

Census 2001 Review and Evaluation

October 2003

Downstream Processing: Executive Summary

Content	Page
Project Objective.....	2
Background.....	2
Methodology	3
How Well Did it Work	5
Lessons Learnt	5
Conclusion	7

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.

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 plan, coordinate and manage the development and implementation of the processes which followed the receipt of data from Lockheed Martin (LM) through to the creation of a clean database for the production of 2001 Census Outputs. These processes and systems were known generically as ‘downstream processing’.

The systems were designed to:

- check for data validity and consistency and to correct data where necessary;
- impose confidentiality;
- assure quality;
- provide information for the production of statistical outputs; and
- establish adequate security and access mechanisms.

The majority of work and resources went into the first of these, data validation, consistency and correction.

Background

In order to produce credible and reliable statistics from the 2001 Census it was necessary to manipulate some of the data provided by the capture system as it was possible that some responses were not captured correctly and the public did not always complete their forms correctly. The suite of systems developed to handle this manipulation of the data came to be known as the ‘Downstream Processing Systems’ during the Open Options Procurement (OOP), when the work had been considered for outsourcing. The result of this consideration was a decision to conduct downstream processing in-house, made largely on grounds of cost and in-house expertise available. The project was primarily concerned with:

- collecting, agreeing and signing off requirements for each component of the suite, including Edit and Imputation, Disclosure Control and a system to support Data Quality initiatives. (Note that this report deals with the technical delivery and implementation of these systems. The individual project reports for each component cover the processes within them.)

- designing and specifying the systems to meet requirements;
- developing, testing and implementing the systems to meet requirements; and
- operational running of the systems to timetable.

Pinning down requirements was essential and relied upon the cooperation of a number of other related projects. The design, development and implementation required a mix of business analysts who had a good understanding of Census matters and skilled Information Technology (IT) resources.

The operational live running was dependent on the phased development and delivery of the downstream systems, (over which the project had total control) and the coded data from LM. The UK was divided into 112 processing areas called Estimation Areas (EAs). The first planned delivery of EAs from LM was 31 July 2001, with deliverables contracted from then until 31 March 2002. However, the delivery from LM was delayed and apart from a handful of EAs, the bulk of the delivery was between January and May 2002. This effectively halved the available time for downstream processing and dealing with this revised schedule placed significant demands on the project.

The initial deadline for the live running was to supply the first tranche of Local Authority (LA) output data by July 2002 (to enable the publication of Census First Results in August 2002 to support the Standard Spending Assessment (the process of resource allocation from Central to local government by the Office of the Deputy Prime Minister)) with the final delivery of the Output Supply Database, from which all standard output is produced, at the end of March 2003. Because of the revised schedule for data delivery from LM, the initial deadline was revised to August (to enable publication in September) with the delivery date for the Output Supply database remaining the same.

The project ran from July 1999 (when the decision was made to develop the downstream processes in-house) to June 2003 and cost around £7m. During this time the Office For National Statistics (ONS) Information Systems Division underwent a number of

Census 2001 Review and Evaluation

changes, including a change of name to Information Management Division. Throughout this report all references to Information Systems, Information Management and Information Technology will be referred to as IT.

This report covers England and Wales only although there were many interfaces with the General Register Office for Scotland (GROS) and the Northern Ireland Statistics and Research Agency (NISRA) that had to be considered within the scope of this work.

Methodology

Project management

At the start of the project governance of the work was controlled within the overall Census Programme and the project reported to the Census Operations Board (COB). This Board was disbanded in early 2000 and the project subsequently reported to the Census Programme Board (CPB). COB served as a useful mechanism for bringing together all project managers from related projects.

Staffing

During the life cycle of this project, the ONS appointed KPMG to carry out an efficiency review of the Office. One of their recommendations was that the IT specialists should be consolidated away from the business areas into one IT Division. Up until then the IT staff had been managed and co-located within Census Division. The Project Manager continued to 'manage' the IT staff day to day and physically they remained situated within the business area, although they were line managed by an IT specialist.

Staffing for the project was made up of IT/business analysts, skilled programmers and administrative staff to run and monitor progress of the live running.

Coherence

A key activity for this project was to ensure that all systems within the scope of this project interfaced effectively, between projects, between central systems in

ONS and GROS, between systems developed by LM and systems developed in-house and between packaged and bespoke software.

Requirements Analysis and Design

In consultation with the Business and IT managers the technical Design Team produced an overall system design covering all Census IT systems showing their relationships and dependencies. This formed the basis for dividing the system into discrete components for development. Lower level requirements for these components were gathered at Joint Requirement Planning Sessions.

In June 1998 the forerunner of this project, the IT Strategy Project, made the following recommendations, which were strictly adhered to during the life of the project;

- the Rapid Application Development (RAD) approach would be used;
- the Technical Design Team would work with the relevant development team to produce an outline design to include all known interfaces; and
- the Technical Design Team would ensure the coherence and integration of all census systems.

The RAD approach was described in internal working guidelines, which were based on a well established and widely used analysis and design methodology DSDM (Dynamic Systems Design Methodology). The main emphasis was on iterative development, although guidelines allowed for a more traditional approach where iterative development was not suitable.

The IT Strategy Project also agreed;

- all data would reside on the Sybase database [Adaptive Server Enterprise (Version 11.5)] ;
- Powerbuilder (Version 6) development language would be used. Exceptionally, for reasons of efficiency or interfacing, Visual C++, Visual basic or Access could be used with the agreement of the project manager; and
- Windows NT and PC architecture would be used.

Census 2001 Review and Evaluation

This approach to requirements analysis worked well for many of the Census processes. The early JRP sessions and the subsequent Processing Flow diagram created from that exercise formed a sound bedrock for subsequent analysis. There was also a strong element of prioritisation in processes such as the Edit and Imputation system which echoed the 'Must have', 'Should have', 'Could have', precedence laid down in RAD guidelines.

Software Development and Testing

All the processes that ran under the umbrella of 2001 Downstream Processing, with the exception of the One Number Census (ONC) System and the processes to support Census Geography were developed and run against the Sybase database. The majority of programs were written in a combination of C, Transact SQL and Powerbuilder.

The batch systems were developed in a 'traditional' manner. Each module/function within the system was developed separately and tested in isolation. Test plans, test data and expected results were produced by the programmer who was developing the module.

The on-line systems were developed with a RAD approach. So, basic screens were developed, then additional functionality was built on to them as required by the customers.

System testing carried out by the development and coherence teams ensured that the modules / functions all linked together. Customer testing was used to confirm that the system did what the system owner expected.

The systems developed ranged from the very complex to relatively straight forward batch processing. There were over 40 processes that each EA had to pass through before releasing to the Output Database.

System Architecture for Live running

The main issue was coping with the volume and complexity of live running. In order to plan and set up the operational environment, the known processes,

processing order and comparative predicted run times for each process were listed with assumptions about which processes could be run concurrently on a single server. A volumetrics exercise to predict the size of all databases, input files, output files and error logs etc, based on information from volume testing and rehearsal running was carried out. Analysing this information along with estimates of predicted delivery rates of 3–4 EA's per week, an initial proposal for processing capacity and data and image storage was produced. The resultant architecture also had to be scalable to allow for contingency.

In order to run the 40+ processes for all 112 EAs to timescale a Process Control System was developed which enabled all the runs to be scheduled and monitored effectively to enable maximum throughput. The control system also enabled the effective management of all back-up and recovery processes.

Issue Management

Inevitably in a project of this nature there will be conflicts of interest and the need to prioritise requirements. The Data Quality Review Panel system was implemented to provide a mechanism for reporting and resolving issues.

Other related responsibilities

The following were also within the remit of this project:

- Ensuring that appropriate security measures were in place to protect access to the data, particularly implementing the necessary communications links between the three external sites, GROS, NISRA and Widnes (the Lockheed Martin processing site).
- Providing the necessary technical and business expertise to design and implement the Output Database.
- Management of census form images.
- Providing the necessary support to other related systems which use census data, such as Longitudinal Study, Survey Non-Response and Sample of anonymised records (SARs).

Census 2001 Review and Evaluation

How well did it work

Achievements

Overall, the project was a huge success. The key date of August 2002 for delivering the basic age and sex counts was met, despite the delays in receiving the data from Lockheed Martin, providing for the publication of First Results on September 30th. The project then dealt with the need to re-run a major part of the downstream processing due to the additional adjustment for One Number Census (ONC) dependency. All revised dates were met and the re-running of a large amount of processing was accomplished in a very tight timescale. One important element contributing to these achievements was the commitment and dedication of some key individuals.

Staffing

There were a number of difficulties in retaining skilled resources for the duration of this project. For a future exercise of this size and complexity, ONS need to ensure there is a strategy in place to attract and retain key specialist/technical staff for the life of the project.

Coherence

The setting up of a coherence team was a key factor in the success of this project. During early planning for 2001 the need for a role to ensure coherence between all systems and technical environments was identified. A Coherence team was set up which gathered requirements from all projects; it did this, by taking the lead in Joint Requirement Planning sessions and converting the requirements into the 2001 Census Information Systems Flow Diagram. This diagram represented all the flows of information between projects, with a high level view of the processes involved and proved a highly effective means of facilitating coherence. It was crucial in enabling a small group of highly skilled and knowledgeable people to have a full understanding of all the issues and interfaces between systems and projects. Their ability to respond to problems, deal with requests for change and generally react and report on any potential impact of issues, proved invaluable during both the development and live running cycles.

System Architecture - Sybase Environment

The setting up of a new technical environment for ONS, primarily to process the 2001 Census was another achievement for this project. The technical environment was chosen through an open procurement project with the objective of procuring a corporate ONS Database and Development tool. Sybase were awarded the contract to supply the hardware, software, and training and consultancy. In the event we required more consultancy than originally intended, because of the large numbers of leavers we had in the initial stages. The consultants supplied by Sybase were generally of a high calibre and assisted us in achieving a robust and reliable environment throughout the live running cycle.

Image Management

The need to be able to effectively store and manage the form image data, so that various users and technical support staff could access the images efficiently, was identified early on in the project. The initial recommendation for a system to support this activity was to contract out the development. However, this proved to be an expensive solution and a cost benefit analysis recommended that an in-house solution could be developed. This proved to be an effective and reliable solution.

Lessons Learnt

Project Management

There were weaknesses in the Board structure especially after reorganisation which led to difficulties determining who had responsibility for taking decisions. Therefore Downstream Processing often had to work on assumptions and informal decisions rather than clear-cut decisions, resulting in an element of risk and uncertainty. In a future Board structure it must be clear to all what the mechanism is for making timely decisions. This could be supported by a more robust Programme Management set-up with the expertise to understand and advise on interdependencies and impacts across the whole programme.

Census 2001 Review and Evaluation

Staffing

As mentioned earlier, ONS centralised the provision of IT resource during this project. From the perspective of a particular large programme such as the Census, it will always be necessary to have dedicated IT resource and a considerable degree of direct control and management. For the Office as a whole, however, there are advantages in having the IT staff centrally managed. For the future, much will depend on the way the Office as a whole is organised. An important factor in the success of this project was that it had the priority and funding to pay for additional skilled resources as required. In order to retain good, appropriately skilled staff a bonus was offered to those in-house staff who stayed committed to the project. This was hotly debated at the time, but none the less worked well and proved not to be divisive as some had feared.

Requirements, Analysis and Design

A realistic assessment of resources required for such a project is needed early on in the project cycle. The requirements for some systems were too complex given the level of resource and time available. For example, the imputation system within the Edit and Imputation part of Downstream Processing was far more complex than the imputation system developed within the One Number Census (ONC) System. However, it could be argued that the overall impact of a simple imputation system within ONC had more of an impact on the overall quality of the data in terms of missingness ie. missing people versus missing data items. Ensuring that all requirements gathering and resulting design is managed and coordinated by the same design team may lead to a better understanding of the potential issues and their impact.

Requirements and decisions need to be agreed and signed off at an earlier stage. The late decision resulting from a change in strategy as to what would be outsourced and what would be delivered in house meant that the project faced severe time constraints. Additional funding and resource had to be acquired to overcome these constraints.

The design work carried out in the downstream processing project did not encompass data collection and for the future we recommend that it should. Form content, including form identity, form control and

geography interfaces were fundamental to the processing system but the project had little involvement with the design of those. For example the form identity code was a series of alphas and numerics which had to be hand written by the enumerators on each form. This data item was fundamental to a number of systems. Bad hand writing along with mis-recognition at the scanning stage caused processing problems further down the line. A coordinated approach across all projects needing to use a common identity code would have prevented this.

Development and Testing

Although the original requirements for the edit and imputation system were reduced, the resulting system was complex and difficult to develop and therefore test, effectively. The bulk of the testing was carried out by using IM resource. For any future exercise it will be important to ensure adequate customer testing resource is available in the business area and to ensure that those with responsibility for this important role are familiar with the concepts of version control and rigorous testing needed for large scale production systems. There would also be advantages in including all development activities within the Downstream Processing Project, or its equivalent. Building and implementing large computer processes to effectively manage, control and process this amount of data needs a strong IT background.

Accountability and ownership for these processes needs to be with the project manager responsible for delivering the systems and not the people responsible for the methodology.

Data Quality Review Panel (DQRP)

During development and live running inter-project issues were raised through the DQRP system. The purpose of this system was to alert all those who were likely to be impacted by an issue so that they could inform the decision making process. The options for resolving the issue were presented to the Quality Manager. The basic idea was good, but in the cut and thrust of a large operational environment it is important that decisions are made quickly. On several occasions this was not achieved, which resulted in operational difficulties. The balance between (a) resources allocated to this work, (b) the depth of assessment required and (c) the timeliness of decision making must be carefully considered in deciding any future processes.

Census 2001 Review and Evaluation

GROS / NISRA Interface

The resource needed to implement the different requirements for the 3 offices was governed by the complexity of the requirement rather than the volume of records to be processed.

There needs to be appropriate governance and funding agreements put in place at the outset, to influence the setting out of requirements and any subsequent change requests.

Conclusion

Overall this has been a successful project, with all key dates within its control met, quality systems designed and developed, no major processing problems and coming in slightly over budget because of the need to recruit external IT resources.

Census 2001 Review and Evaluation

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