

THE CONDUCT OF THE 2011 CENSUSES IN THE UK STATEMENT OF AGREEMENT OF THE NATIONAL STATISTICIAN AND THE REGISTRARS GENERAL FOR SCOTLAND AND NORTHERN IRELAND

The passages in italics are a commentary on progress at September 2011.

Introduction

1. This is a statement of agreement between the National Statistician and the Registrars General for Scotland and Northern Ireland about the conduct of the 2011 Censuses which it is the intention to conduct simultaneously throughout the UK in 2011. *All three censuses were conducted on 27 March 2011 and all three UK Census offices continue to work together on Census outputs.*
2. A Census is taken by the Statistics Board (UK Statistics Authority) and the Registrar General for Scotland under the Census Act 1920 and by the Registrar General for Northern Ireland under the Census Act (Northern Ireland) 1969. The necessary subordinate legislation, relating to the specific arrangements for the Census in each country, requires the approval of the UK Parliament in Westminster for England and Wales the Scottish Parliament and the Northern Ireland Assembly respectively. Welsh Ministers are consulted on the Census for England and Wales, and are responsible for making Census Regulations for Wales. *The Census Orders and Census Regulations for the respective Censuses were all put in place in 2010.*

Principles

3. Each country will be autonomous, with the final decision lying with the relevant Registrar General, Ministers and legislature.
4. Subject to that autonomy, the National Statistician and Registrars General agree that the three Census Offices (the Office for National Statistics (ONS), General Register Office for Scotland (GROS) and the Northern Ireland Statistics and Research Agency (NISRA)) will aim to work in unison to ensure that the 2011 Censuses are a success in providing high quality population and housing statistics, meeting the needs of data users and reflecting UN-ECE and Eurostat requirements.
5. In particular, the National Statistician and the Registrars General have agreed that the three Census Offices and the Welsh Assembly Government will work together and will reach mutual agreement wherever possible on the points in this Statement of Agreement, which will facilitate harmonisation where that is in the interest of Census users. Other points may be added over time where that would facilitate harmonisation.

Harmonised Aspects

6. The National Statistician and Registrars General have, subject to the need for approval where appropriate by the relevant legislatures, agreed the following aspects of the Census where they will aim to achieve harmonisation:-

- **Date.** The three Censuses should be carried out on the same date in 2011. *The Census took place on 27 March 2011.*
- **Population Base.** A common population base should be agreed. *A common population base (usual residents plus visitors) has been agreed.*
- **Topics and Questions.** Common questions should be agreed wherever possible, diverging only in response to clear user needs, with such divergences minimised and the scope for cross-comparison maximised. *Extensive consultation was carried out with users and other interest groups by each of the three Census Offices and a good degree of harmonisation was achieved, helped significantly by the decision to include a fourth page of questions per person. The questions were approved by the relevant legislatures. There were 65 questions asked in at least one of the census questionnaires. These included questions relevant to only one country, such as knowledge of Welsh or proficiency in Gaelic. There were 43 questions which were asked throughout the UK and in 33 of these the wording of the question and the response options were identical. Of the remaining ten where the question wording or response options differ, seven could be aggregated to provide consistent statistics across the UK.*
- **Census materials.** The design of the Census forms and other materials should be co-ordinated. *A common design was agreed for census questionnaires. Field support materials were developed as required for each country.*
- **Definitions.** Common definitions and classifications should be agreed and published. *Key definitions were published alongside the proposals to the UK legislatures on census topics in December 2008.*

- **Disclosure Control and Estimation Methodology.** Common methodologies for disclosure control and for estimation should be evaluated, tested and agreed in advance of the Census. *A disclosure control policy was agreed by the National Statistician and the Registrars General, which was received well by users. The policy statement forms an Annex to this agreement. Record swapping was agreed as the primary method of statistical disclosure control.*
- **Publicity.** A common approach to publicity, tailored to local needs, should be agreed. *Publicity for the Census Tests and Rehearsals was approached separately, but a coordinated strategy was developed for 2011.*
- **Procurement of Systems and Services.** The different scale of the Census Offices may dictate different procurement solutions but a common strategy should be adopted and a common approach generally taken where that maximises value for money, operational effectiveness and (especially) consistency of outputs. *A single contract, with core elements and variations for each Census Office, for the key operational services for England, Wales and Northern Ireland was awarded to Lockheed Martin UK in August 2008. A separate Scottish contract was awarded to CACI (UK) Ltd. in June 2008.*
- **Outputs.** The final product should be consistent, coherent and accessible statistics for the UK and for each component country, a joint database (and/or a common data schema) being a desirable way of facilitating that outcome, with a common approach taken to output specifications, quality, data format and timing of releases. *A joint statement on Output Aims has been agreed by the National Statistician and Registrars General and forms an Annex to this agreement. Each Census Office has published an output strategy. All countries are taking part in consultations on the detailed content of the census outputs, including the content of a set of UK outputs, which will be produced and published by ONS from a single unit record file to ensure harmonisation.*

Cost Sharing

7. The costs, both contractor and Census Office, of systems and services will normally be shared between the Census Offices on an actual cost basis, where appropriate, or apportioned according to the population totals (2001 Census). Differences in approach will be kept to a minimum.

Governance and dispute resolution

8. To promote UK harmonisation, every effort will be made to reach agreement through the cross-working mechanisms of the three Census Offices and any differences of opinion will be discussed and resolved at the UK Census Committee.

**Office for National Statistics
General Register Office for Scotland
Northern Ireland Statistics and Research Agency
February 2005, Revised September 2011**

Annex to the Registrars General's Agreement

2011 CENSUS OUTPUTS: AIMS

1 INTRODUCTION

The UK Registrars General have collectively committed to joint working on the 2011 Census with the aim of maximising harmonisation across the separate Census operations. The objectives of the Registrars General are set out in the Registrars General's Agreement. This document is an annex to that agreement and sets out key high level aims for ensuring, so far as practicable, that the outputs from the 2011 Census meet user needs.

2 UK 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:

- a common population base;
- common questions to be agreed wherever possible, diverging only in response to clear user needs, with such divergences minimised and the scope for cross-comparison maximised;
- common methodologies on disclosure control and estimation;
- consistent, coherent and accessible statistics for the UK 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

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.

Through the Census Access project (England and Wales), SCROL (Scotland) and NICA (Northern Ireland) the three Census Offices delivered all standard 2011 output free to users at the point of delivery. The funding mechanism for these projects differed across the Census Offices; for example the bulk of the England and Wales project was funded through the Invest to Save Budget (ISB) initiative with further funding from a consortium of stakeholders – Department of

Health, Local Government Association, Economic and Social Research Council, Office of the Deputy Prime Minister (ODPM).

Further work is required to determine which outputs will be laid before the Parliaments, 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 section 4.1 of the two Census Acts. 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. (Facilities of this type are already available for local 2001 Census statistical output currently disseminated on the NeSS/SCROL/NICA websites);
- The facility for user-defined tables utilising 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 (ABS) is currently testing this with the aim of utilising it in its 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. (The ABS 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 per cent 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/SCROL/NICA websites 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

The 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 breakdowns. 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 could be the Origin/Destination tables which are very sparse. It may be appropriate to make these matrices available publicly at Super Output Area (SOA)/Data Zone level, and available under licence at Output Area (OA) 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 UK

This aim is in accordance with the Registrar Generals' agreement.

3.9 Aim: Statistics will be released concurrently for all areas within England and Wales/Scotland/Northern Ireland

It is probably inevitable that publication of statistics will be staged, for example in 2001 there were separate releases for population estimates, Key Statistics, Census Area Statistics, and Standard Tables. As in 2001, even though there may be a staged approach to the release of different geographies (this will be the case for England and Wales) at each stage statistics 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 UK outputs

There will inevitably be some differences in questions across the UK 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 UK comparability at some level.

3.12 Aim: Joining-up Census outputs with other statistical outputs

A high level aim is to join-up 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 the Census Offices 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 outputs in the two year period around the Census may not be produced.

4 STATISTICAL DISCLOSURE CONTROL

4.1 Aim: Common UK statistical disclosure control methodology for Census 2011 outputs that minimises disclosure risk whilst maximising data utility

The Registrars General's statement of agreement for 2011 UK Census includes aiming for a common Statistical Disclosure Control (SDC) methodology.

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

The goal for 2011 Census is for a common SDC methodology to be adopted across the UK 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 UK 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 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

ONS Census Division has been working with Neighbourhood Statistics and ONS Geography to establish the high level principles within which the 2011 Census output geography for **England and Wales** will be developed. Three key options have been identified

- a) Define new geographies for 2011 Census outputs
- b) Keep 2001 geographies largely unchanged (OAs, SOAs)
- c) Keep 2001 geographies at some level (e.g. SOAs) and redraw OAs within this constraint

The current ONS recommendation for England and Wales 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 paper on output geography for England and Wales is available at http://www.statistics.gov.uk/census2001/cn_142.asp.)

In **Scotland** decisions have not yet been taken on high level geography principles, though it seems likely that they will be largely similar to those employed in recent Censuses.

In **Northern Ireland**, such decisions have yet to be taken and will need to be viewed in the context of the ongoing Review of Public Administration which is expected to result in the number of Local Government Districts being reduced from 26 to 7.

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 full (100 per cent) coding of questions.

However, a cost/benefit analysis and confirmation of funding will be required before a final decision is reached.

6.2 Aim: 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, so the aim is to make no revisions to the 2011 Census outputs. (*Revisions* refer to changes to headline census counts as a result of coverage error). This aim is underpinned by a commitment to put even greater effort into the quality assurance of results (see 6.4 below). Any necessary subsequent revisions will be made to mid-year population estimates if required.

While the timing of initial output release is likely to be similar to that of 2001 Census, 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.

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

Quality Assurance of Census-based population estimates

The three Census Offices across the UK recognise the importance and potential benefits of a comprehensive and coordinated programme of quality assurance prior to the release of the 2011 results and are committed to identifying all available potential sources of useful information and relevant internal and external experts/bodies to assist with the work. For example, before the 2011 Census, ONS will liaise with Local Authorities to quality assure address lists for their areas, utilising alternative LA 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. GROS also intend to work with local authorities on quality assuring address lists, while NISRA will do likewise with Ordnance Survey.

In any areas where a Census Office 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 QA process more manageable. In addition to internal QA 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

- 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 UK
- Any given set of statistics will be released concurrently for all areas within a country of the UK
- Appropriate associated metadata published concurrently with statistical output
- Joined-up and comparable UK outputs
- Joining-up Census output with other statistical output
- Common UK 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.

ONS/GROS/NISRA

September 2011

Annex to the Registrars General's Agreement

UK Statistical Disclosure Control Policy for 2011 Census Output Background

The Registrars General of Scotland, England and Wales and Northern Ireland have agreed to aim for a common Statistical Disclosure Control (SDC) methodology for 2011 Census outputs. This will help achieve the aim of harmonising the three Censuses where that is in the interest of Census users.

Adoption of a common SDC methodology across the UK will only be possible if there is an agreed SDC policy position across the Census Offices, i.e. agreement about what constitutes a disclosive risk in a census context and tolerable risk thresholds. This statement sets out the SDC policy position that has been agreed by the Registrars General.

UK SDC policy position

The UK 2011 Census SDC policy position is based on the principle of protecting confidentiality set out in the National Statistics Code of Practice which includes the guarantee that 'no statistics will be produced that are likely to identify an individual unless specifically agreed with them'.

Because the key strength of the Census is its complete coverage, and its ability to generate statistics about very small groups of people (as is necessary to ensure that Government and other policies take account of the needs of small groups and communities), it is impracticable entirely to remove the risk of disclosure, without harming the utility of the data. With that in mind, the Registrars General have concluded that the above Code of Practice statement can be met in relation to Census outputs if no statistics are produced that allow the identification of an individual (or information about an individual) with a high degree of confidence. The Registrars General consider that, as long as there has been systematic perturbation of the data, the guarantee in the Code of Practice would be met.

It is considered that "attribute disclosure" (i.e. learning something about an individual) as opposed to "identification" of an individual is the key disclosure risk, because identification reveals no new information to the user. 'Attribute disclosure', however, involves a user discovering something new from the census data that was not previously known to them.

In a Census context, where thousands of tables are generated from one database, the risk of attribute disclosure occurring can be addressed by introducing uncertainty about the true value of small cells.

In order to meet the agreed interpretation of the Code of Practice, it has thus been agreed that small counts (0s, 1s, and 2s) could be included in publicly disseminated Census tables provided that

- a) uncertainty as to whether the small cell is a true value has been systematically created; and
- b) creating that uncertainty does not significantly damage the data.

The exact threshold of uncertainty required has not been decided. The Registrars General will make this judgement at a later stage in the context of results from methodological research into the balance of protection afforded, and damage caused, by various SDC methods.

Different levels of disclosure control are applied to Census outputs according to the mode of access. In general the aim will be to make as much as possible of the Census tabular output publicly accessible. However, if tabular outputs are likely to be seriously compromised by SDC (for example Origin/Destination tables at low geographical levels) then these could be released under other access arrangements (licence or safe setting) where restrictions on access to the data allow less stringent levels of SDC to apply, in order to protect the utility of the data.

As a result of the Chancellor's decision to legislate for ONS independence, the current NS Code of Practice: Protocol on Data Access and Confidentiality will be replaced. But the obligation to preserve the confidentiality of census outputs is likely to be heavily informed by the current Code of Practice.

Implications of the proposed SDC policy position for SDC methodology

The decision to allow small cells in publicly disseminated tables means that no methods of SDC (pre-tabular, post-tabular or combinations of the two) have been ruled out and all methods will be evaluated. The Registrars General have expressed a preference for pre-tabular methods, provided there is not undue damage to the data.

To ensure that the public, and expert audiences, are confident that confidentiality will be preserved by the measures taken to avoid disclosure, clear explanations would be given on the protection afforded by the SDC strategy, and other steps which protect confidentiality, that had been applied.

The choice of SDC methodology for 2011 Census outputs will be based on evaluation of the risk and utility of the various possible methods. Methods will be recommended that afford an acceptable level of protection and preserve the highest level of utility of outputs. Consistency and additivity across tabular output is a priority for users and these will be given a high priority in the assessment of the utility of SDC methods.

Next steps

The principles outlined in this statement provide a basis for consultation with users of Census data and for a two year period of methodological research. The research will assess pre and post tabular SDC methods in terms of the protection they afford together with their impact on the integrity of the data (a risk/utility framework). Because of the interdependence between disclosure control of (pre-defined) census tabular data and disclosure control for other types of census outputs (microdata samples and flexible user defined tabular outputs), SDC methods for all types of Census output will be assessed concurrently and a key consideration in evaluating SDC methods for tabular data will be the potential impact on these other types of census output.

Users will be updated and consulted during the research period. There will also be an independent review through the UK Census Design and Methodology Advisory Committee.

The Census White Paper (and parallel documents relating to Scotland and Northern Ireland) are timetabled to be published in October 2008 and will formalise the agreed policy position of the Registrars General by the inclusion of an SDC policy statement. A recommended SDC strategy for all types of 2011 Census outputs was finalised in Summer 2009. Further work on the details has continued since, and this has included public consultation.

Office for National Statistics
General Register Office for Scotland
Northern Ireland Statistics and Research Agency

September 2011