

CENSUS ADVISORY GROUP

AG (10) 23 Outputs - Consultation, Evaluation and Statistical Disclosure Control

2011 Census Outputs - Consultation, Evaluation and Statistical Disclosure Control for England and Wales – autumn 2010

- This paper has been produced to update Census Advisory Group members of the progress made in relation to the consultation on 2011 Census outputs over the last six months, and to inform members of the proposed plans for further work over the next period. The paper includes:
 - A high level summary of initial findings from the consultation;
 - The methodology used to evaluate users' requests;
 - Some key decisions on factors affecting the output design including progress made in relation to statistical disclosure control and its likely impact upon the design of the outputs;
 - Plans for the next round of consultations (Annex A).

Action requested of Advisory Groups:

Advisory group members are asked to note the progress made over the last six months and are invited to comment on the future programme of work.

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**

2011 Census output consultation

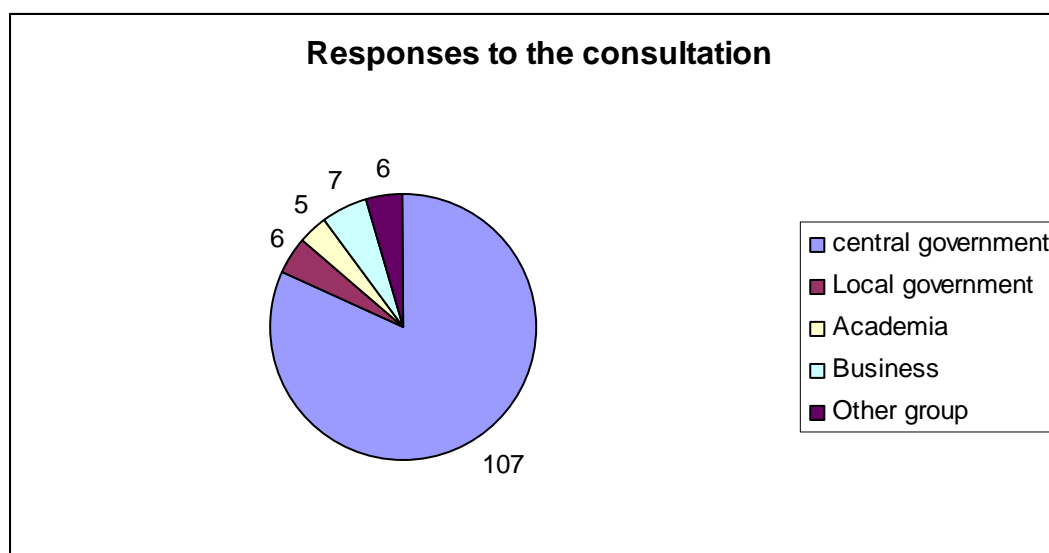
1. Introduction

Advisory group members were previously informed of the 2011 Census consultation plans and timetable. The project requests advisory members to note that although many of the planned activities have been completed to time, some slippage has occurred. As a result, the second round of consultation which was due to take place in the autumn will now take place early next year (2011). A revised timetable is included in Annex A. Advisory group members are encouraged to respond on any aspect of the paper making recommendations or suggestions for further development.

2. Consultation evaluation

Between December 2009 and April 2010 users were asked to comment on and make amendments to a 'baseline' set of specifications and encouraged to add any additional requirements they wished including new tables, geographies, derived variables and classifications. The exercise was publicised widely among all census users, and responses representing over one hundred and thirty different interested parties were received (Figure 1). Feedback from the consultation has been positive, with many users commenting on the usefulness and accuracy of the documentation, whilst taking advantage of the opportunity to comment on the specification and make additional requests. The majority of responses came from local government; 107 submissions represent the interests of local or regional government and PCTs.

Figure 1: Number of responses to the consultation

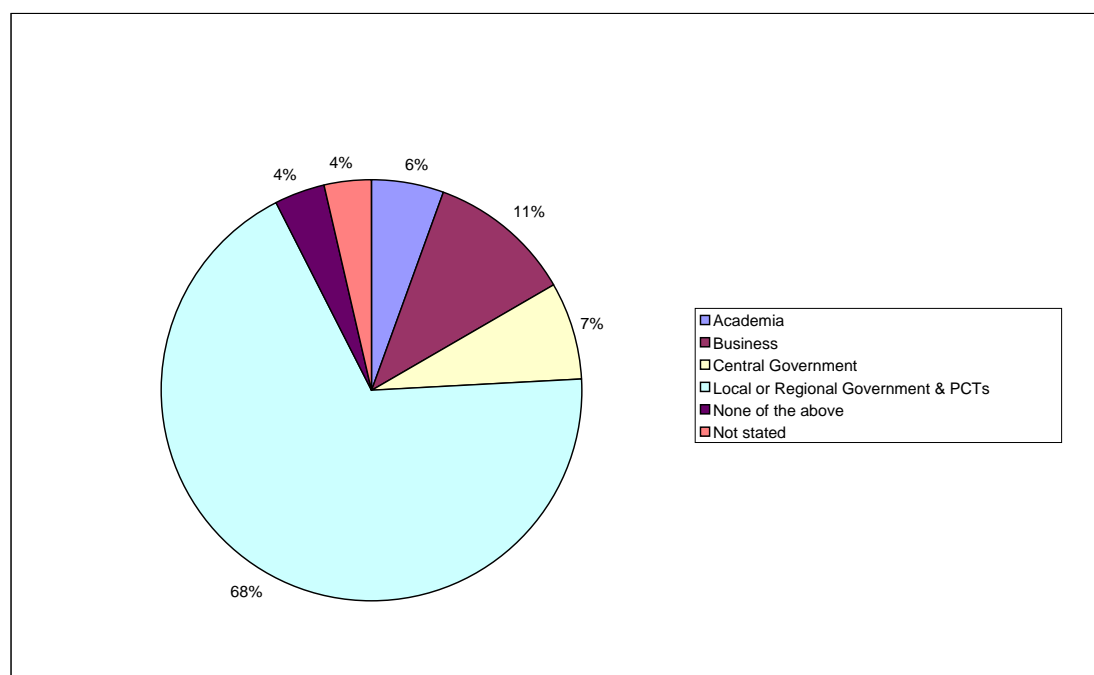


More surprising was the response from central government with six submissions, considerably less than anticipated. This was probably due to a combination of factors; however one key reason may be that many central government departments with an interest in the census have been directly engaged in census planning over a long period of time and this may have led to the reduced level of engagement. ONS has, as a result, contacted several departments directly asking them to submit a late specification of requirements or respond via the next round of consultation early in the new-year. ONS will also be making direct representation to all analytical Heads of profession (HOP'S) in the next round of the consultation process, with the aim of promoting the final round of consultation and thereby ensuring central government requirements are delivered in the final specification.

2.1 Feedback on existing tables

Approximately 50% of the consultation respondents completed the 'feedback on existing tables' worksheet. The greatest number of completed responses was received in the category of local or regional government and PCTs. In all 68% of responses to the feedback sheet were received from this sector (Figure 2).

Figure 2: Responses with completed "Feedback on Existing Tables" worksheet, reported by type of organisation



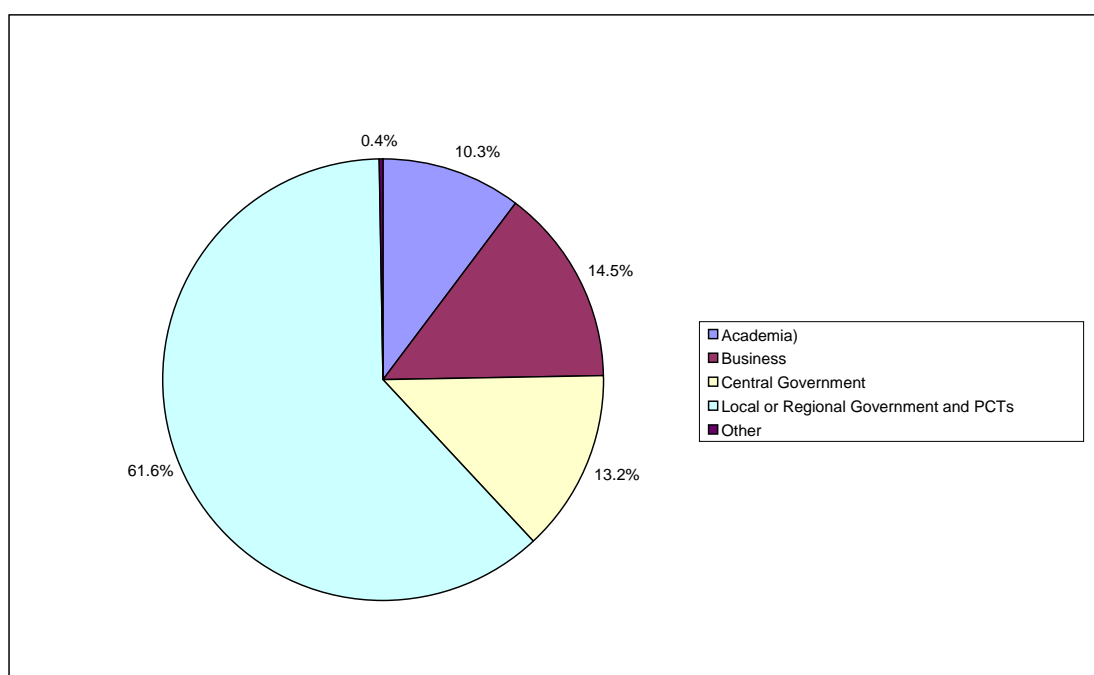
The majority of existing tables were in high demand. Approximately 340 of the 400 tables had at least 37% of users indicating that they would use the table if produced with 2011 data. Tables with a lower level of requirement were mostly those which are relevant only to specific user groups, such as Welsh language tables. The key finding arising from the feedback on existing tables is that there remains a substantial requirement for all 2001 tables¹.

¹ With the exception of those tables for which data are not being collected in 2011, such as lowest floor level.

2.2 New Tables

There were 242 total requests for new tables, made by 34 organisations. A significant number of these specify similar tables resulting in distinct requests for approximately 150 unique tables. Figure 3 illustrates the share of the total requests between different organisation types.

Figure 3: New tables requested reported by type of organisation



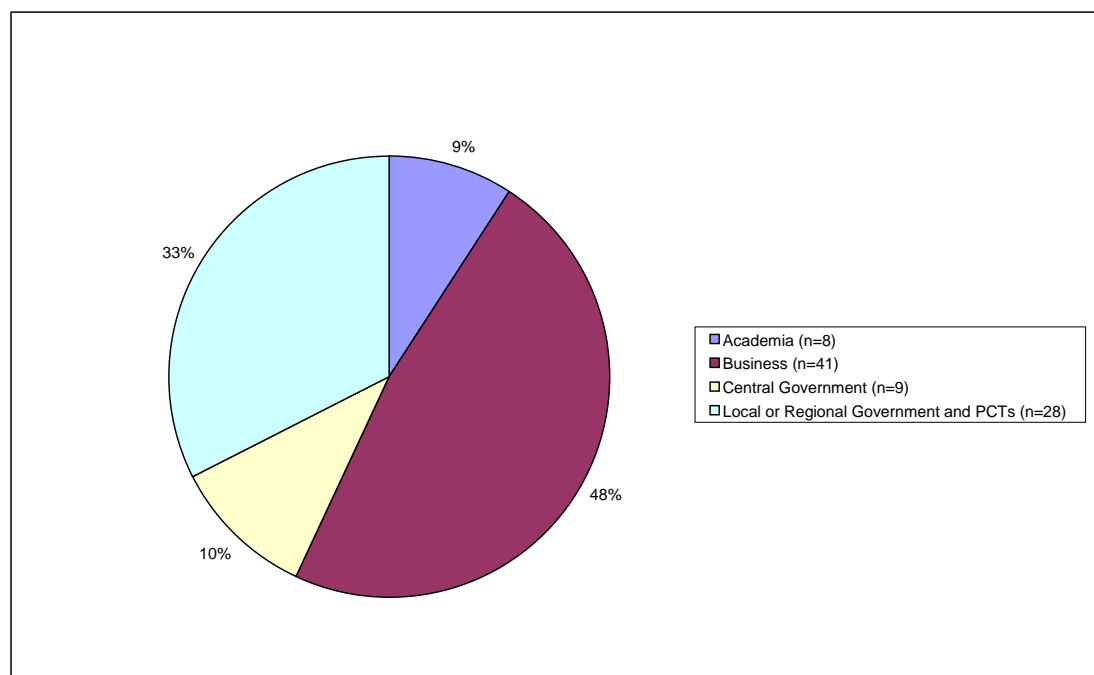
The most commonly requested new tables were univariate tables relating to new questions such as main language, citizenship, national identity and intention to stay. There was also significant requirement expressed for cross-tabulations involving main language and theme tables on the topics of equality and labour market. Again, it will be taken into consideration that many of the tables which received fewer requests did so because they relate only to the requirements of a specific set of users, which should not prejudice any decision to include or exclude such products.

2.3 New derived variables (DVs)

There were 86 requests for new derived variables; approximately 50 of these were unique requests (since a number of variables were requested more than once). Figure 4 illustrates the share of different organisation types in requests made.

The requested derived variables include method of return of census questionnaire (paper or internet), workforce information for those of working age and basic demographics of visitors.

Figure 4: New derived variables requested, reported by type of organisation



3 Consultation evaluation methodology

In order to make certain the responses to the consultation on outputs are evaluated in a fair and transparent manner a set of criteria have been developed by which the responses to the consultation have been assessed. Each criterion has its own individual weight based upon the level of importance assigned to it. The criteria and weights vary depending on whether the user is requesting information carried in both the 2001 and 2011 Censuses, or for new information carried only in the 2011 Census.

This segmentation of the outputs into two groups for evaluation is essential. Evaluation of 2011 outputs using information collected in both 2001 and 2011 Censuses needs to address issues of comparability with respect to time, whereas for new information no such requirement exists. In addition, the consultation enables a measure of the utility and popularity of existing information whereas new outputs have typically been suggested for a particular purpose by a user, or have been postulated as a way to present information newly collected by the census.

A similar weighting strategy was previously adopted to assess the requirement for the census questionnaire content. As a result, any information carried on the census form suggests that univariate information and some level of cross-topic analysis is required in the outputs. The extent of cross-topic analysis and the geographical level at which such information is required is not so clear and therefore defining such requirements is a primary purpose of the consultation.

3.1 Criteria and weighting strategy for new outputs

Criteria	Weight
Range of user interest	3
Business case/requirement for the output	3
ONS consideration	3
Cost/resource implications	1

Primarily the weights were developed to look for the most meaningful ways to present new and existing data in order to maximise user-utility and value for money. The criteria with the greatest weights were therefore based on the range of user interest (how much the information is likely to be used across a wide array of users); the business case for the information (how important it is and what it will be used for); and ONS' own consideration. This third criterion reflected ONS' role as the national body responsible for government statistics and included factors (in no particular order) such as:

- Population and demographic requirements;
- ONS own internal requirements;
- A judgement relating to public interest/education and awareness;
- Other uses and requirements not gathered directly via the consultation process.

Finally cost and resource issues were considered, although these were given the lowest weight.

All the topics were given a score of 1-10 for each criterion, using the evaluation mechanism in Annex B. The scores were then multiplied by the weight, summed to give a score out of 100 and ranked.

3.2 Criteria and weighting strategy for outputs based on information carried forward from 2001

Similar weights were adapted for tables based on information carried forward from 2001.

Criteria	Weight
Business case/requirement for the information	3
Range of user interest	2
Comparability with 2001	2
ONS consideration	2
Cost/resource implications	1

The major difference was that an additional criterion is added that relates to comparability. A key use of census information is to see how the country has changed/developed over the previous decade. This requirement has been built into the main series of outputs, and therefore needs to be included as an important criterion. Only the level of requirement for the information in the first instance outweighs the need for comparability, whilst the range of user interest and ONS' own consideration of the value of the output weigh the same. Cost and resource issues are again the final consideration.

The topics were once more given a score of 1-10 using the evaluation mechanism in Annex C. The score was multiplied by the weight and summed to give a value out of 100.

3.3 How the rankings are to be used

It is inevitable that some requests may not be considered as suitable for inclusion. In such instances the ranking system should make it abundantly clear that there is insufficient general need for the output. In this instance, the user could, if required, submit a request via the census tabulation service.

The main purpose of the ranking is to ascertain the range of outputs that users require and to assist ONS in determining the extent to which various dissemination mechanisms meet the needs of users. Following the evaluation the outputs will be designed, factoring in user requirements alongside other key constraints including: disclosure control; dissemination methods; and financial considerations. The ranking will be essential at this stage of the development to understand the degree to which each of these constraints affects the design of the outputs and the extent to which they infringe on users' needs.

2011 Census outputs –statistical disclosure control

4 Progress made in relation to statistical disclosure control and its likely impact upon the design of outputs

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

Over the last six months the National Statistician and the Registrars General have met on a number of separate occasions to consider these matters. Significant progress has been made and a final decision on the details of the policy is likely to be made at the next meeting scheduled for the end of November 2010. Once in place this will enable the outputs team to fully consider the impact disclosure control will have on the proposed set of outputs derived from the consultation with users.

Although the details of the disclosure control policy are not finalised, it is estimated that the impact of disclosure control on the detail of the outputs will be minor; affecting mostly the detailed multivariate cross classifications at the lowest levels of geography (output areas).

5 Other key decisions on factors affecting the output

There are a number of key factors which also shape the 2011 Census outputs. These include:

- The decision whether to adopt best fit or exact fit ward level data;
- The adoption of new pensionable age criteria; and
- The development of workplace zones.

Exact or best fit wards - The issue over whether ward level data should be produced on a best fit basis alone, or both a best fit basis and an exact basis, has still to be decided. The policy of providing exact fit data for wards runs contrary to NS policy, which advocates the use of output areas as the building block to all higher level geographies. The aim of which is to produce a common geographic base and a harmonised approach to disseminating statistics. Methodologists are currently engaged in analysing the accuracy of outputs on both methods and will report back shortly. Following this evidence the Census will make a final policy decision as to whether ward level data should be produced on both a best fit basis and an exact basis, or on a best fit basis alone.

Pensionable age criteria - In 2001 the census outputs used a pension age of 65 for males and 60 for females. Since 2001 government policy has been to raise pensionable age for women to 65 years of age on a progressive basis, starting in 2010. This has implications for the 2011 Census outputs, in terms of the age ranges used in tables and underlying derivations.

In consultation with the labour market topic lead, and to bring the outputs in line with other ONS products and services, the 2011 Census outputs will use a pensionable age of 65 for both males and females. This will be applied at both the household and the individual level and adopted in all derivations used in the outputs.

In addition to the derived variables used in outputs, other derived variables will be created which will exist on the outputs database. These will identify people according to the 2001 definition of pensionable age, and also to show pensioners by exact age on census day, according to the government's policy of progressive change over time. These will be available for analysis via the commissioned table service and the samples of anonymised records.

Workplace zones - In December 2009, the University of Southampton concluded that it was technically feasible to create workplace zones (WZs) for England and Wales by splitting and merging existing output areas (OAs) based on predefined workforce thresholds. The resulting WZ geography will therefore contain both subdivisions and aggregations of current OAs. No final agreement has currently been reached on the statistical disclosure control (SDC) policy for the generation of WZs, although the research has continued under the working assumption that the workforce population threshold will be equal to that of residential OAs, with a lower threshold of 100 and an upper threshold of 625.

As place of work is not explicitly identified by a postcode, it is not possible to identify a workplace threshold as has been done with households in residential OAs. There is however, a legal obligation under the Statistics and Registration Services Act, to protect the identity of "bodies corporate" – individual businesses – and prevent them being disclosed.

As WZs are being used to protect the identity of something which is not identified within the census data, the proposed solution is to combine a number of workplace postcodes. As a single postcode has the potential to represent a single business, the combination of at least three workplace postcodes should protect the identity of businesses. A single workplace postcode will have at least 100 additional workers (the non-disclosive workforce population threshold), added from at least two additional postcodes, to prevent disclosure.

SDC are currently evaluating the proposal as well as those workplace tables requested for the 2011 Census, to determine whether they will be sufficient to prevent disclosure.

To support the development of WZs and to engage stakeholders in the creation of the policy and methodology for how a geography for workplace statistics should be produced, a Workplace Zone User Group was convened in August 2009. Its responsibility is to ensure that user requirements are fully represented in the creation of a methodology for the development of WZs, and to comment on any work that takes place within the project to ensure that it meets user needs.

The group is currently represented by ten local authorities, five central government agencies and two organisations from the commercial sector.

Furthermore, the 2009-10 Census Outputs Geography Consultation was also an opportunity for users to voice their requirements, and in addition to asking users whether there was a requirement for the geography, the opportunity was taken to consult users on what constraints should be used, what tables they would like to see published at that level and what overall they considered would make the geography a success. The results from this consultation are currently awaiting publication, but have been considered in developing the methodology for WZs.

Annex A: Consultation on outputs activities 2009-2011

Topic	Detail	Dates
Main statistical outputs	Regional consultation events	October 2009
	Baseline specification issued	December 2009
	Consultation closes (after 15 weeks)	April 2010
	Assessment of user requirements	May - December 2010
	Revised specifications issued and report	January/February 2011
	Second regional consultation events	February 2011
	Consultation closes (after six weeks)	April 2011
	Second assessment of user requirements	June 2011
	Proposed specification issued and report	June 2011
Census geographies	Regional consultation events	October 2009
	Baseline specification issued	December 2009
	Consultation closes (after 15 weeks)	March 2010
	Assessment of user requirements	April 2010
	End of consultation report	November 2010
	Second regional consultation events	February – March 2011

Annex B: Scoring template for new outputs

Criterion	Weight	Scoring Ranges	Score (0-10)
Range of user interest	3	User interest across all user types 7-10 User interest across some user types 4-7 User interest from sole user type 1-4 No user interest 0	
Business case/requirement for the output	3	Central government resource allocation 7-10 Local or regional resource allocation 6-8 Policy development or monitoring 5-7 Research requirement 3-5 Other interest 1-4 No User requirement 0	
ONS consideration	3	Key output 7-10 Highly priority output 6-8 Moderate priority output 5-7 Low priority output 3-5 Very low priority output 0 -1	
Cost/resource implications	1	No resource implication 7-10 Low resource implication 6-8 Moderate resource implication 5-7 High resource implication 3-5 Very high resource implication 0-1	

Annex C: Scoring template for new outputs, based on information carried forward from 2001

Criterion	Weight	Scoring Ranges	Score (0-10)
Business case/requirement for the output	3	Central government resource allocation 7-10 Local or regional resource allocation 6-8 Policy development or monitoring 5-7 Research requirement 3-5 Other interest 1-4 No User requirement 0	
Comparability with 2001	2	Essential user requirement for comparability 7-10 Comparable some user requirement for comparability 4-6 Comparable, little or no user requirement for comparability 1-3 No longer comparable -0	
Range of user interest	2	User interest across all user types 7-10 User interest across some user types 4-7 User interest from sole user type 1-4 No user interest 0	
ONS consideration	2	Key output 7-10 Highly priority output 6-8 Moderate priority output 5-7 Low priority output 3-5 Very low priority output 0 -1	
Cost/resource implications	1	No resource implication 7-10 Low resource implication 6-8 Moderate resource implication 5-7 High resource implication 3-5 Very high resource implication 0-1	