

2009 Census Rehearsal Evaluation Paper Data Capture and Coding

January 2010

Table of contents

1. Definition and scope for rehearsal.....	4
2. Evaluation findings	6
3. Other evaluation points	15

2009 Rehearsal – Paper Data Capture and Coding

1. Definition and scope for rehearsal

There is a vast amount of information to be processed once the completed census paper questionnaires have been returned. They are first processed through automatic data capture and then an exercise to code data is carried out in a pre-determined sequence for each capture/coding type.

Questions containing text responses (such as those relating to occupation and industry) are coded by a combination of automatic and computer assisted manual coding. The automatic coding systems use classifications and indexes which are, so far as possible, consistent across the UK Census Offices.

During this rehearsal, General Register Office for Scotland (GROS) statisticians monitored the captured data as it was processed to check progress and data quality. Progress was monitored by the operational team on-site after which data was delivered as outputs to GROS for further data quality checking.

Once the data was checked and securely archived, the paper questionnaires were securely stored and retained for further evaluation. Following the completion of the 2011 Census, they will be destroyed in line with government security guidelines, protecting the privacy of census information while allowing paper questionnaires to be recycled. All systems and storage media used in the rehearsal will be securely erased, in accordance with government security standards.

The solution can be broken down into 3 key stages:

- Warehouse;
- capture and coding and
- data output.

The Paper Data Capture and Coding (PDCC) Team co-ordinates and manages, in conjunction with the service provider, the delivery of each key stage, which includes:

- requirements clarification and service provider education and training;
- detailed specification and design;
- implementation (including Operational Readiness Testing);
- creation and co-ordination of coding aids including classifications, indexes, geography files and a Gaelic glossary;
- management and monitoring of operational running;
- quality assurance of services and outputs; and
- evaluation of operational services and outputs.

What was tested:

- specification of service;
- testing and implementation of service; and
- delivery of service.

What could not be tested

All items above were tested.

Summary of Rehearsal Outcome

The rehearsal successfully demonstrated the capability of the Paper Data Capture and Coding (PDCC) operation to process census returns and output this data to GROS as required. The joint evaluation carried out by GROS and the service provider has been carried out in a positive and open manner and has highlighted a number of useful improvements and enhancements that will be put in place for the 2011 Census. These are detailed in the following sections.

2. Evaluation findings

- Pre – determined evaluation points

Description	Success Criteria	Outcome against success criteria	Recommendation	Timeframe
Warehouse - Specifications	<p>a) Requirements and service specifications provided within agreed timescales to relevant standards.</p> <p>b) All related service catalogue items provided within agreed timescales to relevant standards.</p>	<p>Achieved.</p> <p>a) All Service Level Agreements (SLAs) and requirements were provided.</p> <p>b) There was some minor timescale slippage, but it was possible to prioritise by need. All items were provided at point of need.</p>	<p>The major requirements of the warehouse facility were met. However, two minor recommendations have been added at point 11. and 12. of the “other evaluation points” section below.</p>	N/A
Warehouse - Testing	<p>System testing, Operational Readiness test (ORT) & Operational Test (OPT) all successfully concluded.</p>	<p>Achieved.</p> <p>Testing went well because there was a good planning document developed for the ORT phase and this document was followed.</p>	<p>Although no specific issues were raised in the warehouse testing phase, the need for dedicated testing resource and effective scheduling outlined in the Capture & Coding section below is relevant for all testing for 2011. An additional testing manager has been recruited for 2011.</p>	<p>See capture and coding section below.</p>

Description	Success Criteria	Outcome against success criteria	Recommendation	Timeframe
Warehouse – Delivery of Service	<p>a) Joint working between all relevant parties.</p> <p>b) Boxes for all delivery types are all correctly registered and available for next phase.</p> <p>c) Management information reports produced as required and requested.</p> <p>d) Service levels met.</p>	<p>a) Achieved. Minor issue raised over:</p> <ul style="list-style-type: none"> • timing of sending the consignment file to the service provider when the field office was close to PDC site; • handling of exceptions at point of delivery; and • registering errors in the receiving system. <p>b) Achieved.</p> <p>c) Achieved.</p> <p>d) Achieved.</p>	<p>a) The service provider and GROS need to investigate a solution in cases where the field office is closer than the minimum consignment file notification period;</p> <p>b) The service provider, GROS and the logistics service provider (LSP) will discuss, develop and agree a process/method for recording exceptions at point of delivery. The LSP to ensure that the receiving system can register errors and allow them to report issues back to LSP.</p>	

Description	Success Criteria	Outcome against success criteria	Recommendation	Timeframe
<p>Capture & Coding and Output (i.e. Phase 2) - Specifications</p>	<p>1) Requirements and service specifications provided within agreed timescales to relevant standards;</p> <p>2) All related service catalogue items provided within agreed timescales to relevant standards.</p>	<p>Partly Achieved.</p> <p>a) This area was more complex than at first expected but a lot of good work has been done by both parties and we expect the co-operative approach that developed to continue.</p> <p>b) The system and service design was a lot more complicated than originally envisaged by all parties, and that ultimately put pressure on the resources and schedule going forward. The remaining time available to deliver the project was very challenging and led to a number of issues with delivery.</p> <p>c) However, despite the pressures and problems, both parties worked hard, in a collaborative manner to deliver against the original objectives of the project.</p>	<p>a) There will be pre-determined check points during the 2011 development phase to ensure that the solution will deliver the service required.</p> <p>b) As assurance of the service (or confidence building) is a process that permeates the whole project, GROS will seek appropriate types of assurance from the service provider during each phase of the work. For example; assurance during specification and design stage will be via review and walkthrough.</p> <p>c) The timetable must be agreed with each stakeholder before it is finalised. Progress on delivery of items must be monitored by the integrated project team (IPT) lead.</p>	

Description	Success Criteria	Outcome against success criteria	Recommendation	Timeframe
		<p>d) The service specification documents delivered by the service provider focussed almost entirely on the technical solution. With hindsight, both parties agreed that the service specifications should have focused less on the system itself and more on how the service would be delivered and SLA's met.</p>	<p>d) The service specification for 2011 must provide assurance for the delivery of the service and not the technical solution. Product descriptions will be agreed for the products that form the service specification;</p> <p>GROS requirements specifications must be complete and consistent and the process must include a clarification process with the contractor to mitigate the risk of service delivery errors.</p>	
<p>Capture & Coding and Output (i.e. Phase 2) - Testing</p>	<p>System testing, ORT, OPT and Optical Character Recognition (OCR) Tuning (capture) all successfully concluded.</p>	<p>Achieved.</p> <p>a) Both parties agreed that the testing phases should merge, due to the timescale issues outlined in the specification section above.</p> <p>This was not entirely successful given the short timetable within which GROS and the service provider had to work following contract award.</p>	<p>a) A realistic schedule for these testing phases (including contingency) will be developed for testing for 2011.</p> <p>Properly resourced testing teams will be planned for by GROS and the service provider.</p>	

Description	Success Criteria	Outcome against success criteria	Recommendation	Timeframe
		<p>b) Testing through all phases did complete but issues arose, which have been addressed through the recommendations identified in this report.</p>	<p>b) Clear processes will be agreed for the release and regression testing of any changes made to the systems. Sanity check of system at start of system test. This would use a small test set to test simple scenarios before moving on to more complicated scenarios.</p> <p>A formal regression test approach will be agreed.</p>	
<p>Capture & Coding – Delivery of Service</p>	<p>a) Joint working between all relevant parties, including communication with Geography Team.</p> <p>b) All materials including paper and internet processed with minimal need to handle or rescan to achieve required quality.</p> <p>c) Correct application of coding rules, sample coding</p>	<p>a) Achieved see comment 4 in “Other Evaluation Points” section below for comment on Geography.</p> <p>b) Achieved.</p> <p>c) Partially Achieved - Minor problems found with some</p>	<ul style="list-style-type: none"> • GROS will consider how best to deliver consistent coding indexes to the service provider. • better automation within the IT solution will be provided to speed up and increase accuracy of the key issue of coding Industry & Occupation; • Implement changes proposed in the quality assessment report. 	

Description	Success Criteria	Outcome against success criteria	Recommendation	Timeframe
	<p>checks conducted.</p> <p>d) Reference materials including coding materials updated as per specifications and procedures.</p> <p>e) Management information reports produced as required and requested.</p>	<p>variables during load and validation stage and initial quality checks. Quality assessment carried out using the independent quality assurance system confirmed that further work would be required to:</p> <ul style="list-style-type: none"> • improve service levels; • improve collection, recapture/recoding and reporting of the quality assurance process; and • identify the causes of service level failures. <p>d) Achieved.</p> <p>e) Partially Achieved Both parties agreed that report development and production would be rescheduled due to the timetable issues outlined in the specification section</p>	<ul style="list-style-type: none"> • GROS and the service provider to consider ways of reducing the number of cases referred to GROS for resolution during data capture and coding. • GROS to review report scope and content to ensure that they provide a complete, consistent and accurate summary of services and outputs delivered. 	

Description	Success Criteria	Outcome against success criteria	Recommendation	Timeframe
		<p>above. Although reports were delivered as requested anomalies and inconsistencies were found that could not be fully investigated during the rehearsal.</p>		
<p>Output – Delivery of Service</p>	<p>Images, microfilm and files are produced in line with agreed quality standards and processing schedule timetable.</p>	<p>Achieved</p> <p>a) Although this has ultimately been achieved, there have been a few issues with the quality of output leading to the necessity of re-output of data to ensure the quality standards were met.</p> <p>b) Data transfer was secure at all stages.</p> <p>c) Once data output was received by GROS, an ad hoc process was used to classify the severity of any failures in the output file. Whilst this worked in the rehearsal, a more formal process needs to be developed for 2011.</p>	<p>a) The solutions noted above under “Capture & Coding” relating to a more formal and better resourced approach to testing, project management and change control will also be followed for output.</p> <p>b) Further work will be carried out to define a more formal assurance process once data output is received by GROS.</p> <p>c) Both parties will give consideration as to how the transfer of output data could be streamlined due to the large volumes of data.</p>	

Description	Success Criteria	Outcome against success criteria	Recommendation	Timeframe
		<p>d) Whilst the physical delivery process of data worked well, it was cumbersome and resource intensive.</p> <p>e) For the rehearsal a sample of 2000 census images and associated metadata were recorded onto two rolls of microfilm. The imaging service provider prepared the images for input, produced the two reels of film (positive and negative) and processed the films.</p> <p>There were a few problems encountered at the imaging service providers:</p> <ul style="list-style-type: none"> • Colour pages were inserted by the service provider to indicate the start and the end of a questionnaire, the archive writer was not able to deal with colour. 	<p>d) The parties will consider a more sophisticated validation processes that can be carried out before the data is output to GROS.</p> <p>e) Further investigation on the film produced using the image reader will result in a recommendation on the use of positive or negative film for 2011.</p> <p>GROS will draft the quality checking criteria to be applied to the microfilm in 2011.</p> <p>GROS will draft at a high level the full processes required to meet the requirements for a microfilm archive.</p> <p>These documents will be discussed and agreed with the service provider and any necessary changes agreed to the final version for 2011.</p>	

Description	Success Criteria	Outcome against success criteria	Recommendation	Timeframe
		<ul style="list-style-type: none"> • No header and trailer pages had been produced by the service provider, so the rolls of film do not contain this information. • There was also a problem when the film was being recorded (an error kept occurring). This happened several times, although the second reel was produced with no problem. • There should have been hardware and software available to produce labels for the microfilm cartons. <p>There were no problems encountered with the processing at the imaging service providers although we still have to check the images.</p>	<p>Based on the above recommendations the service provider will produce documents defining how they will provide the microfilm service for 2011.</p>	

3. Other evaluation points

Description	Outcome/Issue	Recommendation	Timeframe
<p>1) Understanding of contract/requirements.</p>	<p>Initially, there was a lack of clarity, with both GROS and the service provider having different views on the services to be delivered and the requirements to be met. However, as the rehearsal progressed a clearer understanding was achieved.</p>	<p>See recommendations at capture & coding output specification above.</p> <p>Creation of a “Dependencies Workstream” will ensure that all requirement dependencies are fully understood by both parties.</p>	
<p>2) Impact of service specifications.</p>	<p>Need to understand the purpose and impact of each service specification before development starts.</p>	<p>The purpose and impact of each specification needs to be understood and agreed by both parties to allow these to be prioritised to ensure development and testing phases can meet the timescale available.</p>	
<p>3) Project Management.</p>	<p>During the rehearsal, project management was, at times, light on both sides (the service provider and GROS). This affected delivery of the project as some dates were allowed to drift and there was not always a strong feeling of working towards common goals.</p>	<p>Strong project management is important in a project of this nature and both sides need to provide dedicated and experienced project management.</p> <p>It has been recognised that the PDCC management structure used in the rehearsal was not suitable for a project of this size and complexity. A new structure has now been put in place with separate workstreams (e.g. System Development, Testing, Resources, Facilities, Dependencies) all managed by a lead from each organisation and reporting to a joint IPT.</p>	

Description	Outcome/Issue	Recommendation	Timeframe
4) Geography Team.	<p>a) There was an inefficient and cumbersome interface for query resolution between the GROS based Geography team and the on-site team.</p> <p>b) It was also clear that the Geography team had not been kept up to date with changes in the system (e.g. the list of offshore departure points available in the system).</p>	<p>a) GROS will consider the potential to streamline the interface with the Geography team.</p> <p>b) The Geography team will be integrated into the project team throughout the life of the project.</p>	
5) Quality Assurance.	The complexity of ensuring that the right assurance was provided, coupled with pressures on time and resources, meant that GROS was late in making the quality documentation available to the service provider.	A different approach is needed for 2011 and workshops need to be arranged to look at options and look at the impact any decisions would have on development and testing.	
6) PDCC Quality Management.	Pressures on time and resources during the rehearsal meant that it was not possible to specify quality in the way that GROS had expected.	Quality Management will be better defined across the solution and better managed.	
7) System Training issues.	<p>a) Due to the “just-in-time” nature of some of the development, formal planning and delivery of training proved challenging.</p> <p>b) Changes to systems and coding strategies also impacted on the understanding of those being trained.</p>	<p>a) Training will be given a higher profile in the run up to 2011 and more professional training standards applied.</p> <p>b) The service provider will develop and maintain a fully populated training schedule that identifies and fully resources each step in the training process between now and delivery in 2011.</p>	

Description	Outcome/Issue	Recommendation	Timeframe
	<p>c) The stop-start nature of the system delivery and testing process made it difficult for trainees to consolidate their knowledge.</p>	<p>c) Dependencies with other workstreams (e.g. signed off functional specification of applications) will be pro-actively managed by the service provider.</p>	
<p>8) Configuration Management.</p>	<p>There were occasions when version control of documents was given a low priority due to other project commitments. This occasionally caused problems during the rehearsal.</p>	<p>Both the service provider and GROS will jointly consider the benefits of a more formal Configuration Management approach.</p>	
<p>9) Management Information Reports.</p>	<p>Pressures on time and resources during the rehearsal meant that it was not possible to specify reports as early or as fully as GROS expected. Further work is needed to understand the scope of reports required for 2011 and to ensure that this work is completed much earlier than it was in the rehearsal. GROS to clarify, in a timely manner, their reporting requirements.</p>	<p>Workshops will be arranged to better understand the requirement and propose solutions.</p>	
<p>10) Security.</p>	<p>The GROS independent security advisor and the service provider's security supplier audits and reports showed that security on site was of a high standard and personal data, both in paper and electronic forms, was secure at all times. A few small issues were highlighted, all of which were actioned quickly. There are a small number of recommendations for 2011.</p>	<p>a) Site access control will to be reviewed once new site chosen.</p> <p>b) A separate security induction will be given as part of the staff induction process to emphasise its importance.</p> <p>c) A debrief on confidentiality obligations will be part of the exit procedures for staff leaving the team.</p>	

Description	Outcome/Issue	Recommendation	Timeframe
11) Warehouse - Materials from remote locations.	It was felt that getting materials from remote locations was a risk (e.g. late or cancelled ferries). The warehouse needs to be flexible and able to accommodate these occasions.	All parties (the service provider, GROS and the logistics service provider) will work to consider these scenarios and agree solutions.	
12) Warehouse - scheduling.	Any lack of information or accuracy regarding scheduled delivery impacts on the service provider's staffing levels and ability to plan workloads.	Workshops will be set up to examine these factors and develop a schedule for inbound deliveries agreed by all parties.	