Office for National Statistics

CENSUS ADVISORY GROUP

AG (09) 04 CAG

2011 Census Data Quality Assurance Strategy

Introduction

This paper provides an overview of the 2011 Census Data Quality Assurance Strategy. This has been approved by the Census Programme Board and is being prepared for imminent publication. It is aspirational, describing QA activities that have been identified or recommended following experiences from 2001 and before and international peer review. Plans for research and development to implementation the strategy are now being developed, and the final approach will be defined as this research progresses.

2011 Census Data Quality Strategy: Executive Summary

The Data Quality Assurance Strategy should be understood in the context of the 2011 Census Strategic Aims, Critical Success Factors and Quality Strategy.

The strategic aims are:

- to provide high quality statistics that meet user needs
- build confidence in the final results
- provide value for money solutions
- to protect, and be seen to protect, confidential personal Census information

The Critical success factors give tangible measures of whether those aims have been met, for example covering measures of census response rates, question response rates, user satisfaction and security incidents.

The Quality Strategy sets out how we intend to ensure that we will achieve those critical success factors using a quality model with five key strands:

- design quality
- operational quality management
- data quality assurance
- quality measurement and reporting
- project quality management.

The Data Quality Assurance Strategy contributes primarily to the third and fourth strands, and relates to the quality of the data captured from the census questionnaires and the resulting census outputs.

Data quality assurance will start in the two to three weeks before Census day, when management information and early Census returns will provide early evidence of response patterns and characteristics, and will continue through to the publication of outputs. The QA activities described in this Strategy represent what could be done to validate Census results. A process of planning and prioritisation will determine where QA resource is focused. Early research findings could also shift the focus of implementation.

Two distinct but complementary strands of data validation will be conducted in parallel, referred to as 'topic' and 'demographic' QA.

Topic QA involves validating Census data and counts as they pass through the processing systems, using comparator datasets and monitoring changes to data distributions. Demographic QA involves validating, adjusting and accepting Census population estimates at local authority district level, following adjustment for under- and over- coverage.

Both topic and demographic QA will get support from subject experts and both will assess distributions of population sub-groups posing known enumeration challenges such as immigrants, armed forces, students, babies under a year old and young men. Topic QA will seek the early identification of data anomalies so that adjustment can be made to the relevant systems or processes. Demographic QA will necessarily be focused on Census estimates following coverage adjustment and will involve intense input from the Quality Team and a QA Panel. The QA Panel will have responsibility for recommending local authority, regional and national Census population estimates for final sign off by an Acceptance Panel of senior ONS staff.

Topic QA will involve examining data before and after key processes including data capture (for internet responses), the reconciliation of within-household multiple responses, item imputation, coverage imputation, derived variables and allocation of output geographies. As well as comparing pre- and post-process distributions against expected values, cross-tabulations will ensure the internal integrity of the data. Experts in the fields of demography, employment, education, health, housing and identity will guide and inform data validation. Checks will be carried out on variable distributions at the levels of local authority, region, national and cumulative total.

For demographic QA, the Quality Team will build expected Census population estimates using a range of administrative data sources and rolled-forward ONS mid-year population estimates. The Census estimates for each local authority district, region and country will be compared against these and adjustments made as appropriate. A series of demographic indicators will provide further validation, for example sex and dependency ratios. Local authority counts and cumulative totals and cumulative distributions of key population sub-groups will be monitored against non-Census sources. A QA Panel of expert demographers and representatives from the Local Government Association and the Welsh Assembly Government will make recommendations for each local authority district estimate. The QA Panel will operate two parallel sub-groups, one 'priority' group focusing on the local authorities posing particular challenges for enumeration or coverage adjustment, and the other

sub-group approving all remaining estimates. A working assumption is that around 25 per cent of local authority districts will be reviewed by the priority sub-group.

Where there is an unexplained discrepancy between expected values and the Census estimates, the QA Panel will draw on further evidence provided by supplementary analysis which is likely to include use of administrative records. For the 2001 Census ONS carried out a series of studies to improve population estimates in the areas that were hardest to count. This included address matching studies in two authorities, Manchester and Westminster, using administrative address lists from their city councils and the address list collated by the ONS for the 2001 Census. In 2011, the strategy involves bringing this process forward so that new evidence on problem sub-groups or areas will begin in 2010 and be available to inform the decisions of the QA Panel in 2011/12. Where possible, aggregate-level data will be used. Micro-data matching from different sources will also be considered, pending legal gateways including parliamentary data sharing orders if necessary.

Trust and confidence in Census data will be supported by a transparent approach to the methods and results of the QA processes. The Quality Assurance Stakeholder Communications Strategy identifies a wide range of stakeholders, including local authority representatives and other government departments (local and national), academia, professional bodies and international partners. It is envisaged that input from local authority representatives will be coordinated at a regional level through ONS' Regional Statisticians. Comments and suggestions for alternative comparator sources will be actively sought and all QA decisions will be published on the internet. Key deliverables from the QA process will be high quality outputs accompanied by timely metadata in the form of data quality reports and a report for each local authority district.

Prioritisation will ensure that QA activities are delivered on time. Topic QA will focus on validation of early data batches, delivered prior to processing in the case of internet data capture. It will prioritise variables used to generate Census population estimates, directly or indirectly - for example those informing coverage imputation. Management information from the field will signal likely data quality issues and information gaps. Demographic QA will prioritise known problematic sub-groups and local authorities. Supporting evidence from administrative sources, including data matching where necessary and involving evidence from the ONS Longitudinal Study, will identify in advance and fill known information gaps arising from field operations or as a result of QA checks. QA analysis will be pre-programmed as far as possible so that analytic expertise can focus on data anomalies identified through automatic screening checks. This work will be supported by data visualisation and, possibly, spatial analysis techniques. The Acceptance Panel, when considering the recommendations of the QA Panel, will have the options of: i) approving estimates, ii) commissioning further investigations/ analysis, iii) requesting data matching for further evidence or iv) rejecting estimates, even if this means that publication of results will be delayed (although such a decision would of course not be taken lightly).

The 2011 Data Quality Assurance (QA) Strategy has been produced following a review of literature including the 2001 Census Quality Strategy, comments following the 2001 Census and examples of international best practice. Interviews with people

involved in the 2001 Census have also ensured that the 2011 Strategy reflects lessons learned from the last Census. It has been guided by information from the UK Quality Assurance Working Group and the UK Census Design and Methodology Advisory Committee, both involving subject experts and Census managers in Northern Ireland and Scotland and representatives from the Welsh Assembly Government. To check robustness, the QA approach proposed has been applied to information available from the 2001 census processing operation, and this showed that, had the 2011 strategy been in use during the 2001 operation, it would have successfully identified the local authorities in 2001 for which Census estimates were queried.

Looking forward

A period of planning and prioritisation is now required, to implement the strategy within given resources. Alongside this planning, work is underway to prepare for data QA during the 2009 Rehearsal, including development of the Data Quality Monitoring System (DQMS). The DQMS aims to automate QA checking as far as possible and employ data visualisation techniques. Results from the Rehearsal will help to identify the analytic resource required for the QA activities described.

Advisory Group members are asked to note the proposed QA Strategy.

April 2009