

# POPULATION AND MIGRATION STATISTICS COMMITTEE (SCOTLAND)



## ONS's Census 2011 consultation/plans

### (a) 2011 CENSUS ADVISORY GROUPS: UPDATE ON TOPICS CONSULTATION

#### Background

- 1 As part of its research programme for the 2011 Census in England and Wales, ONS launched a programme of user consultation, focused on identifying those topics to be considered for inclusion in the next Census. An ONS Consultation Paper, '*The 2011 Census: initial view on content for England and Wales*', was published on 16 May, and invited responses by 5 August 2005. The programme was launched with a Press Release and received good press coverage.
- 2 To supplement the consultation document, six roadshows were held during June and July in Harrogate, Crewe, London, Northampton, Newcastle and Cardiff with over 300 people attending. Over 1,400 responses, from over 450 individuals/organisations have been received.

#### Action

- 3 **Advisory Group members are asked to note some of the early messages from consultation.**

#### Initial summary of response

- 4 An initial scan of the responses shows:
  - that responses have been received on all topics covered in the consultation;
  - the strength of justification varies a lot and some may need follow-up to fully understand the response;
  - there have been good quality responses on all the issues; and
  - there is continuing interest in all the topics previously covered in the 2001 Census.

Basic demographics are required, for example for service provision (including resource allocation), Migration, Ageing population and Housing demand.

Academic qualifications – although poorly answered in 2001, are still the best source of no/low qualification at small area and to identify areas of deprivation  
Disability and illness; carers information is used for example, to support service provision, and monitoring equality.

Ethnicity is used for cross-classifying with other variables.

#### Requests for new information

- 5 There have been requests for a number of new topics to be included in the Census:

- **Second addresses** – to support coverage; for service provision by, for example, Local Authorities.
- **Visitors** - to support coverage; and may indicate those with no usual address.
- **Sexual orientation** – for service provision and for monitoring equality.
- **Income** – Government departments want data on income in order to target areas of deprivation and accordingly are interested in the lower levels of income distribution. Businesses want data on income to target business services and are interested in full income distribution.
- **Languages spoken** – Language used in home for the purpose of provision of services; Language proficiency, for example, for resource allocation; British Sign Language.
- **Identity categories** – Welsh, Cornish, Kashmiri, Gypsy, Irish Traveller, Arab, Sikh. Request for Welsh/Welsh speaking throughout the UK.

### Next Steps

- 6 The process for assessing responses is for ONS to:
  - summarise requirements for each topic/question;
  - assess against the criteria set out in the consultation document;
  - review summaries with ONS topic experts;
  - hold internal UK workshop on findings and draft response;
- 7 This process will take until spring 2006 when a summary of the responses and the ONS view of the likely content of the 2011 Census will be published. There will also be engagement with census user groups and government departments.

### Questionnaire constraints

- 8 There are, however, always demands for more questions than can be accommodated on a self-completion questionnaire. There are a number of constraints such as the burden on the public and cost, and a number of difficult trade-offs will have to be made. It is likely that the questionnaire design will be broadly similar to that for 2001 with a page of household questions and three pages per person. Demand for new information may mean, however, that some questions from 2001 will be lost.
- 9 Question development and testing has already started, and ONS is aiming to produce a draft Census questionnaire for the 2007 Test by June 2006.



## **(b) 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**





## **(c) 2011 Census: Small Area Outputs Geography Policy**

### **1. Summary**

This paper sets out ONS' proposed small area geography policy for 2011 Census outputs. A principle of stability and strong comparability with 2001 Census Output Areas (OAs) and Neighbourhood Statistics Super Output Areas (SOAs) is proposed.

Further research and analysis needs have been identified to establish how OAs and SOAs should be maintained, and what the scale of maintenance required is likely to be. It is also recognised that future policies in statistical disclosure control and potential time series outputs for the next Census are closely linked with any small area geography policy.

**Advisory groups are asked to give their views on the proposed policy at the meeting and/or in writing by the end of the year.**

### **2. Background**

Output Areas (OAs) for Census 2001 were designed and released in 2003.

OAs were designed to :

- have very similar population and household numbers (tightly grouped around a population mean of approx 300 and a household mean of 125);
- have a high degree of homogeneity in terms of housing tenure and type.

In this respect OAs provided a real 'statistical geography' - an improvement on existing geographies which have broad population ranges and are subject to regular re-organisation.

Several constraints were imposed on this geography. OAs were nested within the wards and parishes that existed at the time. They were also built from the 'soft' geography of whole postcodes. The imposition of these constraints on the OA geography resulted in boundaries that often appear strange on a map as they do not follow a 'hard' physical geography (such as streets, rivers, railways, hills, land parcels etc).

The Neighbourhood Statistics (NeSS) geography policy built upon the idea of OAs and introduced the concept of a stable building block geographical hierarchy, to aid data comparison over time and between areas. In other words, to move away from wards and other local or functional geographies that are designed and maintained for non statistical purposes.

The Super Output Areas (SOA) hierarchy has been designed to provide this stable building block geography. It was built directly from the OA geography. The Lower and Middle Layers were both released in 2004. Again both layers were designed for statistical purposes along similar lines to OAs.

There are critics of OAs, and as a result anything that's built from them. However, the idea of a stable building block hierarchy, potentially allowing a wealth of cross analysis and time series analysis of datasets, is broadly welcomed both by data suppliers and by users.

When designed, OAs were not viewed as the future building block for a stable geo-statistical hierarchy. Indeed, the main emphasis at the time was that the accurate geo-referencing of data would allow flexibility for future outputs. In other words, that alternative output geographies could subsequently be produced, and that past data could be re-released against these newer geographies. Such a concept required further work to be carried out to devise methodologies and policies to handle overlapping geographies, disclosure, and recasting and these have gone forward in the NeSS programme. Recently, however, the emphasis has been as much on outputs built from stable building block geographies (the OAs and SOAs) and it is this that underpins current thinking as outlined below.

### **3. Advantages and disadvantages of stability**

ONS proposes that we should keep stability with 2001 OAs and SOAs.

The **advantages** of keeping a stable hierarchy beyond 2011 are as follows:

- The OA / SOA hierarchy performs well on statistical measures such as having compact household and population distributions, and having greater external heterogeneity and internal homogeneity than wards;
- Whilst there will be some degradation in terms of these statistical measures over time, countrywide this is not expected to be marked;
- A critical driver for SOAs was that they should be stable over time. This will facilitate both time series analysis and also greater cross analysis potential as SOAs become embedded and more data is released for them;
- We will have a stable output geography between Censuses – allowing comparison (within some constraints – see below);
- A stable output geography between Censuses will allow comparisons for the quality assurance of Census results, which has not been possible before;
- Public statements have been made that SOAs are intended to be stable;
- There is no guarantee of success if we attempted to create a further geography;
- There is a broad aim to reduce the number of geographies used, not to increase them.

Some **potential disadvantages** have been identified. They are:

- The possibility that we could design something better, either now or in the next few years with subsequent technical, data and policy developments;
- Some user demands for OAs to be redrawn to better reflect local concepts and definitions of neighbourhoods;
- Uncertainty over future disclosure control policies for small area Census outputs;
- OAs will not relate well to many wards by 2011. There is still a large local demand for “definitive” ward level data to be reported for political and policy setting reasons. However, virtually all wards change at some point, and the precision associated with the counts is probably misleading. No harm seems to have been done in Scotland who realised this decades ago;
- There are likely to be a number of methodological differences between the 2001 and 2011 Censuses. These may be most apparent where estimates are made in areas suffering from lower response rates. Direct comparison between 2001 and 2011 at the OA level may therefore throw up spurious results. However, a consistent geography removes one area of uncertainty with regard to the effects of those differences.

#### 4. Proposed policy

The following principles lie behind the policy being proposed:

- 2011 Census and NeSS geography policies must be aligned;
- We need a geography which is good for statistics and policy development;
- SOAs are the fundamental geography for NeSS and will be a good small area geography for time series analyses between censuses. As a result, the stability of the current OA / SOA hierarchy carries a very high weight;
- The potential for flexible output geographies and the release of past and future data against them still exists. To be realised it requires that unit records be accurately geo-referenced and that enabling disclosure control policies and methods be established.

With these principles as our basis, ONS proposes the following policy for small area geographical outputs.

1. The OA / SOA hierarchy will continue in its present form for the foreseeable future, and certainly beyond the next Census.
2. “Definitive” 2011 Census outputs will be released for this hierarchy (OA – SOA – Local Authority).
3. If disclosure thresholds were raised, OAs will simply be amalgamated where necessary for reporting purposes. If disclosure thresholds are reduced, OAs will remain unchanged.
4. Maintenance of the hierarchy will be required to respond to changes in the real world. Each layer of the hierarchy will have fixed minimum population and household thresholds to facilitate data release. They will also have guideline maximum thresholds to maintain a relatively tight distribution across the country. It is to be expected that as time passes some areas will fall below or above these thresholds. ONS will develop and evaluate approaches to maintenance.
5. Maintenance may also be required to fix existing problems. Any structural changes to fix problems will be limited to less than 5% of OA boundaries. Strict criteria will be established for assessing local requests for changes to be made to the hierarchy. Sufficient time will be given before the next census for requests and assessments to be made. If criteria are met, the task of redrawing will be owned by ONS.
6. Maintenance and reworking of the hierarchy will be carried out post census, well ahead of the release of any small area data.
7. Work will be undertaken to assess the potential for better aligning OA boundaries to features on the ground and to OS MasterMap. A range of different approaches will be considered. This work will have as a prime concern that there should be no loss of IPR to Ordnance Survey and no detrimental impact on our ability to freely disseminate digital boundaries for OAs and SOAs.

One consequence of this policy is that it may not be possible to produce “definitive” ward level data for the next Census. At the very least, however, ward outputs will still be produced by best fitting data produced for OAs to wards.

A further consequence is that methodological differences between censuses may be apparent by comparing data at the small area level. It is argued that this should not be a reason for throwing away the geography, but rather these potential differences should be clearly highlighted upfront.

**Nick Stripe**  
**Office for National Statistics (Geography Branch)**  
**2 November 2005**

### Testing the policy

The proposed policy outlined above has been tested internally within ONS, and externally with our Geography Advisory Group. This is a small group of leading academic, central government, local government, and commercial representatives with a professional interest in geo-statistical issues.

Feedback received has been broadly very supportive. Areas for further research and analysis have been suggested, whilst the point has been made that a lot can change in five years.

A small selection of comments to reflect those received are given below:

*"It is better to have stability in units rather than trying to produce a 'better' set of OAs, as it will never be possible to meet all the criteria."*

*"This position is generally sound and it is helpful to put out some initial statements for public information and comment."*

*"You will remain under heavy pressure to provide ward statistics. ONS needs to consider how it can better manage expectations in its markets. How can it educate users on the subtleties of apparent precision versus statistical reliability issues inherent in any survey work."*

*"There is a case for maintenance of SOAs wherever possible, but I'm less convinced that we need to stay with 2001 OAs."*

*"I strongly agree with the suggestion of sticking with OAs as currently defined for the 2011 census output. I think the benefits for research and policy making will be huge (of the order of many tens of millions of pounds worth of added value to the census if this choice were costed)."*

*"We are unable to judge yet whether SOAs meet our requirements, but the impression is so much better than previously available ward data."*

*"Are ONS weighing up the costs and benefits of disclosure? Not just in a purely statistical environment but as a cost benefit to the nation? Too tight statistical disclosure control undermines the purpose of this data, to help develop policy to help people."*

*"Coordinate referencing of a database allows geography for time series to be kept up to date almost indefinitely. The research effort should concentrate on determining area levels and statistical content for reliable time series, and on addressing the disclosure control issues of changing geographies."*

*"The balance struck between the extremes of absolutely no change and wholesale revision was about right."*

*"Concerns about disclosure by differencing have led to a lot of the current approach. Do we and will we still have the same views about this?"*

*"You need very low level geo-referenced data for flexible geographical outputs and the supply of such data is not yet good."*

*"The arguments for stability and continuity provided by using the existing OAs etc far outweigh any against."*

*"All of this is sensible and supported by us at Marks & Spencer – modifying bricks is a real pain."*

*"There will be a some local disapproval if OAs are maintained."*