



**Palestinian National Authority
Palestinian Central Bureau of Statistics**

**United Nations Statistics Division (UNSD)
Economic and Social Commission for Western Asia (ESCWA)**

**Workshop on Census Data Processing
Doha, Qatar 18-22/05/2008**

**The Palestinian Census
Executive Summary**

May, 2008

Table of Contents

Overview of Census-2007 Data Processing.....	3
Control of Workflow	3
Enumeration Areas Master Book	3
Selection of Identification Fields	4
Data entry:	4
Data Processing Estimates:	4
Data Processing Operation Room (Dpor):.....	6
Quality Assurance	6
Daily reporting	6
Registration Process:	7
Edits:	7
Form design and testing.....	8
Software acquisition and evaluation	8
Lessons learned from 2007 census data processing.....	9

Overview of Census-2007 Data Processing

The 2007 Census data processing included all activities that come after census fieldwork, which include check-in stage, office editing, office coding, data capture, data editing, and tabulation. Also, census data processing was involved other census-related activities such as the development and follow up of census personnel system as well as other administrative supporting systems.

In 2007 Palestinian census, the technology infrastructure needed to process a census was part of the statistical office overall technology infrastructure. This way, the technology was tested continuously since the statistical office relies on it heavily on its day to day processing of surveys and others.

In the 2007 census, the statistical office technology infrastructure was used with little upgrade that would enhance performance of the main server. High-speed computers were procured to be used for census computer editing and tabulations.

The census DP activities were carried out using two shifts a day to ensure results are disseminated in timely manner.

Control of Workflow

Prior to the check-in stage, special committees were formed for the activity of the Receipt and Registration of census questionnaires and other materials. These committees were in charge of receiving the deliveries according to a well-explained procedure. Batches received were flagged out by date into the tracking system as well as into the Master Book and then stored in the storage room according to well-ordering system. The tracking system was essential to efficiently monitor movements of batches between the different data processing activities. The tracking system reflected the concurrent flow of the processes. The tracking system developed for 2007 census was based on the queue concept: Batches that completed one activity would be in line waiting for processing in the next activity.

Controlling workflow was given high attention and was daily evaluated by the DP operation room. Time estimation and project plan should take into account different start date for each activity. A time lag should be set between the dependent activities to ensure workflow especially at the beginning of the process. Daily evaluation of progress plan might uncover possible delays that require the movement of staff accordingly. At early stages, movements of DP staff in editing and coding was required. However, at later stage, such movements between these two stages should be very seldom. Later in data processing, movement of DP staff from editing and coding to computer editing should be expected. Therefore, supervisors must be aware that their best DP staff would eventually be moved to other activities as required, so other staff should be ready to take over.

Enumeration Areas Master Book

A list of Enumeration Areas codes was prepared after the completion of updating all maps for all localities in the Palestinian Territory. These codes were then put in a Master Book to be used in all census stages including the data processing stage. The purpose of the Master Book is to verify the coverage of all Enumeration Areas especially in the Checkout (Sending questionnaires to the fields) and Check-in (delivery of questionnaires to Head office). After the population count and during the check-in stage, all enumeration areas batches were checked against the Master book to ensure full coverage and full delivery to the head office.

After the completion of the check-in stage, the Master Book was placed in the questionnaire storage room. The Master File was also used to control data entry activities as well as questionnaire movements among data processing different stages.

Selection of Identification Fields

Selection of appropriate identification fields to control data on batch as well as questionnaire levels is of utmost importance. Subject matter as well as data processing personnel had participated in the selection of these identification fields during development of questionnaires. Frequent testing of the efficiency of these fields proved to be very valuable in eliminating any chance for data duplication on batch or questionnaire levels. Identification fields on batch level are locality code, Enumeration Area code, total number of questionnaire booklets as well as total number of households (questionnaires) in the enumeration Area.

Booklet number, building number, housing unit number, total households in each enumeration area, household sequence number in the Enumeration Area, and total number of family members were used as identification fields on questionnaire level. Household sequence number was one of the main identification keys used to control more efficiently data on household level.

Data entry:

The system used to capture data from the census forms was developed using the US Bureau of the Census CSPRO. PCBS adopted the centralized approach by installing data entry system on the Network, and all data keyers will be using the same system. Data, which keyed in by data keyers, will automatically be placed in special protected directory on the network server.

Data Processing Estimates:

It is necessary to calculate the time needed to carry out each of the census data processing activities. Since activities in census data processing are interdependent, calculation of time estimates had to be accurate if not precise. Failing to correctly producing time estimates would result in delay in the whole plan as well as in extra burden on the staff implementing these activities.

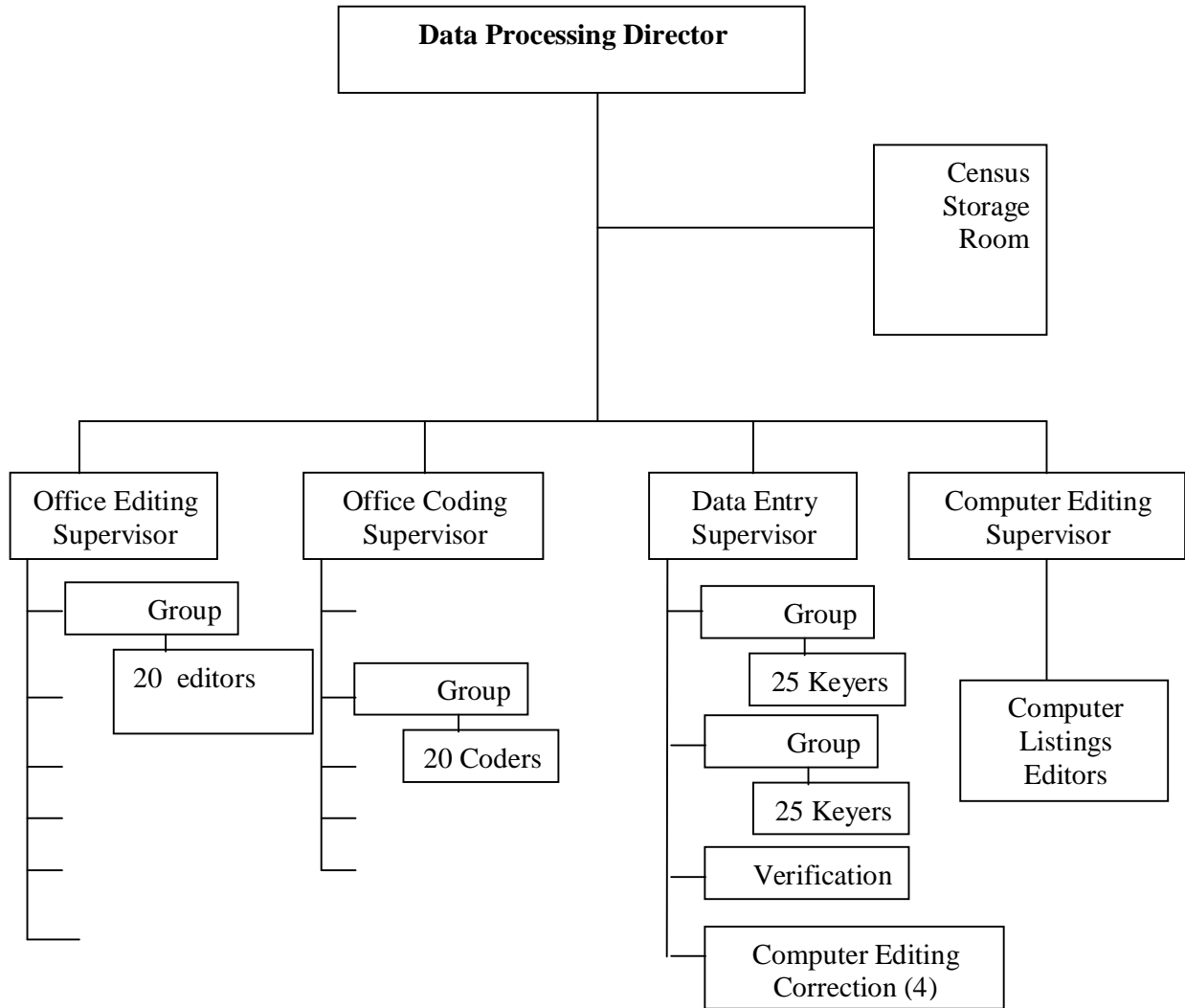
Census activities of office editing, office coding, data entry, and computer editing have to be implemented in parallel with some time lag in between at the beginning of the process. This time delay must be defined and accurately set since all following activities will function accordingly.

This notion of parallel processes with some time lag in between could be looked at as a multiple queuing system where activities that finished processing from one stage will be waiting for in-queue for the next stage. It is worth mentioning that Census Tracking System was based on this notion.

The first stage in census data processing was the Office Editing Stage. Other processes would not start until the time lag calculated during the preparing and planning stage elapses. After the elapse of the time lag, there should be enough work waiting in the "B" queue (See below) for office re-editing. If the time lag elapses with not enough workload available in the "B" queue, all subsequent activities will be delayed and as a result the time schedule of the whole project will be violated.

The learning experience was to monitor progress of all activities constantly and be able to transfer data processing staff from when activity to another as required to keep activities within the time frame. However, such actions should not become frequent in the data processing stage since it requires great amount of effort and attention from the supervision staff.

To ensure accurate time estimates, calculations were based on real data. In 2007 census, the pilot census was conducted before the real census. Statistics were collected during this pilot about the processing rates of each activity in data processing such as editor average rate, coder average rate, data entry average rate and data cleaning average rate. These statistics were one of the most important findings from data processing pointy view.



Data Processing Operation Room (Dpor):

The Data Processing Operation Room (Dpor) was established once data processing activities started. The operation room was headed by the census data processing director and included as members census key subject matter staff, data entry supervisor, census storage keeper, and representative for the administration directorate. The Dpor used to meet daily in the morning to evaluate project progress in all activities, discuss technical problems and provide solutions, advise on moving staff from one activity to another according to workload and priorities, and insuring availability of transportation and other necessary means for the three shifts. The Dpor used to report daily on its activities to the Census National Director through Census Executive Director.

Quality Assurance

In the 2007 Population, Housing and Establishment Census,

1. Office editing was implemented 100%
2. Office editing verification was implemented 100%
3. Office coding was implemented 100%
4. Office coding verification was implemented 100%
5. Data entry verification was implemented 100% at early stages and for new data entry staff, then only %5 random sample was verified
6. DP staff in charge of verification were separated from other staff, and they were the best in the DP pool.
7. Such emphasis in DP activities paid off positively in data cleaning and tabulation.
8. Such mechanism enabled the statistical Bureau to produce final results in record time.
9. DP staff in charge of verification use different colored pens from other DP staff.

Daily reporting**Diagram 1: Report of activity status for each governorate**

Governorate	Total E.As	Office Editing		Office Coding	
		Processed	Accum.	Processed	Accum.
Jenin	298	298	298	298	298
Tubas	58	29	29	12	12
Tulkarem	184				
.					
.					
.					
Rafah	115				
Total	3312	327	327	310	308

- Storage keeper prepared the above report daily.
- This report was daily studied in the data processing operation room.
- Staff moved form one operation to another as work load accumulates

Diagram 2: Report of activity status for each governorate

Governorate	Total E.As	Data Entry		Computer Editing	
		Processed	Accum.	Processed	Accum.
Jenin	298	20	100	10	25
.					
.					
.					
Rafah	115				
Total	3312	327	327	310	308

- This report was prepared daily by data entry supervisor
- This report deals with data entry activity and computer editing.
- This report studied daily by DPOR.
- Batches completed the computer editing stage carried the signature of the DP director to be accepted by the storage keeper for final storage.
- A Completed editing computer program would be run on each batch before stamping final storage signature.

Registration Process:

E.A Master Book as well as the tracking system ensured that every E.A was received at the processing center, this was also ensured also by the strict procedures adopted by the check-in committees.

Edits:

Office editing instructions were transformed into computer editing program with within-record check as well as cross-record checks. Continuous quality improvement could only be reached by involving key subject matter staff into the DP process. Once all enumeration areas of a governorate passes the computer editing stage, special data files that cover all census subjects such as education, fertility, labor, migration, disability, and living conditions are extracted and given to subject matter staff who are special in these fields for validation. Comments are encouraged and once received they would be discussed, and once approved they would be incorporated into the computer-editing program as well as into the office-editing manual. Also, instructions will be given to DP staff to check for such cases. This process is continuous and new checks sometimes come up that were not thought of in the early stages. This process enriched the validation process taking into consideration that this was the first census to be conducted by the statistical Bureau. The quality assurance was applied through incorporating key subject matter staff into the overall DP process.

The office and computer editing manual was prepared and tested prior to the main census. For each variable, the accepted values were defined. The relation of this variable to other variables in the same record or to different record in same household were clearly described. Each test or check is given a number that branches form the variable number on the questionnaire.

The computer-editing program produces inconsistencies that carry the same check number as it is in the manual. This approach assisted editors in finding related checks in the manual easily. One computer editing program is used and applied on all data Batches. Before a batch is sent for final storage, the Batch will go through the editing program and if there are no inconsistencies, the batch will be stamped

and then sent to final storage. The decision to use this type of data capture was largely enforced by the size of the census questionnaires. The questionnaire used for the 2007 Palestinian Census was very large compared with other questionnaires in censuses.

Coding is critical step and must be given special attention. Clerical coding requires good training and close supervision. An expert in this field conducted training on coding. Furthermore, this expert worked in the coding hall with Coders all the time. Verification of office coding was implied 100% . In addition, computer-editing manual included checks on coded variables with other variables in the questionnaire such as Occupation and Highest Education, Occupation and Sex, Highest Education and Age, etc.

Form design and testing

In the 2007 census, form design and testing stage implemented with,

- Full cooperation between subject matter staff and DP staff.
- Controls were embedded into the questionnaires to ensure the quality of the data in all of census stages.
- During this stage, several issues were addressed such as handling of questionnaires during fieldwork and data processing especially during data entry stage.

Software acquisition and evaluation

Software had to be extensively tested in census environments. CSPRO was used in the processing of the 2007 census. The software was evaluated against the following criteria before deciding on using it:

1. Designed mainly for DP of censuses, and could be used in other subjects.
2. Frequently tested and used in censuses.
3. Continuously upgraded by responsible agency.
4. Availability of technical support.
5. Contains utilities for data entry, computer editing, forms tracking, and tabulation.
6. Utilities to embed check and controls into the DP systems.
7. Availability of documentation about the Software.
8. Data portability.
9. User friendly.

Lessons learned from 2007 census data processing

1. Calculations of time and labor estimates must be based on statistics compiled from census experiments such as census pretest or pilot census.
2. Factors that might affect production of various processing activities should be taken into consideration in time and labor estimate.
3. Differentiate between rates for data entry of numeric data and alphanumeric data in calculating time and labor estimates.
4. Monitor progress of activities constantly and be flexible enough to switch data processing staff from one activity to another according to workload distribution.
5. Train all staff working in editing and coding on these two activities to be able to switch between them, as workload requires.
6. During the census implementation stage, PCBS heavily emphasized on detailed documentation of all census activities. One central documentation center was established at the beginning of census field activities along with field documentation center in each governorate. About 30 non-statistical reports about all census activities such as fieldwork, data processing, administration, publicity, census various committees and others had been prepared.