



## 2011 CENSUS OUTPUTS: AIMS AND PRINCIPLES

### 1. Introduction

This paper sets out key high level aims and principles for 2011 Census outputs. The aims are aspirations and it may not be possible for all to be delivered, but we would like to start discussions early so that we can understand user priorities for the different aspects to inform our planning process.

The aims are grouped into those that relate to data access, statistical disclosure control, geography and quality.

**Advisory groups are asked to give their views on the proposed strategy for 2011 Census outputs.**

### 2. U.K. Consistency

The Registrars General of England and Wales, Scotland and Northern Ireland have, subject to the need for approval where appropriate by the relevant legislature, agreed to aim for harmonisation on a number of key aspects of the 2011 Censuses. These include:

- agreeing common methodologies on disclosure control and estimation methodology in advance of the Censuses;
- consistent, coherent and accessible statistics for the U.K and for each component country, with a joint database (and/or a common data schema) seen as a desirable way of facilitating that outcome; and
- a common approach to output specifications, data quality, data formats and timing of releases.

The full agreement can be found at

[http://www.statistics.gov.uk/about/census/census2011/2011census\\_prog.asp](http://www.statistics.gov.uk/about/census/census2011/2011census_prog.asp)

## 3 DATA ACCESS

### 3.1 Aim: Outputs free at point of delivery

A key aim underpinning user access to 2011 Census outputs is that all standard output will be free to users at the point of delivery.

Under the 1920 Census Act, funding to produce the main census outputs that are laid before parliament is included in core census funding, but the cost of additional statistics that are not laid before Parliament have to be recovered from customers. For 2001 Census outputs the Census Access project enabled costs for standard output not laid before Parliament to be recovered upfront. The project successfully bid for funds (£2.1 million) from the Invest to Save budget (ISB) which met 75% of projected costs. The remaining 25% (0.7 million) was contributed by a consortium of stakeholders – Department of Health, Local Government Association, Economic and Social Research Council, Office of the Deputy Prime Minister (ODPM). This arrangement meant all users had free access to all standard 2001 Census output at the point of delivery.

Further work is required to determine which outputs will be laid before Parliament, and to develop the funding model for standard outputs from the 2011 Census that are not laid before Parliament.

### **3.2 Aim: Dissemination methods and media to keep up to date with technological innovation**

An overarching aim for the dissemination of 2011 Census outputs is that methods and media keep up to date with technological innovation. However a balance will need to be struck between taking account of any future technological innovations and developing the necessary dissemination tools and systems in sufficient time. Research will also be needed to establish the extent to which Census requirements may be met by corporate dissemination solutions and tools.

The current assumption is that standard pre-planned outputs will be disseminated via a number of media - paper reports, on-line, and writable media (CD, DVD) but with dissemination likely to be web dominated and paper reports kept to the minimum to meet the legal requirements for reports laid before parliament under the 1920 Census Act (4.1). The user requirement for outputs on writable media will need to be established.

### **3.3 Aim: Flexible table generation on-line:**

In addition to pre-planned outputs, the aim is to provide the facility for flexible table generation on-line. This would be at two levels;

User-defined extractions from, and tallying of, standard tables whereby users can specify sub-groups and/or geographical areas from within published tables to construct tables of interest. (This facility is already available for local 2001 Census statistical output currently disseminated on the NeSS website);

The facility for user defined tables utilizing standard variables but used in a bespoke fashion. Such tables could be generated via hyper-cubes or from protected micro-data. Software that enables this functionality has already been developed and is available as part of the "Space Time Research" package of tools, which includes the Super-cross tabulation package used for 2001 Census outputs. The Australian Bureau of Statistics is currently testing this with the aim of utilizing it in their 2006 Census outputs to increase data utility. Statistical Disclosure Control software would also need to be developed to support this facility to apply adequate disclosure protection to the user generated tables, for example population thresholds and average and minimum cell sizes in tables. (The Australian Bureau of Statistics is currently developing disclosure control methodology to support this facility).

The facility of flexible table generation online has the potential to reduce the user requirement for standard output and commissioned output from the 2011 Census. Approximately 80% of commissioned output from the 2001 Census involved standard variables used in a bespoke fashion.

Clearly, this aim is subject to the development of sufficiently robust statistical disclosure control methodology and security systems.

### **3.4 Aim: On-line facility to produce graphs and maps of standard and user-defined output.**

The facility for on-line mapping and graphing of 2011 Census output disseminated on the NeSS website is already available. There are also interactive maps and charts on the NS on-line web pages. If the facility for on-line user defined output is developed there is the potential for joining up all this functionality to provide on-line charting and mapping facilities for both standard and user defined 2011 Census output.

### **3.5 Aim: Comparisons between 2001 and 2011 Census outputs**

Our aim is to produce comparisons between 2001 and 2011 Census outputs following research into reliability of measures of change taking into account issues such as changes to:

- definitions or questions;
- the geography for Census outputs; and

- statistical disclosure control methodology

Areas of poor coverage in either or both Censuses may mean that comparisons over time at OA level would not be robust for some areas.

Further research will be needed to determine what time series are meaningful and at what geographical level and additionally whether it will be possible to enable on-line user defined comparisons between 2001 and 2011 Census outputs.

### **3.6 Aim: To maximise data utility**

Different levels of access for different Census outputs are being considered in order to maximise data utility. Census tables for small populations, for example some ethnic groups particularly at low geographical levels, would be very sparse. Statistical disclosure control can compromise data utility because of the need to suppress detailed breakdown. The aim is to make all tabular output “fit for purpose” within disclosure control constraints. Data likely to be compromised could be subject to lower levels of statistical disclosure control and made available to users only under special licence or accessed in data labs with outputs checked prior to removal. An example would be the Origin/Destination tables which are very sparse. It may be appropriate to make these matrices available publicly at Super Output Area level, and available under licence at Output Area level, and allow users interested in specific journeys access to detailed micro-data in a safe setting.

Clear and equitable criteria for access would govern this arrangement.

### **3.7 Aim: Output prospectus and timetable to be published pre-release**

A pre-release outputs prospectus and timetable for statistical outputs and metadata will be published. Every endeavour will be made to keep to the published timetable but quality will not be compromised. Any unforeseen problems likely to result in an inability to meet the published timetable will be explained and communicated to users as early as possible, with revised publication dates.

### **3.8 Aim: Concurrent first release of statistics across the U.K.**

This aim is in accordance with the Registrar Generals’ agreement

### **3.9 Aim: Statistics will be released concurrently for all areas in England and Wales**

Publication of statistics will be staged, but at each stage statistics at a given level of geography will be released concurrently for all areas

### **3.10 Aim: Concurrent publication of appropriate metadata with associated statistical outputs**

Not all quality measures will be available at the time of statistical release, however the intention is to publish quality measures such as response rates and imputation rates concurrently with associated statistical outputs. Other more specialised evaluation will be published later in accordance with the published timetable.

### **3.11 Aim: Joined up and comparable U.K. outputs**

There will inevitably be some differences in questions across the U.K. countries to reflect local data needs and the decisions of devolved administrations. However, where differences do occur the aim will be to map results to a common framework to enable U.K. comparability at some level.

### **3.12 Aim: Joining Census outputs with other statistical outputs**

A high level aim is to join 2011 Census outputs with other statistical outputs. However this could range from simply a link from census tabular output to other data on same topic through to the production of topic specific reports that draw on Census and other data. One example would be building on the 2001 'Focus On' Reports.

Further discussion is required to determine which sources should contribute to topic output in the future and the extent to which Census outputs should be joined with other statistical outputs. If ONS were to put significant resources into this it is likely that topic experts would need to spend less time on other activities and hence other ONS outputs in the two year period around the Census may not be produced.

## **4. STATISTICAL DISCLOSURE CONTROL**

### **4.1 Aim: Common U.K. SDC methodology for Census 2011 outputs that minimises disclosure risk whilst maximising data utility**

The Registrar Generals' statement of agreement for 2011 U.K. Census includes aiming for a common Statistical Disclosure Control (SDC) methodology.

2001 Census outputs were subject to differing SDC methodologies across the U.K. which led to much discontentment amongst users and impeded U.K. compatibility.

The goal for 2011 Census is for a common SDC methodology to be adopted across the U.K. which protects against the risk of disclosure whilst maximising data utility (taking into account user output requirements) and which complies fully to the commitment to confidentiality on the Census form and to the legislatures.

In 2001 there were different disclosure control methodologies across the UK because of different views about the level of acceptable risk and the perception of what constitutes disclosure. High level discussions across the U.K. offices on these SDC policy issues are planned with the intention of coming to an agreed view on these risks and perception issues as the base for all further methodological research for 2011 Census outputs.

Further research will cover SDC methodological issues for all Census outputs including pre-planned tables, flexible outputs, commissioned output and micro-data and for different levels of access (public, licensed and safe setting)

Common SDC methodologies will be fully evaluated in terms of a utility/risk continuum and in the context of the agreed U.K. SDC policy position,

It is intended that this work will be completed before consultation on detailed output requirements commence.

## **5. GEOGRAPHY**

### **5.1 Aim: An effective and flexible approach to output geographies**

Census Division has been working with Neighbourhood Statistics and ONS Geography to establish the high level principles within which the 2011 Census output geography will be developed.

Discussions have identified three key options

1. Define new geographies for 2011 Census outputs
2. Keep 2001 geographies largely unchanged (OAs, SOAs)
3. Keep 2001 geographies at some level (e.g. SOAs) and redraw OAs within this constraint

The current ONS recommendation is to maintain existing geographies for 2011 Census outputs. Some modification of the current OA geography will be required to take account of any significant population changes that will have occurred since the last census

A separate paper on output geography will be provided to Census Advisory Groups in November 2005.

## **6. ADJUSTMENT AND QUALITY ASSURANCE**

### **6.1 Aim: Fully adjusted database**

The aim is for a database fully adjusted for missing responses to specific questions, (with the exception of any voluntary questions) and adjusted for estimated under and over coverage. It is also planned to undertake 100% coding of questions. However, a cost/benefit analysis and confirmation of funding will be required before a final decision is reached.

### **6.2 Principle: No Revisions**

Census output tables contain some 2 billion individual counts and take 2 years to produce. It would not be possible to update all of these counts therefore there will be a principle of no revisions to the census outputs. (Revisions refer to changes to headline census counts as a result of coverage error). This principle is underpinned by a commitment to put even greater effort into the quality assurance of results (see 6.4). Any necessary subsequent revisions will be made to mid-year population estimates if required.

The current plan is to capture and process data more quickly and complete coverage assessment processes faster than was the case in 2001 to allow more time for follow-up and quality assurance prior to publication. Timing of initial output release is therefore likely to be similar to that of 2001 Census.

### **6.3 Aim: Corrections policy in place in advance**

Corrections refer to changes to specific outputs as a result of coding or tabulation errors. Mechanisms will be put in place for users to report errors and for corrections to be prioritised. Users will be consulted on the most effective means of communicating corrections to the user population. Version control will be utilized to provide an audit trail and to enable users to access previous versions.

### **6.4 Aim: Joined-up and co-ordinated quality assurance policy to involve internal and external experts and bodies**

#### Q.A. of Census based population estimates

Before the 2011 Census, ONS will liaise with Local Authorities to quality assure address lists for their areas, utilising alternative L.A. data sources. This will help inform the census field operation and will also provide additional information for internal ONS experts who will be responsible for quality assuring the estimates.

In any areas where ONS has particular concerns, we may discuss the estimates with the relevant LA or other government departments during the QA process, subject to strict confidentiality protocols and agreements.

#### Quality Assurance of Tables

External bodies quality assured 2001 Census tables in an ad hoc manner with different users using different processes and identifying different errors after publication. If the facility for flexible tabular outputs from the 2011 Census is available there is the potential for a reduction in the volume of standard tabular output. This would make the internal Q.A. process more manageable. In addition to internal Q.A. of the tables produced, it is intended to plan early access to tables for selected users for quality assurance purposes.

## **7. Summary of census 2011 output aims and principles**

- Outputs free at point of delivery
- Dissemination methods and media to keep up to date with technological innovation
- On-line flexible table generation
- Improved data utility of some data outputs via differential access levels
- On-line mapping and graphing of standard and user-defined output
- 2001/2011 comparisons
- Pre-release output prospectus and timetable that is adhered to
- Concurrent first release of statistics across the U.K.
- Statistics will be released concurrently for all areas in England and Wales
- Appropriate associated metadata published concurrently with statistical output
- Joined-up and comparable U.K. outputs
- Joining Census output with other statistical output
- Common U.K. SDC methodology for Census 2011 outputs that minimises disclosure risk whilst maximising data utility
- An effective and flexible approach to output geographies
- Fully adjusted database
- No Revisions
- Corrections policy in place in advance
- Joined-up and co-ordinated quality assurance policy to involve internal and external experts and bodies.

**Angele Storey**  
**2011 Census Outputs**  
**October 2005**