

## **CENSUS ADVISORY GROUP**

### **AG (10) 03 Outputs - Consultation, Evaluation and Statistical Disclosure Control**

#### **2011 Census Outputs – Consultation, Evaluation and Statistical Disclosure Control for England and Wales**

This paper has been produced to update Census Advisory Group members of the progress made in relation to the consultation on 2011 Census outputs, and to inform members of the proposed plans for further work over the next period. This will include:

- further consultation, and;
- revising the main body of statistical outputs taking into account responses from the initial round of consultations and statistical disclosure control issues.

#### **Action requested of Advisory Groups:**

**Advisory group members are asked to note the good progress made in relation to consultation with users, particularly the successful conclusion to the first phase on the main body of statistical outputs and the publication of an Output Strategy for England and Wales.**

**Advisory group members are invited to comment on current thinking with regard to statistical disclosure control and the production of outputs.**

**Questions and comments will be welcome at the meeting or sent within three weeks to:**

**Joe Traynor  
Office for National Statistics  
Segensworth Road  
Titchfield  
Fareham  
Hants  
PO15 5RR  
Email: [Joe.Traynor@ons.gsi.gov.uk](mailto:Joe.Traynor@ons.gsi.gov.uk)**

## **2011 Census output consultation**

### **1. Introduction**

Advisory group members have previously been informed of the 2011 Census consultation plans and the timetable for delivery. The project is pleased to report that good progress has been made over the last period with little or no slippage in the timetable. This paper will, as a result, inform members of those achievements and outline the future programme of work. Advisory group members are encouraged to respond on any aspect of the paper making recommendations or suggestions for further development.

### **2. Work to date**

#### **2.1 2009 Regional consultation event**

To publicise the launch of the consultation on 2011 Census outputs, six regional events were held in October 2009 in England and Wales. The presentations were well received and are available online for download<sup>1</sup>.

#### **2.2 Publication of the 2011 Census Output Strategy**

The 2011 Census Output Strategy<sup>2</sup> was published alongside the consultation documentation on the main body of statistical outputs. The strategy discusses the context and business objectives behind the strategy before going on to discuss the consultation process, vision, and technical solutions that will define 2011 Census outputs.

#### **2.3 Main statistical output**

Between June and November 2009 a team of researchers developed a preliminary set of specification defining the main outputs to be produced following the 2011 Census. The development of these specifications went through several iterations including topic leads and the Census Output Working Group (COWG). These ‘baseline’ specifications were included in the consultation documents and disseminated alongside a user guide, a series of table layouts and a user feedback template.

The consultation ran from the 14 December 2009 until the 26 March 2010 in accordance with the National Statistics Code of Practice. Its aim was to allow census users to contribute to the definition of the main body of outputs. Users were asked to comment on the specification whilst being free to add any additional requirements they wished including, new tables, geographies, derivations and classifications. The exercise was publicised widely and to-date responses from over seventy organisations have been received. Following the end of the consultation period a number of requests were received for additional time to submit responses. As a result, the consultation team informally accepted late responses until the 16 April 2010. Feedback from the consultation has been positive with many users commenting on the usefulness and accuracy of the documentation.

---

<sup>1</sup> <http://www.ons.gov.uk/census/2011-census/consultations/index.html>

<sup>2</sup> <http://www.ons.gov.uk/census/2011-census/consultations/open-consultations/2011-output-consultation---main-statistical-outputs/2011-census-outputs-strategy.pdf>

## **2011 Census Outputs – Consultation, Evaluation and Statistical Disclosure Control for England and Wales**

### **3 Forthcoming work**

The next six months will be a key period in defining the outputs for the 2011 Census. The work will include analysis of the responses from the aforementioned consultation and will be used to define a second set of specifications. This second set will be used later in the year to complete a shorter second round of consultations to confirm that users are content.

The next six months will also see a requirement gathering exercise for information on internal requirements within ONS, and a separate consultation on secondary population bases.

#### **3.1 Main statistical outputs – Evaluation phase**

The evaluation and redefinition of 2011 census outputs will be a complex procedure. The evaluation needs to take account of:

- disclosure control;
- new delivery mechanisms - including the degree of flexibility that users will be offered;
- the level of detail requested by users, and;
- a separate set of evaluations at each geographical level for which outputs will be produced.

The evaluation will take place in two phases.

The first phase will consider the redesign of the baseline specification in line with user requests. The analysis will segment responses into the different user communities and will take account of the level of detail users have requested as well as any statistical disclosure control issues. The second phase will separately consider the outputs in relation to disclosure and the different delivery mechanisms available (standard tables or hypercubes) including any compromise between the provision of flexible outputs and the amount of detail required.

The time required to evaluate such a complex set of relationships is at present unknown. Initial estimates for such an evaluation range between three and six months. Once the evaluation process has been completed, a second round of consultation will take place (late 2010) to ensure that an acceptable balance has been achieved and that user requirements have been satisfied to the best of ONS' ability, bearing in mind the possibility of competing priorities. Following the second round of consultations, it is expected that minor revisions will be made to the specifications and a set of planned 2011 Census outputs will then be issued to users.

## 2011 Census Outputs – Consultation, Evaluation and Statistical Disclosure Control for England and Wales

### 3.2 Disclosure control

The starting point for the evaluation of outputs must include a practical consideration of disclosure control. Members will be aware that record swapping, as the primary disclosure control method for the 2011 Census was agreed by UKCC in August 2009. Work has continued in this area, including consideration of the interaction between record swapping with geographic thresholds, and the provision of increased flexibility with the introduction of hypercubes in addition to static tables. Underlying this work is the need for clarity over the level of 'sufficient uncertainty' required to provide an acceptable UK SDC Policy position.

Central to any description of 'sufficient uncertainty' is a requirement to consider the underlying attributes which would lead to information being disclosed about an individual or organisation together with the nature of that disclosure. With this in mind SDC methodologists have come up with a classification system of census variables using two descriptors:

- visible variables** where attributes of the variable are 'constantly or frequently in the public view'
- sensitive variables** which include 'information of a delicate nature that could cause harm or distress if revealed'

Visible variables are effectively a key by which an individual can be identified. Sensitive variables are the ones to which ONS would want to give particular attention. The risk of disclosure increases when both visible and sensitive variables appear in the same table or hypercube; the visible variables increase the risk of identification which can then lead to disclosure of the sensitive variables, and the presence of sensitive variables may increase the probability of an intruder being motivated. Thus with combinations of visible and sensitive variables more 'uncertainty' is required and can be provided via swapping records or rules around output design. The risk of disclosure is less when there are no visible variables in the table (hence risk of identification is reduced) or when there are no sensitive variables (one might be able to identify an individual but can only disclose non-sensitive variables, and the probability of an intruder being motivated is lower).

Some preliminary work has been undertaken to classify Census variables as being visible (or not) and sensitive (or not). Note, only individual rather than household variables are shown as an example in Table 1 which is a preliminary outline. Members are invited to consider and respond with suggestions on which information carried by the census should be considered as sensitive or visible. Members should also be aware that as the number of variables classified as 'visible' or 'sensitive' increase the amount of statistical utility and or flexibility and level of detail in the final outputs will reduce.

## 2011 Census Outputs – Consultation, Evaluation and Statistical Disclosure Control for England and Wales

**Table 1 - categorisation of variables**

| <b>Individual Variables</b> | <b>Sensitive</b>  | <b>Not Sensitive</b>  |
|-----------------------------|---|---|
| <b>Visible</b>              | Long-term limiting illness/disability<br>Ethnic group<br>Religion<br>Economic activity status<br>Workplace address<br><b>A</b>  | Marital status<br>Age<br>Sex<br>Student status<br>Second residence type<br>Welsh<br>Main language<br>English ability<br>Method of travel to work<br><b>B</b>  |
| <b>Not Visible</b>          | Country of birth<br>Intended length of stay in the UK<br>Second residence address<br>National identity<br>General health status<br>Non work status<br>Ever worked<br>Job title (occupation)<br>Industry<br>NS-SEC<br><b>C</b> | Date of arrival in the UK<br>Passports held<br>Address 1 year ago<br>Hours care-giving<br>Qualifications<br>Seeking work<br>Availability for work<br>Waiting to start work<br>Self-employed<br>Employee supervision<br><b>D</b> |

The definition of visible used in the categorisation above assumes that the variable is 'in the public view', so essentially public knowledge and known widely. Of less concern is 'private' knowledge, whereby an individual can ascertain some characteristic of a husband/neighbour/friend because they have privileged 'private' knowledge of that person. If we give concern to 'private' knowledge in the design of outputs then the scope of visible variables increases significantly. It is worth noting the Statistics and Registration Services Act only requires 'published information' to be taken into account.

### 3.3 Census geographies

Consultations relating to Census geographies have been conducted alongside those on the main statistical outputs. The consultations were run in relation to three developments. These are

- The need for an Upper-layer Super Output Area (USOA)
- The need for Workplace Zones
- Instances of OAs/SOAs that are unfit as a statistical geography

The feedback from users will be evaluated over the coming weeks, and results produced in early summer 2010.

## **2011 Census Outputs – Consultation, Evaluation and Statistical Disclosure Control for England and Wales**

### **3.4 Secondary population-based outputs**

The design of the 2011 census instrument enables outputs to be produced on a range of population bases in addition to the main population base. Following recommendations made by ONSCD outputs relating to the following population bases are being considered:

- Work-day population
- Out-of-term population
- Majority of time population
- Overseas born short-term (3-12 month) immigrant population

In order to capture the requirements of a diverse set of users the ONS will engage in a separate consultation on these four recommended population bases.

The consultation will ask users to define in broad terms:

- the type of information they would wish to see produced in relation to the four population bases
- the level of detail they would require
- the geographical level for such information
- the business case for such information

Since an evaluation of the returned census data will be required in order to determine the accuracy of these populations, and the exact level of detail that can be provided, users will be informed that the exact specification of the outputs relating to these bases will not be known until well after the census data has been returned. The consultation will also include a general statement that all outputs relating to these population bases would be produced as experimental statistics, and that only the main residency population based outputs would carry the National Statistic stamp of approval.