national
STaTiSTiCS

2001 CENSUS PROCESSING: PLANS AND PROGRESS

1.   This paper outlines plans and progress for the processing of the 2001 Census. It covers:

- **an outline of key processing stages**

- **an outline timetable**

- **progress**

2.   Advisory Group members are asked to note the paper.

Simon King
Head of Census Processing Branch
ONS
30 October 2001

1 Outline of key processing stages

1.1   **Capture, coding and delivery of data**

Lockheed Martin capture, code and deliver data (computer records for each household and person, together with images of the forms, and an audit trail to show where each field was created) on an area basis. There are 101 such areas in England and Wales, deliveries for which are planned to arrive on a schedule spanning September 2001 to April 2002.

1.2   **Downstream Processing**

Downstream Processing comprises a complex set of computing and manual activities which ultimately create a full and final database of Census records from which all Census outputs can be generated. These activities are run by area and cover loading and checking, edit and imputation, and the One Number Census processes.

1.3   **Load**

The files are loaded onto a database during which they are checked at two levels. First, to ensure that all files are present, readable, and that the structure of files and records meets the interface specification. (e.g. is every person linked to a household?) Secondly, checks are made to ensure that the data content and data values are within specification (e.g. nobody of 150 years old!). Problems are reported to Lockheed Martin who will re-supply if necessary.

1.4   **Edit and Imputation**

If a member of the public has not answered a question or has made an inconsistent response (e.g. four year old doctor) the imputation system establishes a valid entry by drawing on valid responses from demographically similar households. Further details on edit and imputation were given in Advisory Group paper (00) 13.

1.5   **Adjustment for undercoverage**

The Census Coverage Survey (CCS) was conducted a few weeks after the Census for about 350,000 households. Data from the CCS is captured and coded in a similar way to Census data by Lockheed Martin. It then passes through load and edit and imputation, and the ONC process, and estimates can then be made of under-enumeration. After that, the Census database is augmented with additional records to represent those estimated to have been missed. Further details of the ONC are available on the website and in Advisory Group paper (01) 03 and in the ONC User Guide.

1.6   **Data Quality Management**

For each area, a range of quality checks and comparisons are carried out for two main purposes. The first is to look for any systematic error which might have been introduced at the Lockheed Martin Processing stage. (for example all garden centre employees coded as childcare <u>nursery</u> employees). We have ninety days in the contract for this activity and at any time we can seek to have Lockheed Martin correct and re-supply.   This supplements an array of quality control checks carried out by Lockheed Martin themselves – the quality plan was agreed between the Census Offices and Lockheed Martin.

1.7   **Create output database**

The data is then restructured suitable for output. At this stage a number of new variables are derived (e.g. family relationships) and disclosure control measures applied.

2   Outline timetable

The following table shows the key dates in the processing plan.

Lockheed Martin commence scanning	June 2001
Lockheed Martin deliver census data area by area	August 2001 - April 2002
Lockheed Martin deliver CCS data	October 2001
Downstream Processing area by area	September 2001 - June 2002
Release of mid-year estimates	August 2002

3   Progress against plan

Lockheed Martin started processing in June 2001 as planned. Their strategy was to have a controlled ramp-up through June and be able to handle the full production throughput rates in July and beyond. The full throughput rates have not yet been achieved due to a number of factors including inefficiency in the 'key from image' process. That is the process which presents keyers with the images of the forms where the automatic character recognition software could not determine the response, so that it can be keyed in manually.

Lockheed Martin have delivered the CCS data and have so far captured and coded approximately 25% of census forms which is somewhat behind target. In order to complete the capture and coding by the end of March 2002 they need to increase throughput. They have provided us with a rectification plan which involves recruitment of additional keying staff and working additional shifts. To date over 200 additional staff have been employed. We are monitoring the effectiveness of these rectifications weekly so as to ensure they hold to the end date.

A further impediment is that the software which is used for collating and exporting the data to us is working more slowly than required. This is being addressed with the deployment of additional hardware. At present, data from two areas have been delivered and processed through the Downstream systems. Because of this, the timeframe for Downstream Processing has been condensed. A set of measures is therefore being implemented in order to increase the Downstream throughput so that the key target for the release of mid-year estimates can be met.

Simon King
Head of Census Processing Branch

30th October 2001