



Census 2001 Review and Evaluation

Census Output Production: Executive Summary

ONS is carrying out a review and evaluation of the 2001 Census in England and Wales which will culminate in a Data Quality report and a General Report being published.

Plans for individual reports on specific aspects of the Census operation and a timetable for release have been published.

Each report is written in isolation and is subject to amendments as processing progresses and further information comes to light.

Reports will be released on the ONS website in the form of a high level Executive Summary and a more detailed Evaluation Report.

November 2003

Content	Page
Project Objectives.....	2
Background.....	2
Methodology.....	2
Assessment and Lessons Learned.....	5
Conclusions.....	6

Census Customer Services
ONS
Titchfield
Fareham
Hants PO15 5RR

Telephone: ++44 (0) 1329 813800
Fax: ++44 (0) 1329 813587
Minicom: ++44 (0) 1329 813669
E-mail: census.customerservices@ons.gov.uk
Website: www.statistics.gov.uk/census2001

Census 2001 Review and Evaluation

Project Objective

To release a comprehensive and accurate set of statistics from the data collected from the 2001 Census held on 29 April 2001, in a number of products in a range of media.

Background

The project began in April 1995 and was set up to work in conjunction with the Census Output Policy and Dissemination Project (OP&D) (can be found at www.statistics.gov.uk/census2001/proj_opad.asp) in the identification of strategies for producing and disseminating the results of the 2001 Census. In 1999 ONS made a successful bid to the Treasury Invest to Save Budget (ISB) which funded a third project, Census Access, to enhance the dissemination of Census output.

One of the key objectives of the project was to provide more of the results in electronic format, moving away from the largely paper based reports of the 1991 Census. Printed reports, however, were still to play a significant part in disseminating 2001 Census output such as the reports to the Westminster Parliament and the National Assembly for Wales.

The strategy of electronic dissemination was originally based on the assumption that by the time census results were released, most census customers would have access to personal computers with CD-ROM capability. The strategy had sufficient flexibility to enable ONS to take advantage of developments in dissemination technology. As the 2001 Census progressed, the internet became a strategic tool for the Office for National Statistics (ONS) and output on CD-ROM became a complimentary medium to web based delivery via the National Statistics (NS) website, specifically through the Neighbourhood Statistics Service (NeSS).

The project was divided into four main activities.

- Tabulation of the Census results, including commissioned tables.
- Supplementary geographic material for output.
- Supporting Information.
- Publication in print, CD-ROM, and web.

Tabulation and publication required the procurement of suitable software, and the training and development of staff to produce the results of the 2001 Census. The creation of geographic material required close liaison with teams in other areas who had responsibility for geography in census enumeration and ONS generally. A range of supporting information needed to be gathered from statistical areas to support the tabulations (metadata), and this had to be stored and prepared for inclusion with the data.

Project Management

The project needed rigorous management procedures in place and the Government's Projects IN Controlled Environments (PRINCE) 2 methodology were followed, using staff who were fully qualified in Project management procedures. A Project Board was set up to oversee progress by the three UK Census Offices and to resolve any issues as they occurred. More information on this can be found in the Programme Management Evaluation report at www.statistics.gov.uk/census2001/pdfs/programme_management_es.pdf.

Methodology

Tabulation of the Census Results

In 1998 and 1999 a thorough and wide-ranging review of fast tabulation tools was carried out specifically to assess their potential for use in the tabulation of 2001 Census results. In May 2000, in conjunction with the General Register Office for Scotland (GROS) and Northern Ireland Statistical Research Agency (NISRA), ONS purchased the STR product SuperSTAR. This software suite comprised two major components that were to be used by the Census; SuperCROSS for tabulation by the Census Offices, and SuperTABLE for manipulation of the data by customers.

SuperCROSS allows enormous numbers of census records to be processed at very high speed. Each table produced was examined closely by trained staff to ensure the counts were accurate and correctly presented.

The software was enhanced to incorporate an adjustment method which ensures the confidentiality of the results. This is in addition to the data being ma-

Census 2001 Review and Evaluation

nipulated in advance of tabulation which also protects confidentiality. For more information, see the Disclosure Control Evaluation summary.

An engineer from STR was assigned to ONS to support the Census and ONS generally. This had the advantage of ensuring technical queries were dealt with quickly by highly skilled support staff.

The tabulated results of the 2001 Census fall into five categories.

- Key Statistics

Key Statistics were presented as a set of 30 tables, providing a summary set of results for areas with populations above a threshold of 40 households and 100 people.

- Standard Tables

More detailed cross tabulations were created in Standard Tables for areas with populations exceeding 400 households and 1000 people.

- Census Area Statistics

The need for cross tabulated data for smaller populations was met through the creation of Census Area Statistics. These were created for the same areas as the Key Statistics.

- Area Profiles & Headcounts

Headcounts were created for all areas, either within the tables themselves, or separately where areas contained populations too small to allow for any other information to be released. Area Profiles provide general information for most areas, and also for areas where populations were small and more detailed output would call for disclosure of confidential information.

- Commissioned Tables

Commissioned tables are created in response to customer request when the standard set of output cannot provide the data required. Confidentiality constraints are rigorously applied and some requests therefore cannot be met.

Geography

Each type of table produced was available for a number of geographies in England and Wales.

- Local Government Areas, eg Wards
- Health Administrative Areas, eg Primary Care Trusts
- Parliamentary Constituencies (including Area Profiles for all Westminster constituencies)
- Postal Sectors
- Parishes
- Urban and Rural Areas
- Output Areas

In addition to producing data for the geographical areas listed above, ONS developed a system for automatically creating Census Output Areas. These are small sub-ward areas that form the building blocks for larger geographical areas. The system was designed and programmed with the substantial assistance of Professor David Martin then of Southampton University. A prototype system generated sample output for consultation purposes from 1998 to 2002, and feedback from customers was taken into account in the production version of the system. The final system used 2001 Census data to identify areas with populations above a given threshold, and that also shared similar characteristics. Further information can be found in the Census Geography evaluation.

Supporting Information

Metadata

A large amount of information, collectively known as metadata, is required by customers to enable them to understand the Census results. Information can vary from evaluation reports, such as this, to detailed classifications and glossaries. ONS has gathered, stored and categorised this information in internal databases, and manipulated this material using internet based software in order to disseminate it via web and CD-ROM. ONS also used a range of publishing software, such as Adobe Indesign 2, to prepare the text for inclusion in printed reports.

Census 2001 Review and Evaluation

Geography

In addition, a host of supporting information about the geographies used in the Census has been prepared for release in tandem with the results. In the main, these consist of Output Area boundaries in digitised form and look up files which describe the relationships between these areas and other areas such as Parliamentary Constituencies and Parishes.

Publications in print, CD-ROM and web

Publications are available as printed reports with CD-ROM supplements, and via the web. Further results are made available on CD-ROM and DVD depending on the volume of the data.

Timetable

Results of the 2001 Census have been released to a challenging timetable, with more than half the products being released between September 2002 and September 2003. Descriptions of all existing and planned census products can be found in the Census Output Prospectus available at www.statistics.gov.uk/census2001/op.asp, which includes direct links to all of the Census data available online.

- National Statistics (NS Online)

In order to provide a broad range of access to Census data, significant use of website delivery mechanisms was employed. The first results of the 2001 Census were released via NS Online on 30 September 2002. Age by sex counts for all Local Authorities (LAs) were published on the web and enhanced by the use of graphics such as 'population pyramids' available at www.statistics.gov.uk/census2001/population_data.asp.

Later, census material became available from the Neighbourhood Statistics Service. However, other complementary web delivery systems continued to be employed. For example, the Key Statistics for local authorities in England and Wales was released on 13 February 2003, and involved the development of easily accessible web based summaries.

In addition, electronic 'pdf' versions of all Census 2001 printed reports can also be downloaded via NS Online, some of them accompanied by downloadable sets of tabulations.

- Neighbourhood Statistics (NeSS) website

2001 Census output was made available via NeSS from 13 February 2003, with the release of Key Statistics for Local Authorities. This free to use, publicly available website, brought together the 2001 Census output with local statistics from a range of other government sources. It allowed ready access to the statistics along with associated interactive mapping, enabling a user to be able to search and view the area of interest. The website also allowed easy comparison between areas. Output from the 2001 Census has been a major component of NeSS, with Census Output Areas being the building brick geography for the longer-term future of the system. Further selected output from the 2001 Census (predominantly the Key Statistics series) are being released throughout 2003.

- Other Media

To complement this delivery, a range of CD-ROMs and DVDs have been produced which provide the tabulations in SuperSTAR formats along with free software to view and manipulate the tabulation. Many of the CD-ROMs are supplements to the printed Reports to Parliament and are provided with an easy to use interface that enables tabulations to be interrogated using only a web browser, provides the user with search and navigation facilities, and enables access to the SuperTABLE files for manipulation. To create these supplementary CD-ROMs, Census Output engaged in partnership with CTPi; a company specialising in the development of web and CD-ROM products for Government.

Census 2001 Review and Evaluation

Assessment and Lessons Learnt

How well did it work

The major challenge was to meet the timetable for delivering the results. The tabulation software proved to be capable of meeting the requirements of the project, but a series of enhancements were necessary to meet the strategic and policy demands of the Census and ONS. These made the tabulation work more demanding, resource intensive and time consuming. In particular, a set of disclosure control measures, introduced after the software was selected, required significant changes to software and production procedures. Additional specialised staff were also needed.

The award of contract to CTPi for the production of interactive CD-ROM to supplement the Reports to Parliament proved to be productive and valuable, with lessons learnt on both sides of the partnership.

Most significantly the ONS flagship web dissemination system (NeSS) was used from February 2003 onwards to provide ready access to census results, and associated mapping, by all those with internet capabilities.

Over two billion counts were produced and supplied to customers on the web, on CD and in print, between September 2002 and September 2003. Over 5,000 files of data were released during that time.

The main Census reports to Parliament were released within two months of the target despite some upstream processing delays. First results were released in September 2002 (planned date August 2002); Key Statistics for Local Authorities in England and Wales were released in February 2003 (planned date December 2002); National Report for England and Wales was released in May 2003 (planned date April 2003); Key Statistics for Output Areas and above were released in June 2003 (planned date April 2003). Given the upstream processing delays, the revised release dates achieved were impressive but from the users point of view these were still delays to key outputs, however minimised.

Census Area Statistics (CAS) were released from August 2003 and a programme of future releases of CAS and Key Statistics for area levels is in train. At the time of writing, November 2003, a review of disclosure risk in workplace tables coupled with some technical difficulties in the other products has necessitated a review of the timetable for future releases.

Many technical challenges have been met and a robust production system is in place operated by a large number of skilled staff.

The first results of the 2001 Census released via NS Online on September 30th 2002, set a new record for the number of visitors to the NS website (www.statistics.gov.uk). The Key Statistics for local authorities in England and Wales released on 13 February 2003 set a further record. Levels of interest through the NeSS website have also been high, with a peak of over 50,000 visitors to the site on the day following press coverage of the release of Key Statistics to ward and Output Area level.

Lessons learnt

The full and wide ranging review of potential tabulation software was proved worthwhile; the tool selected has been very successful in producing a huge number of tabulations from a massive amount of data in a relatively short time. The software was later taken up for wider use throughout ONS. The choice of SuperSTAR as a tabulation software enabled dissemination via CD-ROM and DVD along with free software to view and manipulate the tabulations. There was a mixed reception to output on DVD, but in general most customers saw it as a forward looking venture with low implementation costs.

The involvement of academic experts in the development of the Output Area Production System ensured a statistically sound methodology was embedded in the production process. A significant amount of customer evaluation of prototype systems enabled ONS to incorporate feedback and raise awareness of the system and its purpose. Census Output Areas are the current building block for a range of statistics supplied via NeSS, and further development is being conducted across ONS.

Census 2001 Review and Evaluation

Commitment to a range of media for delivering the results was maintained, and a close eye was kept on web delivery mechanisms, especially in the academic sector with which there has been close involvement. Web development was therefore initiated and subsequently dovetailed with the corporate NeSS System.

The process of getting census data from SuperCROSS format to the format required for the website was, however, resource intensive. Systems were developed within ONS to enable and manage the process of taking census statistics from SuperCROSS format through to Beyond 20/20 format - the software used on NeSS for holding and displaying data.

Liaison with census customers who wish to add value to the data by loading it into their own database systems resulted in a specific data format being created and supplied free of charge by census. This facilitated the wider distribution and use of census results.

Changes to geographic policy and disclosure control procedures created challenges for the Output Production project. Dealing with these has been a factor in the failure to meet the original timetable. The impact of changes to policy, in terms of timetable effects, should be a high factor in the decision making.

Resource allocated to output production for 2001 was reduced significantly from that allocated in 1991. Technological advances certainly warranted a significant reduction but with hindsight the reduction was probably too great, leading to under resourcing. The main impact was that work was generally done, in serial rather than parallel, which reduced flexibility and meant one delay had a knock on effect on subsequent outputs and on other projects. Although steps were taken to remedy this once it became apparent, it contributed to some of the timetable delays.

Conclusion

The production of census output is a massive undertaking. Developing and running a vast number of tables, ensuring that confidentiality measures have been applied and checking the validity of the output, has been a challenging task. The sheer scale of output production work cannot be over emphasised. An incredible two billion count cells have already been produced, with more to follow, providing both specialists and the public with an unrivalled information resource.

Rigorous evaluation of the software and hardware to be used for Census Output, and selection of the right tools has enabled an enormous amount of data to be processed as complex tabulations in a relatively short time.

Robust systems were put in place to ensure data was processed accurately and quickly. Staff have been well trained and highly motivated. However, there have been challenging periods when tabulations have been released to customers accurately and on time only by the teams working very long hours.

The needs of customers and the feedback from them was noted and incorporated where ever possible into the development of systems and products.

The full range of media, print, CD-ROM, and web, ensures that all types of customers are provided with data in forms that are accessible and easy to use, taking account of a broad range of skills and understanding of the data. Supporting information is provided in all media to assist the use of the statistics.

Census 2001 Review and Evaluation

Census Topics	Target Dates for Release
Legislation	Published
Non-Compliance (Executive Summary Only)	Published
Data Needs	Published
Geography	Published
Publicity	Published
Data Collection Development	Published
Data Collection Support	Published
Census Coverage Survey	Published
Processing	Published
Annex: Quality of Data Capture and Coding	Published
Downstream Processing	Published
Data Quality	(Executive Summary)
- Question non-response rates	Published
- Disclosure Control (Executive Summary only)	Published
- Data Validation (Executive Summary only)	Published
Edit & Imputation	Published
One Number Census	Published
- Quality Assurance	Published
- Lessons learnt (Executive Summary only)	Published
Output Policy	Published
Output Production	(Executive Summary)
- Part 1:Review of Output Released to date	Published
- Part 2:including Sample of Anonymised Records (SARs)/Origin Destination Matrices	Published
Census Access	Published
Programme Management	Published
Quality Report	(Executive Summary)
General Report	Published

Please note that the dates for release of individual evaluation reports noted above are target dates, and therefore subject to change. For the latest information please visit www.statistics.gov.uk/census2001/reviewevaluation.asp