2009 Census Rehearsal Evaluation
Downstream Processing

December 2009

Table of contents

2009 Rehearsal – Downstream Processing

1. Definition and scope for the rehearsal

Census data, once captured and coded upstream by the data capture and coding service provider requires careful checking and modification before reliable outputs which meet users' needs can be produced. Modification is required to take account of any missing responses on the census questionnaires; to remove duplicate questionnaires; to correct data captured erroneously during the data capture operations; to remove inconsistencies (such as children born before their parents); and to prepare the data for dissemination to users. The systems developed to clean and adjust the data, so that a fully consistent database of census records can be prepared as the basis for output dissemination, are collectively known as 'downstream processing' (DSP).

For the rehearsal, the aim was to develop steps 8.1 to 8.3.1 of the DSP system and to run rehearsal data through these steps, while also testing the full functionality of all other system elements. This would allow the evaluation of the methodology behind DSP, the quality of the Statistical Analytical Software Electronic Data Interchange (SAS EDI) software tool for the system and the quality of the upstream data.

Steps 8.1 to 8.3.1 are:

- load & validation (8.1);
- remove false records (8.2.1);
- multiple response (8.2.2);
- derived variables – simple input (8.3.1); and
- filter rules (8.3.1).

The system was expected to have full functionality with regard to:

- process control;
- metadata management;
- data quality management; and
- data security.

Whilst the design of the system should provide a platform for the continued development of DSP for 2011, the evaluation of this is outside of the scope of the rehearsal.

What was tested:

- Methodology and IT systems for each of the steps outlined above
- Operation of the overall DSP system for each of the steps outlined above

What could not be tested:

Steps 8.4 onwards. These will be developed during 2009 and 2010 and will be tested using rehearsal and other suitable data sources.

2. Evaluation findings

Pre – determined evaluation points

Description	Success Criteria	Outcome against success criteria	Recommendation	Timeframe
Step 8.1 Load & Validation	Data loads correctly No invalid data passes to later DSP stages.	All data was successfully loaded into the system. The validation and profiling rules worked effectively and no invalid data was passed to later stages.	No changes for 2011, though see data transfer method in other evaluation points below.	
Step 8.2.1 Invalid Questionnaires	A sample of records that the downstream processing system identified as passed or failed records are found to be genuine when quality checked manually.	The original specification for invalid questionnaires forced too many genuine records to be rejected as 'invalid'. The methodology was revised based upon this and re-run on all deliveries, and as a result of this worked successfully.	Keep the revised algorithm for 2011, ensuring a consistent approach across the UK.	March 2010
Step 8.2.2 & **8.2.3** Multiple Responses (MR)s	Sample of MR decisions shows consistency and accuracy.	Unable to develop the automated MR system in time for Rehearsal, due to the complexity of the MR resolution methodology. A 'manual' system was used instead. This worked successfully and allowed	Continue to develop the MR methodology and automated MR system.	March 2010

Description	Success Criteria	Outcome against success criteria	Recommendation	Timeframe
		evaluation of the methodology. The methodology will be improved on the basis of the findings.		
Step 8.3.1 Filter Rules	Sample of filtered records are correct.	All filter rules were successful, although some minor methodological and IT tweaks were made during processing.	Further evaluation of the filter rules methodology is required, particularly with regard to the whole questionnaire, to ensure they function appropriately.	January 2010
Step 8.3.1 Key Derived Variables (DV) – Simple Input	Sample of DVs are correct.	All derived variables worked successfully.	No changes for 2011, although ongoing review will continue, especially as the other DVs (e.g. Output) are developed.	March 2010
Data Quality Management System (DQMS)	DQMS provides required functionality for analysis of data quality.	We did not develop a DQMS for rehearsal, instead we carried out data quality analysis using a manual processes.	Complete the development of the DQMS system and test using rehearsal data.	August 2010
DSP System	System provides full functionality for process control, metadata management, data quality management and data security.	The system successfully delivered metadata management and the required level of data security. Process control was achieved by a manual rather than automated system.	Develop automated process control functionality.	August 2010

3. Other evaluation points

Data Transfer - The secure transfer of rehearsal data from the Paper Data Capture (PDC) site to DSP system was achieved by the physical transport of encrypted media carried out by in-house staff, which meets Her Majesty's (HM) Government security standards. This process was labour intensive and required the co-ordination of a number of staff, with the result that each output took over a day to be loaded. An automated electronic transfer method would be more efficient and enable greater integration between upstream and downstream processes.

Recommendation: Investigate the feasibility of a secure automated electronic data transfer process which meets HM Government security standards, but which also significantly reduces transfer times.