

**Economic and Social Council**

Distr.: General
14 March 2012

Original: English

Economic Commission for Europe**Conference of European Statisticians****Group of Experts on Population and Housing Censuses****Fourteenth Meeting**

Geneva, 24-25 May 2012

Item 5 of the provisional agenda

Internet data collection**2011 Census in Portugal – implementation of the e-Census system****Note by Statistics Portugal***Summary*

This document describes Portugal's experience with using the Internet (e-census) to answer the 2011 Census questionnaires. This document follows up on that submitted in the twelfth meeting of the Group of Experts on Population and Housing Censuses held on 28 - 30 October 2009 (<http://www.unece.org/stats/documents/2009.10.census.html>).

I. Introduction

1. Population and housing censuses have been conducted in Portugal since 1864, in accordance with international recommendations. Up to 1960 account was only taken of the population census, which included only a few questions on housing. As of 1970 the housing content was significantly enhanced and autonomously organised, for which reason two censuses went on to be simultaneously conducted (Population and Housing). Hence, the censuses held in 2011 (2011 Census) correspond to the fifteenth General Population Census and fifth General Housing Census and have always been conducted simultaneously, given the strong relationship between the two contents.

2. The census model used has always been the traditional one, insofar as all statistical units (building, living quarters, household and individual) have been observed on an individual basis and exhaustively through specific questionnaires also treated exhaustively. Therefore, no sampling methods were used in both the collection and processing of

questionnaires. The only exception occurred in 1970 due to problems in the data processing system. Thus, most of the available results for 1970 are based on a sample of 20 per cent of the collected questionnaires.

3. Ever since the inception of the 2011 Census in 2006 it was decided that an online response system would be implemented, taking into account the growing possibilities of using this type of responses (easier access by citizens and a growing number of people who were familiar with this technological infrastructure). In turn, the known experience of a number of countries such as Canada, New Zealand and Australia was not only encouraging in terms of the population's participation in answering the census over the Internet, but it also clearly showed that, regardless of the census model, Internet use in such statistical operations clearly met the demand for comfort and privacy of many citizens.

4. The results obtained with citizens' participation in answering the 2011 Census over the Internet (around 50 per cent of the enumerated population) were quite significant and even somewhat surprising given those seen in experimental operations. These results owed to a strong commitment of the entire fieldwork structure, motivated to promote online responses, notably due to a remuneration model that did not distinguish between response types. As online responses rendered the work of enumerators easier for reducing the number of visits to housing units, they encouraged people to answer over the Internet, both individually and with the help of family members or local administrative entities.

II. Experimental operations

5. In the course of the preparation of the 2011 Census three experimental operations were conducted: two tests to the questionnaires in 2008 and 2009 and a pilot survey in 2010. Experimental operations were always held in the period of the year forecast for the final operation's census day, so as to ensure four key conditions:

- (a) Weather and population mobility conditions similar to those expected for the final operation;
- (b) Growing size and diversification of the observed population and areas;
- (c) Distribution of areas to be observed across all the main regions (7), to ensure the participation and experience of the respective regional structures responsible for the final operation;
- (d) Appropriate time span to prepare reports on each experimental operation and incorporate the results into the following operation.

6. Answers to the 2011 Census could be given online (e-census) or through paper questionnaires, which required defining a specific period to end e-census responses and approach all living quarters that had not yet answered. Hence, a more limited period of time was set out for answering the e-census, as of which it would only be possible to do so through paper questionnaires.

7. The table below shows a set of indicators on how experimental operations were organised and the main results.

Table 1

| <i>Experimental operation and year</i> | <i>Census day</i> | <i>Size (number of living quarters)</i> | <i>Data collection period</i> | | <i>Response rate to e-census</i> | <i>Average time completion on e-census</i> | <i>Final report</i> |
|--|-------------------|---|-------------------------------|-------------------------|----------------------------------|--|---------------------|
| | | | <i>E-Census</i> | <i>E-Census + paper</i> | | | |
| Test 1 – 2008 | 7 April | 6,911 | 7 April to 4 May | 14 April to 4 May | 14% | 46 minutes | July 2008 |
| Test 2 – 2009 | 20 April | 17,161 | 20 April to 10 May | 27 April 24 May | 9% | 49 minutes | July 2009 |
| Pilot Survey – 2010 | 12 April | 45,887 | 12 April to 2 May | 19 April to 16 May | 13,4% | 42 minutes | July 2010 |

8. The previous census experience is always a key element. However, given the time elapsed – usually 10 years – one cannot do without significant organised experimentation work. In the case of online responses, in addition to this being the first experience of the kind in Portugal, there was an element of additional complexity for the fact that there was no previously built file on living quarter's addresses. This required a further control of procedures to ensure identification of every response sent and accepted by the system.

9. Therefore, the preparatory stage of this statistical operation was a key experimentation element of all activities, but chiefly of the continuing assessment of developments deemed necessary to be implemented, in order to ensure high levels of participation in the e-census and the appropriate control of responses. It was also clear that the levels of e-census responses relied to a large extent on local commitment by those involved in the fieldwork. In addition, a good advertising campaign was of the essence to involve local administrative structures (municipality and civil parish offices) in supporting the population, namely by setting up "e-census counters" with equipment with Internet connection and accredited persons to help. Limiting the collection only to the e-census on the first week was a systematic procedure in experimental operations targeted at promoting online responses.

III. Brief description on the organisation of the fieldwork for the final data collection

10. Since 1981 in Portugal, these censuses have been held with the support of an important cartographic infrastructure dividing the territory into statistical areas per each enumerator, relying on cartography. In 2011 use was made of digital cartography with a 50 cm resolution, allowing for quite an easy identification of the limits of each building in the respective territory. Since there was no prior file on postal address of living quarters, the identification of buildings and living quarters within each work area was made during the questionnaire distribution process through a specific code incorporating the whole administrative and statistical location up to the housing units.

11. Organisation of the 2011 Census data collection complied with the following criteria:

(a) The census day was 21 March 2011;

(b) The distribution of questionnaires took place two weeks prior to the census day: between 7 and 20 March. During the distribution of questionnaires living quarters were identified through the registration on a cover for buildings retained by the enumerator and containing the identification of the living quarters and the envelope left with the identification and secret codes (ID/PIN) for answering over the Internet. The ID/PIN code

was calculated so as to identify the municipality (308 across the whole country) of each living quarters. Each living quarters received the following material:

- (i) Housing unit questionnaire with the identification area duly filled in by the enumerator;
 - (ii) Private household questionnaire;
 - (iii) Individual questionnaires for residents;
 - (iv) Closed envelope with the ID/PIN code of the respective municipality, containing the identification on the outside and the individual response code on the inside;
 - (v) In housing units with more than one household and in collective living quarters no codes were distributed for online responses due to the additional complexity that this might cause. However, such living quarters account for less than 1% of the enumerated population.
- (c) Questionnaires were collected as follows:
- (i) In the week from 21 to 27 March only online responses were allowed, and enumerators sought to collect questionnaires in those living quarters where such responses were not possible;
 - (ii) In the period from 28 March to 12 April simultaneous online and paper-based responses were allowed; however, preference was given for online responses whenever that intention was expressed; there was no discrimination in the remuneration of the different types of 2011 Census responses;
 - (iii) Whenever a living quarters answered the e-census, the respective enumerator would receive a text message on their mobile phone with the identification of the relevant living quarters;
 - (iv) As of 12 April only paper-based responses were allowed.

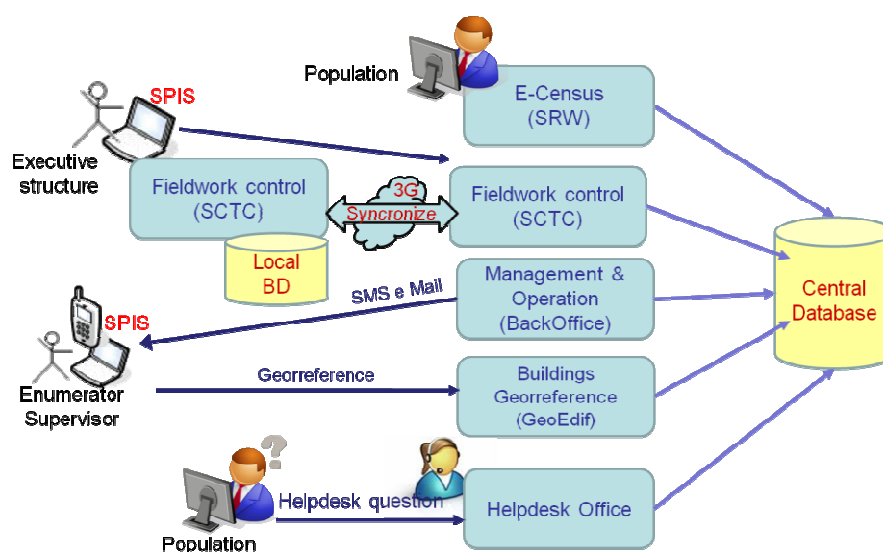
12. At the end of April around 97 per cent of the living quarters distributed were enumerated. As usual for this type of statistical operation, a small proportion of living quarters led to a considerable burden, due to rechecking works, which extended until late May. At the end of June 2011 early results for the 2011 Census (preliminary results) were made available and the provisional results were released in the beginning of December 2011.

IV. Organisation and functioning of the final technological infrastructure

A. Core technological structure

13. The 2011 Census supporting architecture was composed of the following functional applications or modules.

Figure 1



E-census or web collection system: application that made the electronic questionnaire available to housing units, respective households and individuals;

Fieldwork control: made it possible for the field staff structure to enter data on the work carried out, perform support tasks and monitor work progress;

Management and operations (Back-office): supplied information to the internal structure as well as the whole operation parameterisation and configuration;

Helpline (Helpdesk): made it possible to answer the questions and requests for assistance of respondents. A technical support website was also provided for the field structure.

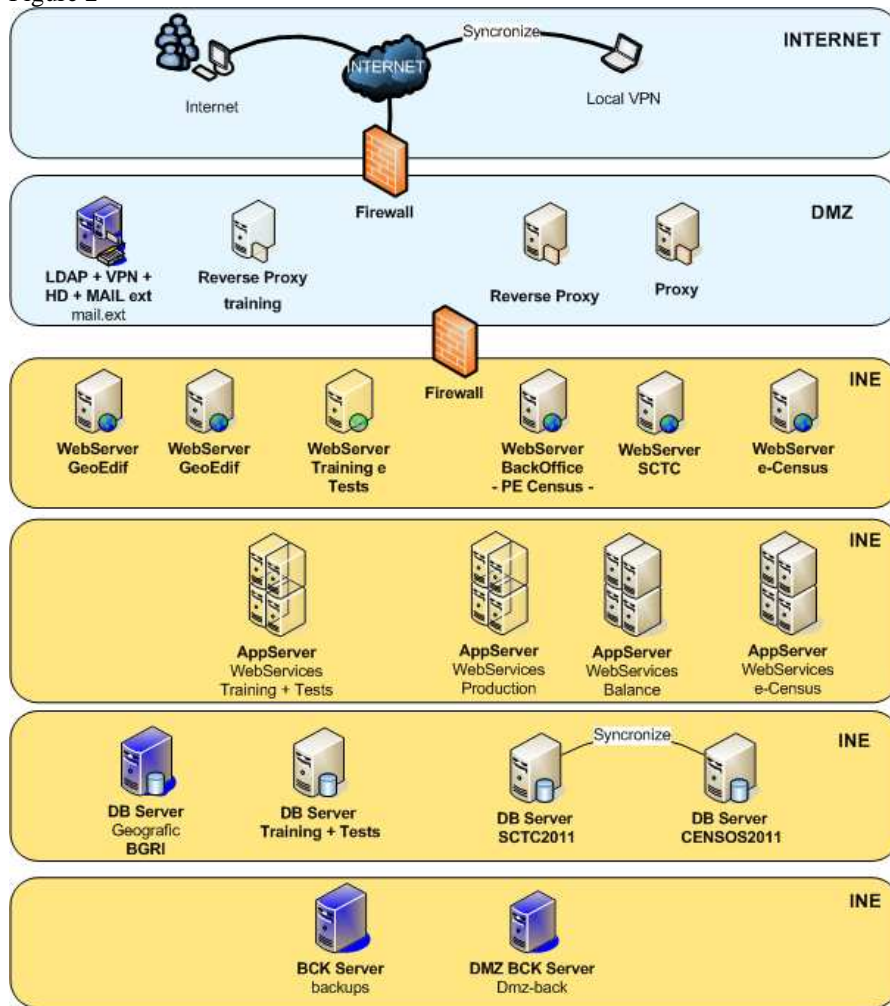
Building geo-referencing: made it possible to mark the coordinates of the listed buildings and the respective numbers of living quarters.

14. A several number of jobs were developed to trigger alerts to the fieldwork collaborators, including SMS with the information of each household that fill the e-census questionnaire.

15. The design of the technological infrastructure used was progressively consolidated in the course of experimental operations, allowing for the optimisation of a number of functional and technological features.

16. The various technological infrastructure components were subject to independent audits to their safety and reliability. When the production equipment was installed the whole structure underwent safety and stress tests. Due to its importance and criticality, the e-census application – which would be subject to a more intensive use – was tested the most.

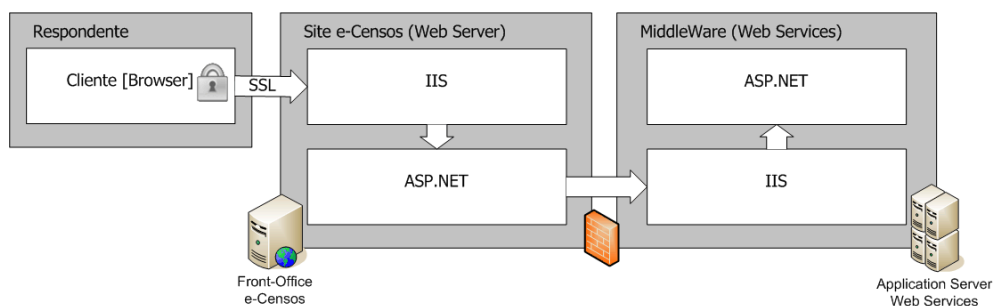
Figure 2



17. During the operation of the technological infrastructure IT support was provided to the different system components. Hardware, software and communication equipment support was reinforced and ensured through extended operating times during the e-Census answering period.

18. The image below briefly illustrates access to the e-census website, authentication in the system and the information flow.

Figure 3



Respondents accessed the e-Census website through a browser, on a secure address (HTTPS/SSL). In turn, the server supporting the e-Census website only has access (via a firewall) to the application server, when the required rules are implemented on the firewall. Only authorised servers have direct access to the application server. Webservices are implemented with recourse to WCF technology (Windows Communication Foundation).

B. Security

19. Before filling in the questionnaire, respondents went through an authentication process by introducing an access identifier (ID), with checkdigit, plus a pin code (PIN), which had been distributed to the population by the enumerator during the questionnaire delivery stage.

20. For this purpose use was made of a closed opaque envelope with the ID printed on the outside and the access ID and PIN on the inside.

21. An access identifier (ID) blocking scheme was implemented in the validation process, so as to protect the database from possible attacks and to have control over the maximum number of active sessions, thereby assuring a minimum level of quality for the service to be provided to respondents filling in the electronic form.

22. Before filling in the questionnaire, the geographic code of the living quarters (validated up to the sub-section geographic area) is also required, and so is the living quarters' address. Following correct authentication, respondents may change their PIN code.

23. Once the questionnaire is completely filled-in and sent by respondents, the Identifier/PIN pair used will no longer be accepted and respondents are informed thereof.

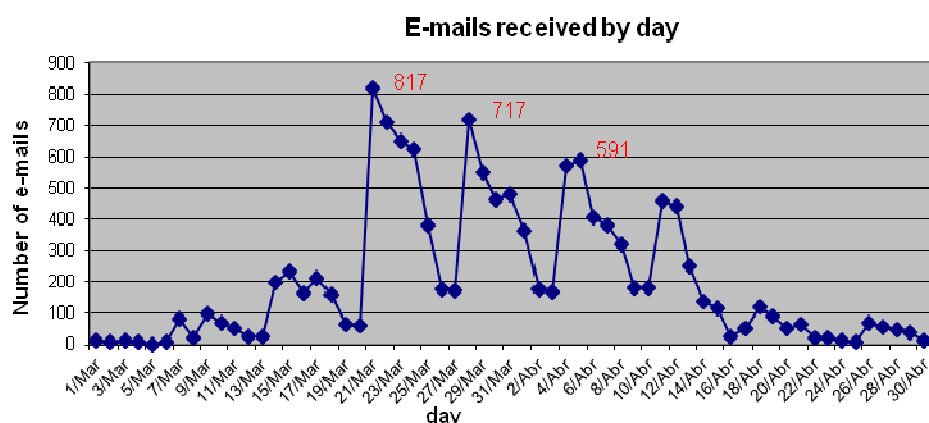
24. An error log and centralised operation system was implemented for the different applications, allowing for the monitoring and characterisation of the errors generated.

25. The security of services produced (Web services and WCFs) was implemented through a user password and the authentication among machines. A safety level was created to authenticate the issuer of the information, e.g. to guarantee that the information transferred to and from the services implemented would be secure. The access process thus had its security implemented, as well as the information amendment process.

C. Back-office and operational indicators

26. An e-mail address was provided to the population for contact and requests for assistance. Between 1 March and 30 April 12,713 e-mails were received, 81 per cent of which (10,236) during the period in which the website was open (21 March to 12 April). On average, 504 e-mails were received per day on the first week, whereas on the second week this indicator declined to 417 and on the third week it fell to 376.

Graph 1



27. Most e-mails received (70%) are related to fieldwork, and not directly to difficulties in answering over the Internet:

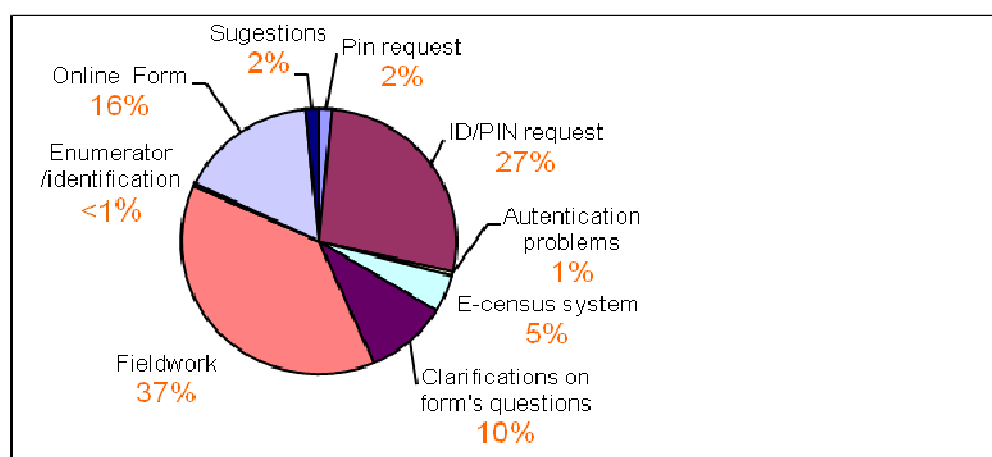
28. All e-mails were replied to. On average, the time taken to reply, per subject, was:

- Request for ID/PIN pin – reply up to 24 hours in the most critical days.
- Request for assistance related to lack of information to answer over the Internet – 48 hours at the latest
- Other types of requests – variable, but never exceeding 72 hours.

D. Helpline

29. A dedicated free helpline was available daily between 9 am and 10 pm (including weekends). From 11 March to 27 April 127,446 calls were received. The average call time was 3 minutes.

Graph 2. Main reasons for calling helpline



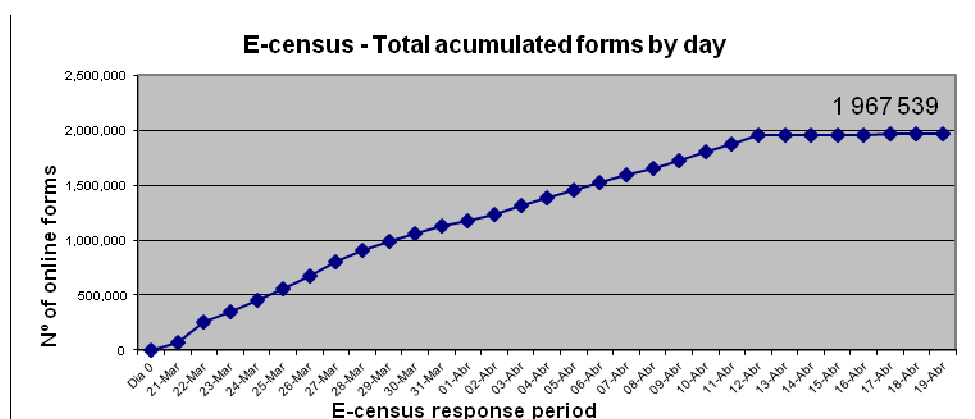
30. Around 40 per cent of online help (distribution and collection, confirmation by a staff member, suggestions) was not related to difficulties in the use of the e-census system.

V. Results

A. National and regional response rate

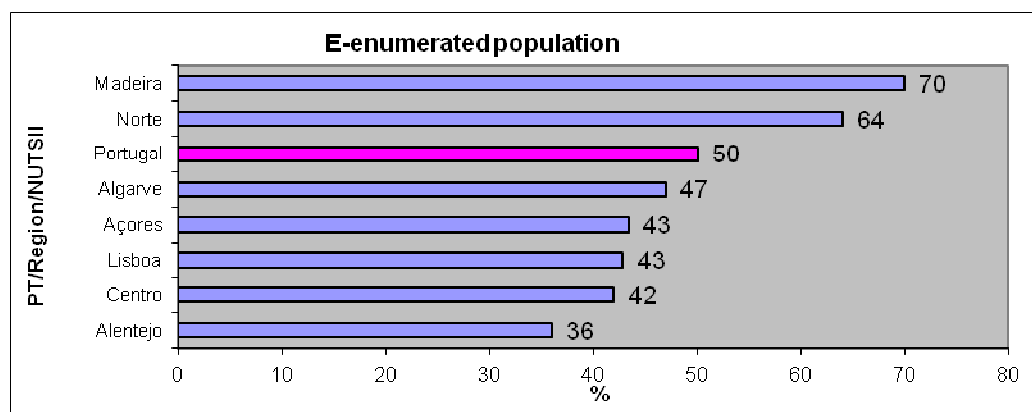
31. The online response rate amounted to around 50 per cent, corresponding to 1,967,539 housing units and 5,328,044 residents. 69,094 answers were received on the first day, and the number of answers a day was 85,500 on average.

Graph 3



32. All regions across the country experienced a high response rate, ranging between 36 per cent and 70 per cent.

Graph 4

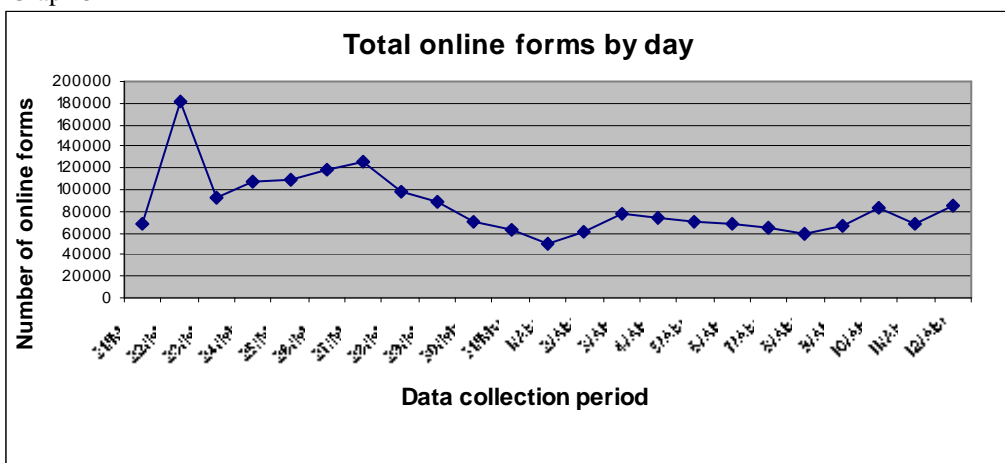


B. Trend of online responses

33. The trend of online responses over the 23 days in which the website was open was similar to that seen during the preparation tests. A considerable share of responses was given on the first few days, and the period of the day with the greatest number of visits to the website was after dinner, between 9 pm and 11 pm.

34. The number of responses sent was more intense on the first week, i.e. around 115 thousand responses on a daily basis. The peak was recorded on the second day, with 182,046 responses sent, i.e. more than twice the number of the first day (70 thousand).

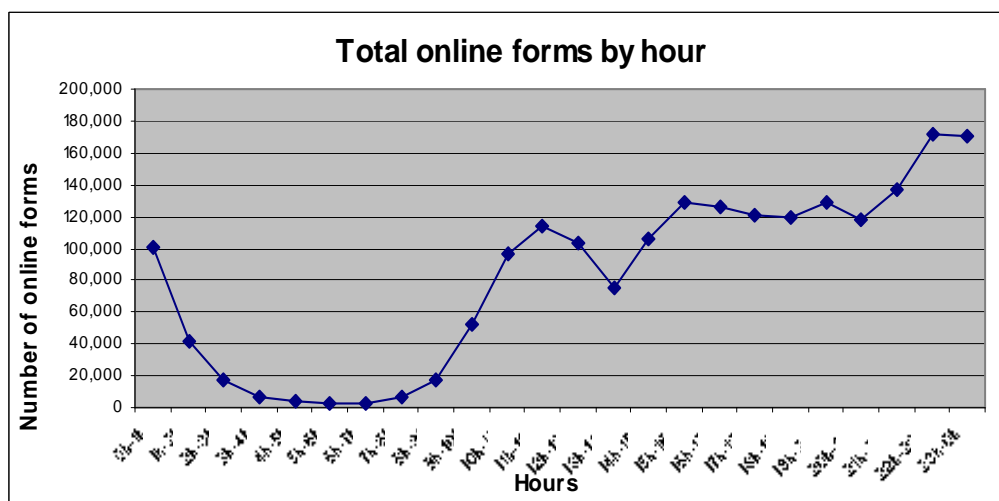
Graph 5



35. In the two following weeks the pace slowed down to a daily average of 72,000 responses (on average, 33 thousand less responses sent per day than in the first week). At the end of the collection period and with the extension of the deadline there was no peak of responses on the last day.

36. The curve of responses throughout the day shows a greater number of visits during the 9 pm-11 pm period, which concentrated around one fourth of the responses of the day.

Graph 6



37. The online collection period, initially forecast to take place over three weeks (from 21 March 2011 to 10 April 2011) was extended for a further two days until 12 April. This extension was chiefly aimed at dampening the peak effect characterising the last day, as well as at allowing more people to answer the 2011 Census over the Internet.

C. Average completion time

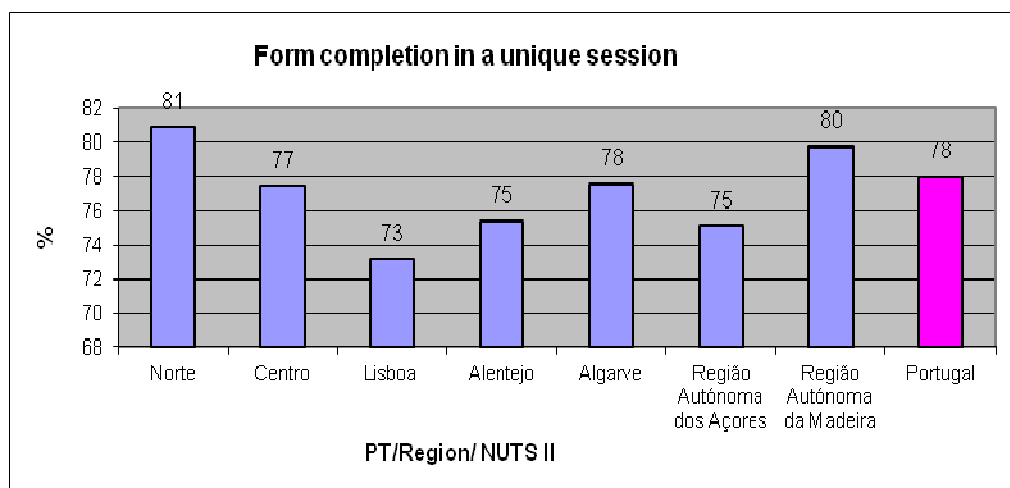
38. The average time spent completing the electronic questionnaire was 45 minutes. This indicator is slightly higher than in the previous tests, namely in the pilot survey (42 minutes), and may be due to slowness of the website on a few days and at certain hours in

the day. The regional average time ranges from 40 minutes in the Norte region to 53 in the Lisboa region.

D. Responses sent on a single session

39. Approximately 78% of responses – i.e. 1,526,924 – were filled in on a single session. This means that most of the population started and finished completing the 2011 Census questionnaire on a single occasion. The Save and return later functionality was used in 22 per cent of the responses sent and allowed respondents to interrupt completion of the questionnaire and resume it at a later stage.

Graph 7



E. Replies to the opinion poll

40. An opinion poll was held with the population answering online to the 2011 Census, composed of only two questions. Answering the questionnaire was optional, and therefore results only represent the opinion of the respective respondents.

Table 2

Results of the online evaluation questionnaire

| | Yes | % | No | % | NA | % | TOTAL |
|--|---------|------|-------|-----|-------|-----|---------|
| Did you pleased to answer the 2011 Census online? | 192.067 | 94,5 | 9.567 | 4,7 | 967 | 0,5 | 202.601 |
| Do you intent to use internet to answer to other statistical survey? | 196.673 | 97,1 | 4.915 | 2,4 | 1.013 | 0,5 | 202.601 |

41. 202,601 living quarters answered this survey, corresponding to a participation rate of 10%, taking as reference the number of online responses to the 2011 Census (1,967,539).

42. Results show that 95% of respondents to this optional survey were pleased for having answered the 2011 Census over the Internet and 97% would choose to answer Statistics Portugal surveys in such manner.

VI. Main conclusions

- The response rate – around 50 per cent – shows a strong participation of the population in the online census data collection.
 - The responsiveness of the implemented system proved suitable to actual demand.
 - The experimental process developed in 2008, 2009 and 2010 led to a good proxy and materialisation of the operation's final design.
-