

**Economic and Social Council**

Distr.: General
19 April 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 for England and Wales: Internet Data Capture****Note by Office for National Statistics, United Kingdom¹***Summary*

For the first time in England and Wales, the 2011 Census of Population offered households and individuals the opportunity to complete their return via the internet, as an alternative to completing a traditional paper questionnaire. The Office for National Statistics (ONS) hoped to achieve around 25 per cent of all census returns through the Internet Data Capture (IDC) system.

This paper outlines how the IDC questionnaires were developed, considering both the challenges faced and the opportunities available to ease respondent burden, and gives an early assessment of how successful it was.

Similar arrangements were made to collect census returns on-line in Scotland and Northern Ireland in order to provide a consistent UK approach.

I. Development

1. This section of the paper discusses some of the key challenges faced when developing the IDC questionnaires, together with some of the improvements it was possible to introduce to reduce respondent burden and improve data quality.

¹ Prepared by Peter Stokes.

A. Background

2. The IDC questionnaires were not developed at the same time as the traditional paper equivalent. Work did not start until the content and design paper questionnaires were almost finalised. Although a range of questionnaires were developed, for completion by households, individuals and communal establishments, the version discussed in this paper is that used to collect information about households, as this is the “standard” questionnaire and was the main focus of development and testing. This questionnaire constituted 32 A4 pages, used to collect information about the property, a maximum of six residents and their relationships to each other.

3. The paper questionnaires to be used in the 2011 Census were developed through an extensive programme of consultation and testing², which gave ONS a high level of confidence that they were understood by respondents and would accurately collect the information intended. Given this confidence, and concerns about potential modal bias that could be introduced if the IDC questionnaires were significantly different, it was agreed that the two modes should be very similar from a respondent’s perspective.

4. This presented some challenges, as consideration was not given to the IDC version when developing the paper questionnaires, and some aspects were not easily transferable. However, it was also possible to make relatively small changes which would reduce the burden on respondents and improve data quality.

B. Moving from paper to internet

5. The main paper questionnaire used in the 2011 Census was developed to make optimal use of the space available, with questions and relevant supporting information presented across the two open pages visible to the respondent where appropriate. It would not prove possible to present as much information on a single page of the IDC, and some significant design changes were necessary. The most significant change in this respect was to the question asking relationship between household members, which used two full pages of the paper questionnaire as shown below:

² <http://www.ons.gov.uk/ons/guide-method/census/2011/the-2011-census/2011-census-questionnaire-content/final-recommended-questionnaire-content-for-england-and-wales.pdf>

Household questions - continued

H6 How are members of this household related to each other? If members are not related, tick the 'Unrelated' box.

- ⌕ If there are more than six people, contact us to request a Continuation Questionnaire
- ⌕ If you live alone ➔ Go to **H7**
- ⌕ If no-one usually lives here and there are no visitors staying overnight here on 27 March 2011, answer questions H7 to H11 on page 6 and then go to the Declaration on the front page

For Person 5 (James), there is a tick next to 'Son or daughter' in the columns for Persons 1 and 2 to show he is the son of Robert and Mary. Columns 3 and 4 show he is the brother of Persons 3 and 4 (Alison and Stephen).

Example: This shows how a household with two parents and four children are related to each other

Name of Person 1 First name Last name	Name of Person 2 First name Last name	Name of Person 3 First name Last name	Name of Person 4 First name Last name	Name of Person 5 First name Last name	Name of Person 6 First name Last name
ROBERT SMITH	MARY SMITH	ALISON SMITH	STEPHEN SMITH	JAMES SMITH	SARAH SMITH
How is Person 1 related to Person 2: → 1					
How is Person 2 related to Person 3: → 1 2					
How is Person 3 related to Person 4: → 1 2 3					
How is Person 4 related to Person 5: → 1 2 3 4					
How is Person 5 related to Person 6: → 1 2 3 4 5					
Husband or wife	<input checked="" type="checkbox"/>	Husband or wife	<input type="checkbox"/>	Husband or wife	<input type="checkbox"/>
Same-sex civil partner	<input type="checkbox"/>	Same-sex civil partner	<input type="checkbox"/>	Same-sex civil partner	<input type="checkbox"/>
Partner	<input type="checkbox"/>	Partner	<input type="checkbox"/>	Partner	<input type="checkbox"/>
Son or daughter	<input type="checkbox"/>	Son or daughter	<input checked="" type="checkbox"/>	Son or daughter	<input checked="" type="checkbox"/>
Step-child	<input type="checkbox"/>	Step-child	<input checked="" type="checkbox"/>	Step-child	<input type="checkbox"/>
Brother or sister	<input type="checkbox"/>	Brother or sister	<input type="checkbox"/>	Brother or sister	<input checked="" type="checkbox"/>

Using the same order you used in question H3 (page 3), write the name of everyone who usually lives here at the top of each column. Remember to include children, babies and people who have requested an individual questionnaire

Tick a box to show the relationship of each person to each of the other members of this household

Name of Person 1 First name Last name	Name of Person 2 First name Last name	Name of Person 3 First name Last name	Name of Person 4 First name Last name	Name of Person 5 First name Last name	Name of Person 6 First name Last name
How is Person 1 related to Person 2: → 1					
How is Person 2 related to Person 3: → 1 2					
How is Person 3 related to Person 4: → 1 2 3					
How is Person 4 related to Person 5: → 1 2 3 4					
How is Person 5 related to Person 6: → 1 2 3 4 5					
Husband or wife	<input type="checkbox"/>	Husband or wife	<input type="checkbox"/>	Husband or wife	<input type="checkbox"/>
Same-sex civil partner	<input type="checkbox"/>	Same-sex civil partner	<input type="checkbox"/>	Same-sex civil partner	<input type="checkbox"/>
Partner	<input type="checkbox"/>	Partner	<input type="checkbox"/>	Partner	<input type="checkbox"/>
Son or daughter	<input type="checkbox"/>	Son or daughter	<input type="checkbox"/>	Son or daughter	<input type="checkbox"/>
Step-child	<input type="checkbox"/>	Step-child	<input type="checkbox"/>	Step-child	<input type="checkbox"/>
Brother or sister	<input type="checkbox"/>	Brother or sister	<input type="checkbox"/>	Brother or sister	<input type="checkbox"/>
Step-brother or step-sister	<input type="checkbox"/>	Step-brother or step-sister	<input type="checkbox"/>	Step-brother or step-sister	<input type="checkbox"/>
Mother or father	<input type="checkbox"/>	Mother or father	<input type="checkbox"/>	Mother or father	<input type="checkbox"/>
Step-mother or step-father	<input type="checkbox"/>	Step-mother or step-father	<input type="checkbox"/>	Step-mother or step-father	<input type="checkbox"/>
Grandchild	<input type="checkbox"/>	Grandchild	<input type="checkbox"/>	Grandchild	<input type="checkbox"/>
Grandparent	<input type="checkbox"/>	Grandparent	<input type="checkbox"/>	Grandparent	<input type="checkbox"/>
Relation - other	<input type="checkbox"/>	Relation - other	<input type="checkbox"/>	Relation - other	<input type="checkbox"/>
Unrelated (including foster child)	<input type="checkbox"/>	Unrelated (including foster child)	<input type="checkbox"/>	Unrelated (including foster child)	<input type="checkbox"/>

ENTER NAME OF PERSON 1 HERE AS IN QUESTION **H3**

6. In situations like this, where it was not possible to replicate the paper questionnaire design, an IDC-specific solution was developed. In this example, the requirement to develop a new version of the question on relationships, which is the most cognitively complex on the paper questionnaire, gave an opportunity to make completion easier for the respondent and reduce the likelihood of them making an error.

7. This was achieved by personalising the question using names of household members, which had already been collected, as shown below:

2011 Census - Microsoft Internet Explorer

File Edit View Favorites Tools Help

Address file:///H:/Prototype_v1.3/Idc-england-content/household/h6.html#

Household Questions

- H1 - H2 Finished
- H3 Finished
- H4 - H5 Finished
- H6 In progress**
- H7 - H8
- H9 - H10
- H11
- H12
- H13
- H14

Individual Questions

Submission

H6 How are members of this household related to each other? If members are not related, tick the Unrelated box.

Help with question H6

Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua. Ut enim ad minim veniam, quis nostrud exercitation ullamco laboris nisi ut aliquip ex ea commodo consequat.

To get more help, use the link at the top right of this page called "Help (opens in new window)"

Person 2: Rene Smith

How is Rene Smith related to Alex Smith

Please select from the list below

Person 3: Steven Smith

How is Steven Smith related to Alex Smith

Please select from the list below

How is Steven Smith related to Rene Smith

Please select from the list below

Person 4: Gail Smith

How is Gail Smith related to Alex Smith

Please select from the list below

8. While this example gave an opportunity to improve the question, it was not possible to modify other long questions, such as those for ethnic groups and qualifications, in the same way. These questions were not always viewable on the IDC questionnaire without scrolling (depending on the monitor and resolution used), and it was necessary to test these to ensure that response options at the bottom of the question were not missed.

9. For the majority of questions, it was possible to replicate them exactly on the IDC questionnaire, although a number of subtle changes were made to reduce the risk of respondent error and improve data quality. These changes are discussed in the next section.

C. Quality improvements

1. Name

10. The paper questionnaire required respondents to record the names of household members three times: Once when listing all household members to ensure coverage, again when recording relationship and finally when completing the main census return for each individual. Although this is usually completed correctly, it can cause inconsistencies if respondents change the order of household members on the questionnaire between the relationship question and the questions about the individual. For the IDC questionnaire it was possible to remove this risk by collecting names only once, and auto-completing the information in subsequent questions.

2. Radio buttons

11. The majority of questions ask respondents to give a single response from the options provided, but this instruction is not always followed. Multiple responses were often received when respondents selected an option before finding that a subsequent option was more accurate. An example of this in the England and Wales census was where respondents recorded that they were “married” but then also recorded that they were “Separated, but still legally married”. The IDC questionnaire was developed using radio buttons for these questions, to ensure that only one response was received.

3. Automatic routing

12. The 2011 Census questionnaire for England and Wales contained 42 questions for each household member, but respondents were not required to answer all of them because not all were relevant, depending on their age and answers to previous questions. However, some respondents missed these instructions on the paper questionnaire and answered questions unnecessarily. For example, children under 16 were not required to answer questions about employment status, but were sometimes recorded as ‘unemployed’. This caused a problem during data analysis, as it was not always clear, for such a child, whether the employment questions had been completed improperly, or that their date of birth had been recorded incorrectly. This issue was resolved by automatically routing respondents past any unnecessary questions, but stating both that this has been done and why, to give an opportunity to change the earlier question if in error.

4. English and Welsh

13. Respondents in Wales were entitled to complete their Census questionnaire in either English or Welsh, and were sent two paper questionnaires, one in each language.

14. The IDC questionnaire included a facility to toggle between languages, enabling different household members to use the language of their choice rather than complete separate questionnaires. This facility was also useful for respondents who did not

understand some of the words used in one language or the other, as they were able to switch between them for a translation.

5. Large households

15. The paper questionnaires included questions for up to six household members, and an additional questionnaire was required for larger households. The IDC system was not limited in this way. This significantly facilitated easier completion for large households.

III. Issues with a mixed-mode census

16. This section of the paper outlines some of the problems experienced in the England and Wales census that resulted from allowing respondents to complete their return using either the paper questionnaire or the internet, and how they were resolved.

A. False responses

17. False responses were a significant problem in the 2001 UK Census, where respondents crossed out pages that were not relevant to them (typically the pages for additional people in small households), and a process was developed to identify and remove these records from the data early in data processing. This process worked well, and was modified slightly for use in the 2011 Census, with genuine person records identified using the following criteria:

- a) Respondent must have answered 'name' OR 'date of birth' in the Individual section; *and*
- b) must have answered one other of: 'name' (in either the individual section or in household member table), 'date of birth', 'sex' or marital status;

18. 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.

19. This system worked as planned for paper questionnaires, but did not always cope well with IDC responses. It was decided that, when the IDC system was to close down, all responses that had not been formally submitted by the respondent should be 'force submitted', to ensure that the data already keyed were not lost if respondents failed to submit the return.

20. When the census data were received for processing it became clear that some respondents had opened their IDC questionnaire and entered their name before deciding not to continue and used the paper questionnaire instead. As the name fields were auto-completed, as outlined in section 2.3, this resulted in them being considered genuine responses because they were considered to have answered 'name' in both the Individual section and the household member table.

21. This resulted in a large number of records containing no data about the respondent, which caused a problem in later processing, particularly item-imputation. As a result, the process was modified to require that at least one other of the key questions was answered for a response to be considered genuine, and the data were reprocessed.

B. Multiple responses

22. The availability of the IDC system significantly increased the number of multiple responses (where respondents recorded themselves on the Census more than once at the same address), and the number of these was increased further when partial IDC responses were ‘force submitted’.

23. Multiple matches were identified by comparing all questionnaires received for each household and the people within them. Duplicate person responses were identified by matching on name, sex and date of birth, with probable duplicates also identified by matching names using Soundex and Levenshtein distances.

24. Where multiple responses were identified, the more complete record was retained and the other(s) discarded. Any questions answered on the discarded record but not on the retained one, were merged into the final record.

C. Addresses and other text responses

25. The paper questionnaire included a number of questions relating to addresses and others requiring text responses regarding ‘occupation’ and ‘industry of employment’. The number of response boxes provided was intended to maximise the value of the information collected within the space constraints of the questionnaire. As the same data capture and coding system was used for both modes of response, the same constraints were placed upon the IDC questionnaire.

26. This decision also reflected a concern that, if IDC respondents could describe their occupation and industry using as many words as desired, rather than being constrained as on paper, a modal bias could have been introduced in how these questions were answered and coded.

27. This restriction was not well received by some respondents, who felt it was unnecessary, and which resulted in additional calls to the census helpline.

IV. Results

28. This section gives an early analysis of how well the IDC questionnaire was completed, and who chose to complete their census return in this way. As, at the time of writing, data processing and analysis are still underway, the information in section is, indicative and provisional and may not reflect the results published later in the 2012.

A. Who used the IDC questionnaire?

29. ONS had set a target of 25 per cent of all responses to be received through the IDC system, but did not achieve this; only around 15 per cent of households used this mode. This is likely to be due to a number of factors including:

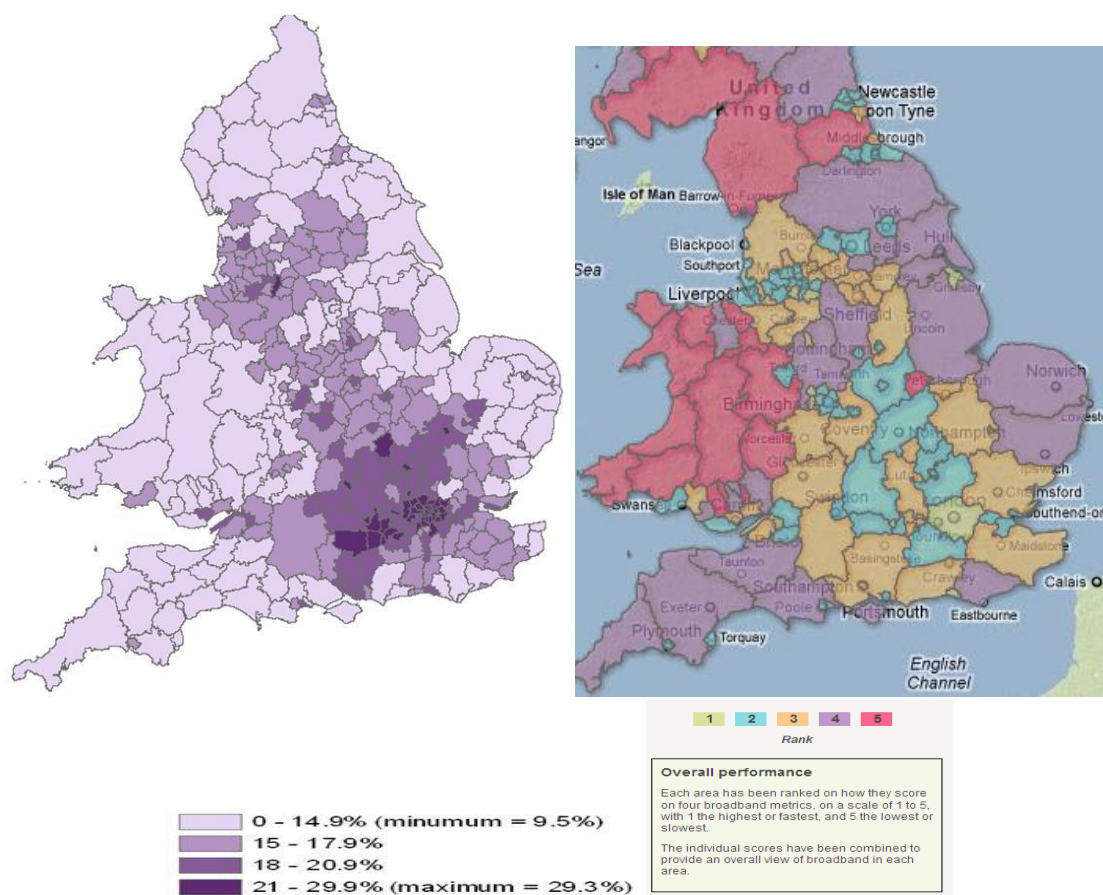
- All households were given their unique IDC code printed on their paper questionnaire, because it was essential to make sure the paper alternative was available to all.
 - This is likely to have increased paper response, as it would have been perceived to be ‘easier’ than turning a computer on.

- The Census media strategy was designed to maximise overall response, and focused on encouraging this and highlighting the benefits to the population of a successful census, rather than specifically encouraging on-line responses.
- The IDC questionnaire could only be easily completed using a broadband connection.
 - Currently only 50 – 80 per cent of households in each region of the UK have a broadband connection.

30. The affect of this last point on overall IDC response can be seen in the correlation between the two maps³ below, which show the level of IDC responses and the level of broadband take up respectively.

31. However, despite the clear link between broadband availability and IDC use, it was not the only factor to have a significant effect. An analysis of the data, focussing on the characteristics of the first person in each household using the IDC system (as they are most likely to have completed the questionnaire), found that a number of these had a significant effect on the likelihood of doing so.

32. The most significant variable, perhaps unsurprisingly, was age, where around 27 per cent respondents aged 26-35 used the IDC system, declining gradually with only 6 per cent of respondents aged over 75 doing so.



³ Broadband performance information from OFCOM: <http://maps.ofcom.org.uk/broadband/>

33. As other variables investigated are correlated with age (for example, marital status and tenure of accommodation), this was controlled for in subsequent tests. However, significant effects were still found on the likelihood of using the IDC questionnaire by a number of variables including: marital status, where people married or in a civil partnership were most likely to do so; qualifications, where higher levels made it more likely; and country of birth, where respondents born outside the UK were more likely to use the IDC system. The significance of country of birth may be because respondents born outside the UK were able to use the internet to translate some questions when completing their census returns.

34. The other factor significantly affecting the likelihood of using the IDC system was household size. As discussed in section 2.3.5, the paper questionnaires were restricted to six respondents, with an additional questionnaire required for larger households. As a result, a significant majority of households with more than six residents chose to use the IDC questionnaire, rather than request and complete more than one paper alternative.

B. Data quality

35. The IDC data contained very low levels of item non-response across all variables when compared with the data from paper questionnaires and with the 2001 Census returns.

36. Although item non-response was lower in the 2011 Census generally, reflecting the improvements made to questionnaire design and navigation, the IDC data were very well completed, despite it being possible to leave questions unanswered. Although the IDC questionnaire did not force respondents to answer all relevant questions, a warning was displayed for each unanswered question, to encourage response. This was considered an effective way of improving item level response, without significantly impairing user experience, and has been successful.

37. In addition to the low levels of item non-response, data quality was also improved by the use of radio buttons to prevent multiple response options being selected in a single question, and typed text being easier to code than hand-written equivalent from the paper questionnaires.

V. Conclusions

38. Although analysis is still underway, and the first results of the 2011 Census for England and Wales will not be published until Summer 2012, it is possible to draw a number of conclusions about both the success of the IDC system and the factors that influenced its use.

39. The IDC questionnaire was not used by as many respondents as anticipated, with around 15 per cent of households using it against a target of 25 per cent, but the quality of data collected appears to be very good, with low levels of item non-response across all variables.

40. A number of factors influenced the likelihood of respondents choosing to use the internet to complete their census return, and a number of these could be further investigated to help increase the level of IDC use in any future census. For example, understanding why factors such as country of birth increase the likelihood of a respondent using the IDC system, would help to target publicity to highlight the benefits to the user of completing their return in this way.

41. Although it is the factor which has the greatest influence on IDC use, the availability and speed of broadband internet, cannot not be controlled by ONS or other NSIs. It could, however, be used to help model likely use of such a system in future.