in a suitable way, a really high percentage of responses would be obtained. Apart from that, the amount of money that could be saved and the quality of the data received through this channel would be two key pillars in order to make the big effort needed to achieve this goal.

### Acknowledgements

I would like to express my thanks to all my colleagues at Census Department in the National Statistics Institute of Spain, especially to Eduardo Teijeiro (teixeiro@ine.es) for being so close and also for answering all my questions.

### References

- [1] CES Recommendations for the 2010 Censuses of Population and Housing. United Nations, New York and Geneva. (2006). [http://www.unece.org/stats/publications/CES\\\_2010\\\_Census\\\_Recommendations\\\_Engli\~sh.pdf](http://www.unece.org/stats/publications/CES\_2010\_Census\_Recommendations\_Engli\~sh.pdf)
- [2] Register-based statistics in the Nordic countries. Review of the best practices with focus on population and social statistics. United Nations, New York and Geneva. (2007). [http://unstats.un.org/unsd/censuskb/attachments/Register\\\_based\\\_statistics\\\_in\\\_Nordic\\\_countries-GUID6c4bc07a61994d2fa92e7c1dd79438fa.pdf](http://unstats.un.org/unsd/censuskb/attachments/Register\_based\_statistics\_in\_Nordic\_countries-GUID6c4bc07a61994d2fa92e7c1dd79438fa.pdf)
- [3] Regulation (EC) No 763/2008 on population and housing censuses. <http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2008:218:0014:0020:EN:PDF>
- [4] 2001 Population and Housing Census. INE. (Spain). <http://www.ine.es/jaxi/menu.do?type=pcaxis&path=%2Ft20%2Fe242&file=inebase&L=1>
- [5] Survey on the Equipment and Use of Information and Communication Technologies (ICT-H) in Households. INE-2008. (Spain). <http://www.ine.es/jaxi/menu.do?type=pcaxis&path=%2Ft25/p450&file=inebase&L=1>
- [6] Handbook on Statistical Disclosure Control. <http://neon.vb.cbs.nl/casc/..%5Ccasc%5Chandbook.htm>

### About the author

**Jorge L. Vega Valle** is Program Director in the Population Department of the National Statistics Institute of Spain. His main research interests are the development of statistical infrastructure processes and the preparation of the 2011 Spanish Population and Housing Census.