ONS(ONC(SC))00/07

ONE NUMBER CENSUS STEERING COMMITTEE

One Number Census Project Update

1. This paper describes the progress on the One Number Census since the last meeting of the ONC Steering Committee on 1 July 1999. It also outlines the work planned over the next 6 months.

2. The Steering Committee are asked to:

a) note the progress made and

b) provide any comments at the meeting on the 9th February 2000, or in writing by 23rd February 2000.

Marie Cruddas
Census Division
Office for National Statistics
Room 4200W
Segensworth Road
Titchfield
Fareham
HANTS
PO15 5RR

January 2000

One Number Census Project Update

Marie Cruddas

1. Introduction

1.1 This paper reports progress on the One Number Census (ONC) project since the last meeting of the ONC Steering Committee on 1st July 1999. Work planned for the next 6 months is also included.

1.2 The work on implementation and development of the Census Coverage Survey (CCS) in 2001 is covered by a separate CCS Project under the management of John Dixie. ONS(ONC(SC))00/05 provides an update of progress on the CCS project since the last meeting of the Steering Committee.

1.3 As reported at the last Steering Committee the work planned for this period was mainly further development of the prototype ONC systems and evaluation of the Census Rehearsal. While we have made good progress on the first of these, delay in the delivery of the data has put considerable pressure on our ability to evaluate the ONC methodology. This has particularly affected the timing of the evaluation of the ONC Matching process.

1.4 The CCS data and Census data for Leeds has now been delivered. Before this data can be used for evaluation there are some data completeness and quality issues to be resolved.

1.5 An additional pressure that has only become apparent recently is the requirement to have the specification of requirements for the CCS in place by the end of April so that the processing system can be fully developed and quality assured. There is a risk that some aspects will have to be decided without evaluation being possible. We are revisiting our evaluation plans to prioritise issues.

2. Update on key areas of research

Census Coverage Survey Design

2.1 Since the last Steering Committee meeting many of the implementation issues surrounding the design and the selection of postcodes have been examined. However evaluation of the design implemented in the Rehearsal has been delayed due to the non-availability of the captured data.

2.2 In the absence of data has been necessary to formulate judgements as to the likely outcomes of the evaluation of the CCS Design in the Rehearsal – for example the composition and number of categories of the Hard to Count Index - for inclusion in the ONC Methodology paper ONS(ONC(SC))00/01.

2.3 The data used in the construction of the postcode sampling frame needs careful consideration and geographical issues need to be taken into account. In particular, postcodes are not controlled by ONS and are not designed for the purposes of carrying out a survey. Therefore, they may change between sample selection and fieldwork. A strategy for dealing with these problems has been developed, which involved working with the CCS Project, Census Geography team and ONS Geography experts. The data sources that could be used to form the sampling frame for 2001 have been examined in greater detail, and based on their availability a timetable for the selection has been drafted in conjunction with the CCS Project.

2.4 These solutions, together with the Rehearsal evaluation will be included in a paper for the June Steering Committee recommending the final sample design.

Matching

2.5 A Computer Assisted Matching System (CAMS) has been developed. Most recently this has been refined in an attempt to make it more user-friendly, an online help facility added and System and User Guides produced.

2.6 However the next stages in the Matching project require Rehearsal data and this is only now starting to become available.

2.7 Jennet Woolford has attended the GSS(M) Taskforce on Matching and its Matching Methodology Subgroup. The ONC Matching project has been taken up as a case study by this subgroup.

Demographic Analyses and Administrative Records in Support of a One Number Census

2.8 The uncertainty intervals for national level demographic estimates are being updated and an age/sex breakdown is also being produced. This work is a development of that described in ONS(ONC(SC)99/05.

2.9 The national methodology for producing uncertainty intervals cannot be extended to the sub-national level. The plan instead is that it may be possible to use the variability between previous Censuses and mid-year population estimates to inform sub-national variability. This work is being taken forward by jointly by P&VS and Census Division with advice from Ian Diamond and Ludi Simpson.

2.10 Progress on evaluating the use of administrative registers is contained in ONS(ONC(SC)00/04.

2.11 Further work is planned to build on the agreed quality assurance and contingency strategy prior to the consultation. This will include:

a) clear descriptions of the data to be used as comparators in the QA process and, where appropriate, an assessment of their reliability for example error bounds. A description of how the comparators will be used, what constitutes a difference and the geographical level for which the comparison can usefully be made. A description of the criteria to be used to decide whether to adjust further the census estimates, and the

action to be taken where it is agreed that the census based estimates require further adjustment. Some examples to illustrate the QA process and contingency action

b) consideration the implications of the timing of the production of mid 2001 population estimates for the quality assurance process.

Imputation

2.12 Development of the Imputation system passed from Southampton University to ONS in June 1999. Work since that date has concentrated on fully understanding, improving and developing the software to ensure we have an operational system to trial in the 1999 Census Rehearsal.

2.13 The system is written in SAS and the code has now been optimised for use with data from the 1999 Census Rehearsal. Improvements to the code have also led to a significant reduction in the need for manual intervention thus making the system as automated as possible. This, combined with the use of higher specification hardware, has led to a significant improvement in processing time.

2.14 Development and testing has used a simulated data set containing 500,000 people which is comparable to a single Design Group and forms the processing block for the One Number Census. A single Design Group currently takes 12 hours to process.

2.15 In the past few months a significant amount of time has been spent considering the wider issues associated with placing imputed households and individuals into the Census database. This has included close liaison with

- those developing the Edit and Imputation System (EDIS). To ensure that both systems adopt a consistent approach and are complementary to each other

- with Census Geography to develop methodology for the placement of imputed households into postcodes

- with staff from Census IS to enable the development of external interfaces between ONC Imputation and the Census and Geography databases.

2.16 The ONC Imputation system is now awaiting the availability of Rehearsal data from the earlier ONC processes. This will provide the first opportunity to fully evaluate the system using 'live' Census data.

2.17 After a full evaluation of the methodology and computer system effort will concentrate on working towards the development and implementation of a fully operational system for the 2001 Census.

3. Update on Implementation

3.1 There has been ongoing development of the ONC system a) for use with the Rehearsal data in order to evaluate the methodology for the ONC and b) to ensure we have an operational system for 2001.

3.2 A ONC Dry Run has been conducted. This involved a simulated data set equivalent in size to a Design Group being run through each of the ONC processes in turn, and demonstrated that all the systems and interfaces work satisfactorily. The Dry Run allowed timings for the exact matching, estimation, and imputation processes to be approximately estimated, and also highlighted a number of issues which have now been addressed.

3.3 Integration of the ONC system with other processes and the development of an operational system for 2001 and is being progressed within the overall Census downstream processing development. The fact that IS resources have been dedicated to the development of the ONC is welcomed and a work plan is being devised currently.

4. Consultation/communication

4.1 Tim Holt presented the paper 'Risk in Official Statistics: A Case Study of the 2001 One Number Census Project' at a plenary session of the RSS Conference on Risk 12-15 July 1999 in Warwick. The paper has since been submitted JRSS series D and Dr. Holt has made a further presentation of the paper to the OR Society.

4.2 Presentation were given at the Federal Committee on Statistical Methodology Research Conference 15-17 November in Washington by Marie Cruddas on the One Number Census methodology and by John Dixie on plans for the Census Coverage Survey.

4.3 Dr Howard Hogan Chief Decennial Statistical Studies Division, US Census Bureau visited Census Division, ONS and University of Southampton on 10-14 January 2000 to discuss the ONC methodology. Of particular interest to him is the Dual System estimation.

4.4 A paper on the ONC methodology has been accepted for the Population Association of America Conference to be held in Los Angeles March 2000.

4.5 A paper on the ONC imputation methodology has been submitted to Series A of the Journal of the Royal Statistical Society.