

CENSUS ADVISORY GROUP

2011 CENSUS OUTPUT STRATEGY UPDATE AG(08)14

Summary

This paper provides an update on the current progress of the Census Output Project and provides an outline of the next steps in the coming six months.

Advisory group members are asked to note:

- The progress to date on user consultation
- The proposed ONS approach to web delivery of 2011 Census outputs
- The proposed ONS feasibility study, due for completion in April 2009
- ONS' aim to make decisions on the way forward for output dissemination by June 2009
- An update on the Statistical Disclosure Control methodological research

Consultation to date

The UK Census Offices began a consultation on high level functionality in Spring 2008 and this is now close to completion. Consultation was web based, with contributors able to complete a detailed on-line questionnaire and make further comment using blogs and forums. The survey had 530 responses, and initial findings showed general support for the RGs high level output aims (see Annex A) such as:

72% of respondents indicated that it was either important or very important to be able to access data from each Census Office from one access point

69% of respondents said that the facility to create user defined tables from given sets of variables (hypercubes) would be very useful

67% of respondents would prefer a smaller set of predefined tables plus user defined output rather than having a predefined set of tables equivalent in scope and quantity to 2001 census tables

89% of respondents would support an arrangement to make sparse census tabular output available under licence or in a safe setting and subject to less statistical disclosure control

83% said they will have a requirement to compare 2011 Census results with those from the 2001 Census.

An internal UK workshop was held in September 2008 at which it was agreed that the aims in the RGs' statement were still sound, but further work is needed to qualify some of the aims e.g. the amount of 2001/11 comparison, and the degree of on-line flexibility in creating user defined tables. This would lead to these aims being subdivided, enabling priorities to be assigned to each subdivision. A further workshop will be held with GROS and NISRA colleagues in early November. This will enable the team to prioritise the aims to identify what we must deliver and what we may be able to deliver, subject to cost, resources and the market place.

Consultation has also taken place with major distributors of census data. ONS Census has represented the UK through membership of a Eurostat Task Force set up to prepare Regulations governing the next round of European censuses and propose technical solutions for data delivery to Eurostat. Representatives of the UK Census Offices are also members of the Data Feeds Network funded by the ESRC that addresses technical solutions for web delivery systems in academia and local government. The requirements of these distributors are well understood and could be met through the proposed ONS approach to web delivery.

Future consultation activities

General consultation on output continues with key users through established channels, e.g. advisory groups, and more widely via roadshow meetings being held in November across England and Wales. Formal consultation on content and detailed functionality is planned from the beginning of 2009.

A new consultation website will be launched in early November, which will open up channels of communication for user engagement on Output issues. This will consist of a blog that anyone may contribute to. There will also be a forum for invited and registered users. Forum members will be able to discuss issues raised through the blog and raise their own points for discussion. A User Council will be established. This will comprise of representatives from a range of User Communities and will enable regular face to face meetings with ONS to address in detail issues raised via the blog and forum.

Further details of the plans for developing our output content consultation will be outlined at the roadshows in early November. This will include developing a framework of core tabular output products on which to consult users through analysis of:

- Business cases made for topic cross-tabulations during the Census questionnaire consultations
- On-line survey results on use and experience of 2001 products
- Tables commonly requested from Commissioned Table Service
- Feedback from Census Customer Services
- EU Hypercube requirements
- Face to face meetings and focus groups across census user communities

ONS corporate plans for web-based dissemination systems

The ONS Executive Management Group have recently endorsed a Corporate proposal to use common, re-usable, and sharable software to support web access to publishable ONS datasets, with an initial system in place in Spring 2009.

The proposed approach is

- a) to continue to use the current data storage ('back end') and presentation ('front end') elements of current ONS web dissemination systems, for example for National Accounts and Neighbourhood Statistics
- b) to separate the two parts of the system, and add a generic interface between the two - an Application Programming Interface (API) that will manage requests from the front end to the back end using a standard query language and data transfer format
- c) to make the API publicly available (on a limited pilot basis initially) to enable partners outside ONS to use the API to access ONS aggregate data, and present using their own front end systems
- d) to further develop the back end systems, giving a wider range of datasets and increasing functionality, potentially using multi-dimensional arrays ('hypercubes')

The current Census proposal is to partner with the ONS pilot project, with a view to understand the feasibility of re-use and to further develop for the 2011 Census.

The pilot project will take the initial steps towards implementing the ONS 'future of the web' strategy. This is likely to include a working 'data explorer' which will allow users to access, explore and download cubes of data from the site. This service will initially be limited to a specific number of 'data cubes', but the availability of ONS data to be accessed in this way will increase following the initial release. The underlying technical solutions for providing this functionality will be the foundations for future developments, including ways for enabling access to 2011 Census data in the future.

Given the scale of census datasets, and the range of outputs and users, there will be significant questions about the feasibility of such an

approach, and the ONS Census Output projects are currently developing a range of questions to be answered through the feasibility and testing stage to April 09.

Depending on the outcomes of the feasibility study, ONS plans to make high level decisions in Summer 2009 on the way forward for Census Output production and dissemination. If the decision is made to embark on a similar approach to that proposed for the pilot, then extensive development will be required, with a clear contingency plan for a less dynamic output distribution method in case the envisaged approach does not succeed.

A series of internal output workshops are planned for November and December to further progress the output strategy work.

Statistical Disclosure Control; Update on Methodological Research

Background

The work programme to decide which method to adopt for Statistical Disclosure Control (SDC) for the 2011 Census consists of three stages :

- Evaluation of the process carried out for the 2001 Census
- High level qualitative evaluation of possible SDC methods for 2011 and agreement of a short-list of methods to undergo full evaluation
- Full quantitative evaluation of short-listed methods resulting in a recommendation to the United Kingdom Census Committee (UKCC)

A UK SDC Working Group has been set up to provide advice, quality review and to steer the work.

Progress

The second stage of the SDC work programme is now complete and the third stage is nearing completion. In July 2007 a review of a wide range of SDC methods was undertaken assessing them against a set of qualitative criteria in line with the policy statement made by the Registrars General. This resulted in the following three SDC methods being short-listed for further evaluation to assess risk and utility quantitatively:

- Record swapping
- Over-imputation
- Australian Bureau of Statistics (ABS) method of cell perturbation

Record swapping and over-imputation are pre-tabular disclosure control methods whilst the ABS method of cell perturbation is a post-tabular

method which uses pre-tabular information. In addition the methods were benchmarked against the methods used for the 2001 Census. The recommendation for the short listed methods received formal approval from the UK Census Committee (UKCC) in October 2007.

The development of the 2011 UK Census Disclosure Control Strategy is progressing well. The quantitative evaluation of the three short listed methods for controlling the disclosure of tabular outputs has considered the assessment of disclosure risk and data utility, assessed the potential for information loss and the overall impact of imposing statistical disclosure control.

The quantitative evaluation was presented to the UK SDC Working Group at the end of September. The UK Census Design and Methodology Advisory Committee (UKCDMAC) are currently reviewing the methodology and will be asked to comment on the validity of the evaluation, results and broad conclusions drawn at their next meeting. The Working Group will meet again in early November to review the UKCDMAC comments then, assuming that agreement is reached, each of the Census Offices will sign-off at their relevant board or committee before a formal sign-off by the UKCC in January 2009.

Future Work

Work to date has focussed on statistical disclosure control strategies for census tables, whilst considering the impact on microdata samples. Detailed work on 2011 Census microdata access and protection is scheduled to start in Spring 2009.

October 2008

Annex A: RGs published aims for 2011 Outputs

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 coordinated quality assurance policy to involve internal and external experts and bodies