

**PROJECT ASSISTANCE COMPLETION REPORT**

**CENSUS DATA PROCESSING**

**OAJC II Project, 272-0105**

**Selma L. Sawaya  
U.S. Bureau of the Census**

**July 26, 1995**

# TABLE OF CONTENTS

EXECUTIVE SUMMARY .....	1
BACKGROUND .....	2
STATUS AT PACD .....	3
<b>I. INTRODUCTION .....</b>	<b>4</b>
<b>II. PLANNING AND PREPARATORY ACTIVITIES .....</b>	<b>4</b>
<b>A. Prior to the arrival of the Data Processing Advisor .....</b>	<b>4</b>
<b>B. Post-February 1993 .....</b>	<b>5</b>
1. MARCH-JUNE 1993 .....	5
2. JULY-SEPT 1993 .....	5
3. OCT-NOV 1993 .....	5
4. POST-ENUMERATION PLANNING FOR LATER PHASES .....	5
<b>III. SYSTEM OVERVIEW .....</b>	<b>6</b>
<b>A. Resources .....</b>	<b>6</b>
1. HARDWARE AND SOFTWARE RESOURCES .....	6
2. HUMAN RESOURCES .....	6
<b>B. System Elements .....</b>	<b>8</b>
1. GEOGRAPHIC DATA BASE .....	8
2. ENUMERATION CONTROL TOTALS .....	8
3. ESTABLISHMENT CENSUS .....	9
4. SUMMARY FORMS .....	9
5. HOUSING AND POPULATION CENSUS .....	9
<b>C. Processing Phases .....</b>	<b>10</b>
1. GEOGRAPHIC DATA BASE .....	10
2. PRE-ENUMERATION CONTROL COUNTS .....	10
3. DATA ENTRY STAFF SELECTION .....	10
4. SUMMARY FORMS (I) KEYING AND TABULATION .....	11
5. SELECTION OF 10% SAMPLE EAS .....	11
6. ESTABLISHMENT CENSUS DATA PROCESSING .....	11
7. SUMMARY FORMS (II) KEYING .....	11
8. POPULATION DATA (10%) PROCESSING .....	11
9. POPULATION DATA (90%) KEYING .....	11
10. DATA EDITING (MANUAL) .....	11
11. DATA CONSOLIDATION .....	11
12. DATA EDITING (AUTOMATIC) .....	12
13. ESTABLISHMENT CENSUS NAMES AND INDUSTRY GROUP CODES .....	12
14. HPC DATA FILE ACCEPTED BY USERS .....	12
15. TABULATION OF 100% .....	12
<b>IV. DETAILED PROCEDURES .....</b>	<b>13</b>
<b>A. Data entry .....</b>	<b>13</b>
1. STAFF SELECTION .....	13
2. SHIFTS .....	15
3. KEYING ROOM SET-UP .....	15
4. STANDARD PROCEDURES DURING PRODUCTION KEYING .....	15
<b>B. Operational Control .....</b>	<b>16</b>
1. GEOGRAPHIC DATA BASE .....	16
2. CENSUS BOOK CONTAINER STORAGE (BOOK CONTROL DATA BASE) .....	18
<b>C. Pre-enumeration control counts .....</b>	<b>18</b>
1. SYSTEM DEVELOPMENT .....	18

2.	FILE STRUCTURE	19
3.	ENTRY OF INFORMATION	19
4.	TABULATION	19
<b>D.</b>	<b>Summary Forms</b>	19
1.	SYSTEM DEVELOPMENT	20
2.	FILE STRUCTURE	20
3.	ENTRY OF FORMS (I)	20
4.	RE-CREATION AND ENTRY OF SUMMARY FORMS (II)	21
5.	USE OF SUMMARY FORMS IN LATER PROCESSING	23
<b>E.</b>	<b>Establishment Census data</b>	23
1.	SYSTEM DEVELOPMENT	23
2.	FILE STRUCTURE AND CONTENT	24
3.	PRE-ENTRY MANUAL PROCESSING	24
4.	DATA ENTRY AND VERIFICATION	25
5.	OPERATIONAL CONTROL	25
6.	VALIDATION OF DATA (MANUAL)	25
7.	FINAL PRE-TABULATION COMPUTER VALIDATION	25
8.	TABULATION	26
9.	SUBFILE EXTRACTION OF ACTIVE ESTABLISHMENTS	26
10.	ADDITION OF NAMES TO SUBFILE OF ACTIVE ESTABLISHMENTS	26
11.	MERGE OF MODIFIED SUBFILES	27
12.	CREATING SECTOR FIELD FOR ACTIVE ESTABLISHMENTS	27
<b>F.</b>	<b>Housing and Population Census data</b>	27
1.	SYSTEM DEVELOPMENT	27
2.	HPC DATA FILE STRUCTURE	29
3.	SAMPLE SELECTION AND WEIGHT CALCULATION	30
4.	PRE-ENTRY MANUAL PROCESSING	30
5.	ENTRY OF CENSUS BOOKS	32
6.	OPERATIONAL CONTROL	33
7.	VALIDATION OF DATA (MANUAL)	33
8.	CONSOLIDATION TO HIGHER-LEVEL DATA FILES	33
9.	FINAL PRE-TABULATION COMPUTER VALIDATION	33
10.	USER ACCEPTANCE OF DATA	33
11.	TABULATION	34
12.	UNSCHEDULED REQUESTS FOR DATA AND TABULATIONS	34
<b>V.</b>	<b>CONSIDERATIONS FOR FUTURE CENSUSES</b>	35
<b>A.</b>	<b>Planning for processing</b>	35
1.	PROCESSING OF DATA COLLECTED DURING ENUMERATION	35
2.	SYSTEMS FOR MANAGEMENT CONTROL	35
3.	STAFFING	36
4.	DATA ENTRY ENVIRONMENT	37
<b>B.</b>	<b>Processing</b>	37
1.	CHECK-IN OF CENSUS BOOKS AND VALIDATION OF GEOGRAPHIC CODES	37
2.	SAMPLE PROCESSING	38
3.	MANUAL EDITING	38
4.	CODING	39
5.	DATA ENTRY	40
6.	DATA EDITING AND CORRECTION	41
7.	DATA TABULATION	41
8.	OPERATIONAL PROCEDURES	42
9.	TIME SCHEDULE	43
10.	POST-ENUMERATION SURVEY (PES)	43
<b>VI.</b>	<b>APPENDICES</b>	45
<b>A.</b>	<b>Data bases</b>	45
1.	OCCUPATION	45
2.	INDUSTRY [ECONOMIC ACTIVITY]	45

3. <b>GEOGRAPHIC</b> .....	45
4. <b>BOOK CONTROL</b> .....	46
<b>B. Abbreviations and Definitions</b> .....	46
<b>C. IMPS (Integrated Microcomputer Processing System)</b> .....	47
<b>CONTACT LIST</b> .....	49

## EXECUTIVE SUMMARY

In January 1993, the U.S. Bureau of the Census signed a PASA with the Omani-American Joint Commission (USAID/Muscat), under the terms of which the Census Bureau would provide to the Development Council (later Ministry of Development) of the Sultanate of Oman a long-term advisor in Census Data Processing. The advisor would be assigned to the Census Project at the Ministry, and would work with the Director of the Census, Ali Mahboob Hassan Al-Raisi, in ensuring the timely and accurate processing of data collected during the Census field enumeration. This assistance would extend through publication of final results, and would encompass various training and documentation activities, as well.

Under the terms of the PASA, the OAJC would also underwrite the cost of microcomputers and related equipment, in quantities sufficient to ensure the timely processing of the data. Estimates of the numbers needed of each type of equipment were based on information given by Census Project staff to the Census Bureau advisor who had made the initial assessment visit in the fall of 1992.

The Advisor [DPA] arrived in the third week of February 1993, and continued work with the the Census Project staff through 26 July 1995 (the original two-year PASA was extended once). The months of March–November 1993 were dedicated to planning for the post-enumeration processing activities, to receiving and installing the procured hardware and software, and to training the staff (Omani and expatriate) assigned to the Census Data Processing Unit [CDPU].

The Census enumeration was conducted during the period 1–10 December 1993. Beginning 11 December, processing began, with the first output (preliminary results based on Summary Forms) produced by 17 December. These were the figures which were announced to the public before the end of that month (December 1993), and on which the allocation of seats in the Majlis A`Shura was based.

Basic processing activities continued uninterrupted from the end of enumeration. With the assistance of the DPA, the CDPU was able to process, in a timely and accurate fashion, data from the Census of Establishments; the Summary Forms (which had provided the preliminary results) a second time (to correct locality code errors); a ten percent sample of Housing and Population Census data from the entire Sultanate (used internally); and the full set of data from the Housing and Population Census, which formed the basis of the results announced to the public beginning in late 1994. In addition, a number of users, both inside and outside the Ministry, had received subsets of the data for research or analytic purposes.

At the time of the Advisor's departure, the Ministry was in the process of transferring the Census data sets to the Ministry's main computer system, where they were expected to function as public-access data bases (with some restrictions). The Advisor completed the Data Processing Procedural History before her departure.

## BACKGROUND

In the Sultanate of Oman, the modern era is considered to have begun in 1970, when the current ruler, Sultan Qaboos bin Said, ascended to the throne. In keeping with the Sultan's desires for the development of his people, he initiated a series of five-year plans to guide the country in the most appropriate use of its resources—human, natural, and financial. From time to time in the period following the Sultan's accession, surveys were carried out in selected areas to provide basic data on housing, health, education, etc., but in the entire history of the Sultanate there had never been a Census to use as a "benchmark" for comparing survey data, or even to provide a proper sampling frame. In addition, even though a statistical unit existed in the Development Council, many of these surveys were conducted under the auspices of other Ministries, with the expected differences in methodology, attention to detail, and precision of results.

Thus, in June 1991 the Sultan signed a Royal Decree which authorized a census of housing, population, and establishments, to be conducted as soon as possible. The Development Council was given overall executive authority for the operation, and a supervisory committee [Census Technical Committee/CTC] was organized. The December 1993 period was decided upon as most appropriate for enumeration activities. In 1992, Ali Mahbeob Hassan was named to direct the Census Project, and he began the task of assembling staff and advisors to complete the planning and design phase.

When the Census Bureau DPA arrived in February 1993, the questionnaire was still not in final form, nor had the publication tables been designed. (It is preferable to draw up publication table layouts before defining the questionnaire, so that questionnaire content may be precisely defined and question categories delimited. However, almost invariably Census planners will design a questionnaire and then consider uses for the data collected. This was the path followed in Oman, and it led, as expected, to difficulties in the editing and tabulation phases.)

Over the life of the Census Project, the Director was assisted by (in addition to the DPA) a Principal Census Advisor (Nader Al-Hallak, from Syria), a Cartography Advisor (Dr. P. Nag, from India), and various short-term advisors from a number of Arabic-speaking countries, who assisted Omani staff in carrying out activities in the field and in the office.

## STATUS AT PACD

The objective of USG assistance to the 1993 Census of the Sultanate of Oman was to ensure that the agency responsible for the Census (Development Council/Ministry of Development) had sufficient resources (including technical expertise) to permit the processing of Census data in a timely and accurate manner. To this end, USG assistance was channeled into the procurement of microcomputers and related processing equipment, and the provision of a long-term technical advisor to assist local Census Project staff with the programs and procedures necessary to produce usable results.

USG assistance to the 1993 Census Project began with the arrival of the Advisor in Muscat in February, 1993 and ended with the departure of the Advisor from Muscat in July, 1995. The microcomputer-related equipment was procured in two phases: a preliminary shipment of equipment needed by the Census data processing staff (including the Advisor) to develop and test system software, and a final shipment, just before the Census date, of equipment needed for data entry and validation. Under the assumptions upon which the estimates of equipment were based, the quantities provided would have been sufficient to carry out the work; however, shortly after the Advisor's arrival, Census Project management modified the work plan and the estimates of volume, and the original quantities of equipment became insufficient. By dint of judicious procurement, the Census Bureau was able to slightly increase the total numbers of microcomputers shipped, and the Ministry itself purchased the remaining equipment needed.

Once the data entry operation was complete, some of the microcomputers used in that phase of the work were set aside to use as backups for data storage, but most were recycled to other offices in the Ministry. The more powerful microcomputers used for system development and processing continue to be used, along with the printers and the Bernoulli storage devices, for generating publication tabulations. Ministry staff estimate that this process will be completed by 31 August of this year, at which time the Census Data Processing Unit will be formally disbanded and the remaining equipment distributed to the Computer Department and other Ministry offices.

In terms of target dates, the Census Project performed extremely well. Data entry for both the Establishment Census and the Housing and Population Census was completed at least two months ahead of the earliest planned date. Results based on a(n approximately) ten percent sample of the data were available within seven months of the enumeration period (for internal use only), and the final results, by sex and Omani/non-Omani breakdowns to the level of locality, were available within one year of the enumeration period. Detailed tabulations are still being generated, but it is clear that—barring some extremely unlikely unforeseen occurrence—they will be completed well within the two-year time limit generally accepted for publication of census results; they will, in fact, probably be ready for formal publication by late August, 1995. At that time the Census data files will be transferred to the MoD's minicomputer, to serve as a data base for approved research and analysis by users both within and outside the MoD.

The detailed report which follows is an adaptation of the Procedural History prepared for the Ministry of Development. A number of attachments and supporting documents of a more technical nature have been eliminated, but most of the narrative history has been retained for purposes of completeness. Points to be considered for improvement during the next census have been included in Section V. At the end of Section VI will be found a list of those persons, involved in the 1993 Census activity at the Ministry, with whom the Advisor had most contact.

1993 Census of Population, Housing, and Establishments  
of the Sultanate of Oman

SYSTEM DOCUMENTATION

**I. INTRODUCTION**

This document is intended to be the record of all activity related to processing the data collected from the 1993 Census of Population, Housing, and Establishments, and all auxiliary computer-related activities. It has been organized so that the reader who wishes only a cursory overview need read only the general description of the system(s) in which s/he is interested, while those who wish to understand the systems at the most detailed level can also find the information they seek.

- NOTE:
- (1) In this document, the use of the word "region" is understood to include the administrative division known as "governorate," as well. Any exceptions will be noted in the specific instance.
  - (2) A list of abbreviations and definitions of terms used may be found in Section VI.B. below.
  - (3) The Royal Decree which ordered a national census charged the organization known as the "Development Council" with the conduct of the operation. In January 1994, after completion of the Census and publication of preliminary results, the Ministry of Development [MoD] was created and assumed responsibility for all 1993 Census-related activities. In this document, references to the "Ministry" or "MoD" refer essentially to the same organization, as it was constituted prior to and after January 1994, respectively.
  - (4) Throughout this document, the reader will find many references to Census field staff. To better understand these references, a brief explanation of the Census field organization may be necessary. For the purposes of enumeration of establishments, housing units, households, and individuals, the Census Project divided the eight regions and governorates in the Sultanate of Oman into fifteen administrative areas, each of which was headed by a Supervisor, who was responsible for all field operations in his/her area. (For the record, no Supervisor Area covered more than one region or governorate, and some regions/governorates were divided into two or more Supervisor Areas.) Each of these fifteen Supervisor areas was subdivided into a number of smaller areas, each of which was headed by an Assistant Supervisor [AS], who was responsible for all field operations in his/her sub-area. The Assistant Supervisor's area was further divided, and each of the resulting sub-areas was headed by a Crew Leader [CL], who was responsible for the enumeration of establishments and the pre-enumeration listing of housing units and households, as well as for all work carried out by the Enumerators under his control. The lowest rung in the field hierarchy was the Enumerator, who was responsible for the detailed enumeration of housing units, households, and population in the Enumeration Area [EA] to which s/he was assigned.

**II. PLANNING AND PREPARATORY ACTIVITIES**

The work performed under the umbrella of Data Processing encompasses a number of procedures, both manual and computerized. Some of the activities were envisioned in the original planning for the Census, but many activities were developed "on the fly," as the need for them arose. This planning and general preparatory process is described in the following paragraphs.

**A. Prior to the arrival of the Data Processing Advisor**

In August 1992, in response to a request from the Ministry of Development [MoD] to the Omani-American Joint Commission [OAJC], the U.S. Bureau of the Census [BuCen] sent a senior data processor to Muscat to review Census planning activities and to make recommendations for processing the data to be collected in the Census. The consultant, Kathy Chamberlain, spent two weeks in Muscat interviewing and working with MoD staff and the Census advisors before preparing a comprehensive report on the status of planning for the Census at that time. The report contained a number of recommendations to the Director of the Census Project, the Census advisors, and the OAJC. Many of the recommendations were ultimately implemented, including the nomination of a long-term Data Processing Advisor [DPA] and the acquisition of microcomputer equipment for processing Census data.

## **B. Post-February 1993**

The DPA arrived in Muscat at the beginning of Ramadhan (22 Feb 93). She began to familiarize herself with procedures at the MoD (the umbrella organization for the Census), to become acquainted with the personnel attached to the Census Project, and to study the reports and documents (including Census forms) produced to that point.

### **1. MARCH-JUNE 1993**

During this period, the major concern of the DPA was to consider the processability of the proposed instruments (that is, the forms for collecting information during the enumeration periods) and to advise on the probable effects on data processing of decisions taken by subject-matter staff. The main questionnaire for the Census was still highly fluid during this period, although by the end of the quarter it approximated the form it would ultimately acquire. In addition, the DPA began to develop the outline of the basic processing system, although details could not be specified until user needs became more clearly defined. The first tentative timetable was drafted, based on estimates of volume (including the proposed tabulation program) in use at that time.

It was also during this period that the first shipment of hardware arrived from the U.S., the Census Data Processing Unit [CDPU] was formed, and the first staff were assigned to the project (see below). The DPA began to define other hardware and software needs (e.g., printers and Arabic-language interface software), so that local procurement of these items might be initiated.

### **2. JULY-SEPT 1993**

During this period the CDPU began creating the geographic data base. Specifications were drawn up for the data entry staff to be hired, and the tender was issued. The staff was introduced to the software to be used for processing, and they began work on procedures to be used for entry of Housing and Population Census [HPC] data. The preliminary version of the timetable went through another iteration.

### **3. OCT-NOV 1993**

At the request of the Census Project, the U.N. Advisor in Sampling Methodology, in Muscat on consultation with another unit of the MoD, drafted a document with his recommendations on the procedures to be used in drawing the sample for advanced tabulations.

For entry of Census data, management decided to use local keyers, in addition to the contract keyers, and the CDPU began planning for Establishment Census [EC] processing.

Management also requested another system for processing information from the pre-enumeration field work. Since the request was made after staff had begun the field work, it cannot be said that the CDPU **planned** for this system. The programs that were developed, in haste, were not part of the original workload or timetable.

### **4. POST-ENUMERATION PLANNING FOR LATER PHASES**

Following the period of enumeration in the field, there were a number of modifications and additions to the previously-established schedule (similar to the example mentioned in the preceding paragraph). The effect of these modifications on data processing resources and the overall timetable depended, to a large extent, on the degree of advance notice given before the modifications had to be implemented. By their nature, these modifications were not planned for, so they cannot properly be discussed under the rubric of "Planning." They will be discussed in the remaining sections, in approximate chronological order.

### III. SYSTEM OVERVIEW

This section is designed to provide a general description of the resources available to the CDPU, and the multiple systems and subsystems that were employed during the various phases of the Census operation. Detailed specifications will be found in later sections.

#### A. Resources

The systems developed were dependent on the resources (human and other) available to the CDPU. These resources are discussed in the following paragraphs.

##### 1. HARDWARE AND SOFTWARE RESOURCES

Following the recommendations in Ms. Chamberlain's report, Census Project management made two decisions of critical importance: first, to establish a Census data processing unit separate from the Computer Department of the MoD; and second, to use microcomputers for all Census-related processing.

These two decisions helped ensure the ultimate success of the Census data processing effort as follows:

- first, by defining Census data processing as a special task, separate from the day-to-day activities of the parent organization, management freed the CDPU to carry out its task (a) without restraints imposed by persons not familiar with the specific characteristics of such activity, and (b) without having to contend for resources needed by other data processing systems; and
- second, by simplifying the hardware option, Census Project management ensured that the complications of maintaining hardware were kept to a minimum. This does not mean that there were no hardware problems, because there were; but it does mean that a more complicated or sophisticated hardware arrangement (e.g., a network) would likely have presented even more problems over the period of operation, because of user unfamiliarity with the hardware and software involved. [*A list of hardware resources can be found as Attachment III.A.1, below.*]

Once the choice of hardware (stand-alone, DOS-based microcomputers) was made, the choice of software was easy. For the operating system, the basic DOS, Release 5, was chosen. Microcomputers that would be adequate for the data entry operation under DOS would be totally inadequate under Windows, principally because Windows requires much greater system resources (disk space, mouse, RAM, enhanced video, CPU speed, etc.) to perform at even the minimum level than does DOS. In addition, a graphic operating system, such as Windows, would bring no advantage to a batch-based operation such as production data entry.

For the processing software, the choice was equally clear. Of all possible off-the-shelf software packages with appropriate facilities for statistical data processing, from data entry through tabulation, the one which combines greatest ease of use with maximum capabilities is the Integrated Microcomputer Processing System [IMPS]. This software was developed by the U.S. Bureau of the Census [BuCen] specifically for processing of censuses and surveys, and contains modules for all functions required for most statistical operations: data definition, data entry and verification, data editing, data tabulation, table retrieval, calculation of variance, and operational control. Non-technical staff can be trained to use many of the modules, and the most important elements (data entry, editing, tabulation) permits customization to a very high degree. The software was made available, at no cost to the Census Project, by BuCen through the OAJC. [*A brief description of IMPS and its component modules can be found in Section VI.C, below.*]

##### 2. HUMAN RESOURCES

Staffing needs were defined by phase, and varied greatly over the life of the project. They can be understood by examining the categories of staff required.

###### a. Programming staff

During the first few months, the DPA worked with one programmer-analyst, while awaiting management decisions on allocation of other staff from the Computer Department. In June, a program-

mer was detailed from the Ministry of Health [MoH], initially for a period of six months. (MoH subsequently agreed to permit the programmer to remain with the Census Project for a full year.) In mid-summer 1993, management assigned three programmers and one systems analyst (from other areas of the MoD) to the Census Project. Staffing continued at this level through the first quarter of 1994. Between May and August 1994, the staff was gradually reduced to the minimum level (one systems analyst, one programmer-analyst, and two programmers), where it remained for the remainder of the DPA's assignment in Oman.

From the beginning of his assignment to the project, the systems analyst functioned as the counterpart to the DPA, taking part in all discussions with Census Project management and participating in all decisions related to data processing activities. In the summer of 1994, he became the principal responsible for all coordination between subject-matter staff and the CDPU, and the DPA was able to concentrate on system development and staff training tasks, participating in meetings and discussions only when necessary. Over the course of the project, the counterpart also acquired a number of supervisory functions with respect to the other CDPU staff.

b. Data entry staff

The single most labor-intensive task in processing large volumes of data (such as a Census) is the transfer of manually-recorded information to magnetic media. During her 1992 visit, Ms. Chamberlain prepared initial estimates of time needed for **only** this operation (commonly known as "data entry"). She based her estimates on the content of the questionnaire, as it existed at that time, and on the estimates [of population, households, and housing units to be enumerated] then being used for planning purposes. It should be noted that she was not made aware of the need to take into account the Establishment Census operation, so none of the estimates of time or manpower in her recommendations reflected this additional workload, nor could they reflect the delays later occasioned by the user not meeting deadlines. With the information at her disposal at the time, she arrived at an estimate in which data entry would be completed within approximately six months, based on 27 keying machines.

Following the arrival of the DPA, the timing estimates were revised and refined a number of times as the questionnaires were brought closer to their final form, until management felt confident they could be used as the basis for determining the number and size of keying groups to be contracted, and the period of time for which they would be hired. Estimates were calculated in a very conservative fashion—that is, average keying speed was set at a rate attainable by moderately experienced keyers (and not at that which could be expected from skilled professional keyers or keyers who have had long training), and one hour out of each seven-hour shift was assumed to be non-productive, or "downtime" (getting started, breaks, closing down, eventual problems, etc.). For ease of control and greatest efficiency, the very large number of keyers needed was broken down into six keying groups, each of which would comprise ten keyers and one supervisor.

After studying the local labor market, which appeared to offer few persons with the necessary keying skills, and taking into consideration the administrative difficulties entailed by hiring in-house personnel for short-term work, management made an initial decision to use only third-party vendors as a source of labor supply. Based on the estimates of volume and the deadlines established by management, a tender document was developed. An announcement of the tender was placed in the local newspapers, and area firms were encouraged to submit proposals. The tender was issued in September, and a very large number of firms responded. The DPA and her counterpart prepared a detailed analysis of the responses, comparing the cost of all proposals submitted, so that the Census Technical Committee [CTC] could make a decision and award the contract(s).

However, after announcing the tender and receiving proposals from local vendors, management reversed its original decision, and notified the CDPU that it must use both contract keyers *and* local keyers. This decision had serious repercussions on the CDPU staff workload, as well as on the eventual timetable for data entry. [*Details will be found in Section II.A.1.b, below.*]

c. Support staff

Beginning in August 1993, existing Census Project staff were reassigned, for varying periods of time, to work on the multiple auxiliary tasks necessary to carrying out the processing of Census data in all its forms. These support staff can be classed, roughly, by function: data base entry, editing and correction, and operational control.

(1) *Data base entry staff*

The first auxiliary staff to be hired were three computer clerks. Until the post-enumeration keying began, they worked at entering information, in Arabic and English, into the data bases (locality codes and names; education codes and descriptions; occupation codes and descriptions; and industry codes and descriptions) needed for later processing. In addition, one of the clerks assisted in the preparation of table formats.

(2) *Editors*

Following the period of enumeration, staff from the group of clerks carrying out the manual editing and coding task were reassigned to evaluate the error reports generated by the first edit program, and to decide what corrections, if any, to apply to the data. Some of them were eventually trained in the data entry software so they could both identify and apply the corrections, thus streamlining the post-keying operation.

(3) *Operational control*

These staff members did not assume their duties until the post-enumeration period, when Census documents began to move into the CDPU for processing. They were responsible for maintaining control over all documents received by and returned from the CDPU, and for guaranteeing an orderly flow of work to and from the keying and editing sections, so there would be neither backlogs nor idle periods.

## B. System Elements

Not all of the elements of the systems developed to process the various kinds of data collected during the Census operation were foreseen from the beginning of planning. In addition to the phases for which data processors actually planned, management made *ad hoc* requests for information which could only be obtained by diverting resources from a scheduled task to the special request. This caused a number of problems in meeting the deadlines of tasks in the original timetable.

### 1. GEOGRAPHIC DATA BASE

In the original planning, it was decided that the geographic data base, when complete, would serve as a gazetteer of all localities in the Sultanate, even those uninhabited at the time of the enumeration. It would also serve as the basic control file during processing of HPC data, against which the completeness of processing could be verified. To this end, the information collected by cartographers during their field visits was used to establish the basic data base, which was updated by the Crew Leaders [CL] and Enumerators during their work in the field. [*A more detailed discussion of this data base is included in Section IV.B.1, below.*]

### 2. ENUMERATION CONTROL TOTALS

This phase of the system was not foreseen in the original plans. On the eve of enumeration, management requested a system to provide daily totals of units identified (that is, buildings, households, population) during the pre-enumeration listing operation. These totals would be generated at the CL level and would be summarized to the Supervisor level, to provide subtotals by wilayat, region, and Sultanate. At the same time, the Computer Department of the MoD developed a parallel system which would provide to Ministry management continuous monitoring of field performance. [*Further details of this subsystem will be found in Section IV.C, below.*]

### 3. ESTABLISHMENT CENSUS

Royal Decree 91/50 (signed 10 June 1991) had specified that a census of "housing, population, and establishments" be conducted. The Census Technical Committee [CTC] had determined that the appropriate time for collecting information on establishments would be the period, prior to the enumeration of housing units and population, during which the Crew Leaders [CLs] were familiarizing themselves with the Enumeration Areas [EAs] under their responsibility. As part of the process of moving from building to building in each EA, the CLs would complete the questionnaires in the Establishment Census books. Information collected would consist of establishment identification (name and location), activity status, principal and secondary activities, and other relevant data. This information would be used to generate reports and sub-files for further investigation and analysis. *[Further details of this subsystem will be found in Section III.E, below.]*

### 4. SUMMARY FORMS

A critical element of the original processing plan was the subsystem to enter and tabulate the Summary Forms. The books used by the Enumerators were designed in such a way that the first page inside the book was a duplicate, in layout and content, of the book cover. On this detachable page, the Enumerator was to copy all information [identification codes and summary totals] recorded on the book cover. At the end of the enumeration period, while the books were still in the field, the summary page from each book used would be detached and forwarded immediately to Census Project management in Muscat. This would permit management to obtain preliminary totals within ten days of the completion of field operations, and would permit a more accurate estimation of resources needed for the processing of the Census books themselves. *[A more detailed description of this phase will be found in Section III.D, below.]*

### 5. HOUSING AND POPULATION CENSUS

As with all other data collected during the Census operation, the data recorded in the Census books on each housing unit, household, and person in the Sultanate had first to be transferred to a magnetic storage medium before processing could be carried out. Once the data had been entered and key-verified, the first edit (with corrections applied manually) was carried out. Once the data were accepted as structurally sound, they were submitted to the final edit, during which corrections were applied automatically (by the computer program). Once the data were accepted by the user, the required tabulations were generated. These procedures were carried out in two overlapping phases: first, for those EAs included in the 10% sample, and then for the remaining 90% of the data. *[More detailed information can be found in Section III.F, below.]*

#### a. Census Book storage and control

When Census books arrived at Census Project headquarters after the enumeration period was completed, they were checked in using the Book Control Data Base [BCDB] and then stored in containers, in order by EA and book number. The BCDB *[see Section III.A, below]* was created from the Geographical Data Base [GDB] with information to the level of EA. Information on book number and total number of books in each EA was added by the storage control clerk as he checked the books into the BCDB. This data base controlled the movement of all Census books during post-enumeration processing, even though actual implementation of the system occurred after the first processing of the Summary Forms and during the processing of the Establishment Census.

#### b. Sample processing for advance tabulations

To satisfy management's demand for key indicators within six months of the Census date, the Census Project staff was instructed to generate selected tabulations from an appropriately-weighted sample of the data. To this end, ten percent of the EAs in each of five regions were selected for the sample data file; in two other areas (Governorate of Musandam and Al-Wusta Region), all EAs (that is, 100% of the data) were included in the processing. (In these two areas, population size was so small that no processing advantage would have been gained by sampling, and in fact, results might have been seriously distorted because of the bias inherent in the structure imposed by the enumeration operation.) In the remaining area (Governorate of Dhofar), selection of ten percent of the EAs

resulted in a sample which omitted two of the wilayats. A single EA was then randomly selected from each of the unrepresented wilayats to complete the sample, making the Dhofar sample slightly more than ten percent of the EAs in the area. [Details of the procedures used for selecting the sample may be found in Section IV.F.4, below.]

It should be noted that the requirement to produce sample tabulations within the specified time period demanded a great deal of additional effort on the part of the CDPU and it added, as well, another level of complexity to the control operation. In spite of all the additional effort expended, the sample tabulations were never published or otherwise utilized.

c. Processing of non-sample data

The remaining [non-sample] data for the six regions sampled followed the same processing path as did the sample data. Keying and key verification of each Census Book were followed by a preliminary edit with manual correction, after which Census Books for each EA were consolidated to the EA level. When all remaining EAs for a given wilayat had been manually edited and accepted, they were joined with the EAs from the sample to produce a single wilayat-level file containing 100% of the data. The final consistency edit (with automatic correction) was then applied to this file. Once all wilayat-level files for a region had been edited, corrected, and accepted, they were consolidated into a single region-level file for ease in generating tabulations.

Following the generation of the final data file, tabulations were produced on a flow basis. Those tabulations which had been produced for the sample were generated first (to permit the subject-matter staff to compare preliminary results with final results), followed by the remaining table series.

### C. Processing Phases

Processing of data from the Housing, Population, and Establishment Census books was essentially the same from first to last, although the editing processes did vary over time, as the specifications progressed from preliminary to final form. However, before generating final tabulations, all data were edited with the final version of the appropriate edit program. As with any large-scale multi-phase operation, processing was not carried out in rigid, straight-line sequence; rather, operations from different phases and different subsystems overlapped each other, to make optimal use of the resources at hand. During production processing, the schedule was in fact highly dynamic, and the eventual record of activities carried out differed more than a little from that which had been anticipated during pre-enumeration planning.

#### 1. GEOGRAPHIC DATA BASE

The first activities to be put into "production" were related to the Geographic Data Base [GDB]. The worksheets from the field operations began arriving in the CDPU in late summer 1993, and continued arriving up to the point of actual enumeration, at which time the GDB was considered temporarily complete. In the year following enumeration, the GDB was repeatedly modified in light of information from other sources, and was in fact not considered complete and correct until well into 1995.

#### 2. PRE-ENUMERATION CONTROL COUNTS

Activities related to this phase were carried out in the period mid-November to mid-December 1993.

#### 3. DATA ENTRY STAFF SELECTION

While not strictly a phase of processing, this activity required a great deal of time and effort on the part of CDPU staff. Management's unexpected last-minute decision to use local keyers meant that the CDPU had no time to plan for this activity before it was required to carry it out, and the addition of this task to the other responsibilities of the CDPU during the period immediately preceding enumeration (late October-early November 1993) caused serious disruption to the schedule of CDPU activities and the daily routine of the staff. [A detailed description of this activity will be found in Section IV.A, below.]

#### **4. SUMMARY FORMS (I) KEYING AND TABULATION**

Systems development for this phase began in mid-summer 1993, and production processing was carried out in the week immediately following the enumeration. Final results from this phase were delivered on 17 December 1993, but there were occasional recursions to these data in the months that followed.

#### **5. SELECTION OF 10% SAMPLE EAs**

Planning for the sample selection began in late 1993, after management had decided to adopt the recommendation of the U.N. Advisor on Sampling Methodology. Necessary programs were designed, but could not be completed until after enumeration was carried out and the precise number of EAs in each wilayat was known. The sample was drawn in mid-January, so that editors would be able to give priority to EAs in the sample.

#### **6. ESTABLISHMENT CENSUS DATA PROCESSING**

Entry of Establishment Census data began in the last ten days of December, as soon as work was completed on the Summary Forms. Manual editing began approximately one week after keying, and all entry and validation was completed by the last week of February 1994. The final computer edit was applied and generation of tabulations began in March 1994. The last of the tabulations, as originally requested by the user, were delivered to the user by June 1994.

#### **7. SUMMARY FORMS (II) KEYING**

The edit program required for the second round of Summary Forms processing was completed in January 1994, and keying, editing, and tabulation was carried out in February 1994.

#### **8. POPULATION DATA (10%) PROCESSING**

Keying and verification of sample data began in mid-February and was completed by mid-March. Manual editing began approximately one week after keying, and required several weeks to carry out. Keyed and edited batches were consolidated into EA files; EA files were consolidated into CL files; and CL files were consolidated into wilayat-level files for purposes of the final edit. The version of the computer edit current at that time was used to complete editing of the sample data and generate a tabulation-ready file. Weights were added in May, and the tabulations were produced by mid-June 1994.

#### **9. POPULATION DATA (90%) KEYING**

Keying and verification of the remaining non-sample data began as the verification of the sample data was being completed (that is, the phases overlapped to avoid any slowdown in production). All keying and verification were completed by late July 1994.

#### **10. DATA EDITING (MANUAL)**

Editing of the 90% data for manual correction of errors began as soon as the sample data editing was completed, and continued through early August 1994.

#### **11. DATA CONSOLIDATION**

Just as data entry and verification were done on a container (and therefore on a geographic) basis, consolidation of batches into higher-level file groups was carried out following the field structure used in the enumeration. That is, individual batches (books) were consolidated into a single file for each EA. When all EA files for a given CL area were complete, they were consolidated into a single file for the CL area. For areas of larger population, when all CL files for a given AS area were complete, they were consolidated into a single file for the AS area, and those AS area files were then consolidated into wilayat-level files. For areas of lesser population, CL area files were consolidated directly into wilayat-level files. Consolidation, therefore, began as soon as the first batches of data had been edited (March), and continued until all data were ready for final editing (August).

## **12. DATA EDITING (AUTOMATIC)**

Editing of the data for correction by computer was carried out in phases, on a wilayat-by-wilayat basis, as the keying and manual editing of each wilayat was completed. Prior to the edit, all EA files from the sample EAs were incorporated into their respective CL files, so that the final edit was performed on 100% of the raw (unedited) data for each wilayat. Editing of the full data file was performed in August/September.

## **13. ESTABLISHMENT CENSUS NAMES AND INDUSTRY GROUP CODES**

At the request of the users, the establishment name (in Arabic) was added to the record for each active commercial establishment in the Establishment Census data file. This activity was carried out during August 1994.

## **14. HPC DATA FILE ACCEPTED BY USERS**

In August 1994, after final editing, the file containing 100% of the housing, household, and population data was approved by management for use in tabulation.

## **15. TABULATION OF 100%**

Tabulation of the final data began immediately upon acceptance of the data by management (September 1994), and continues to the present date. It is estimated that all publication tabulations will be produced by September 1995.

#### IV. DETAILED PROCEDURES

In this section, the procedures followed in carrying out each of the subsystems and phases of processing mentioned above will be described in detail. The description will include appropriate references to the software modules used, although technical details of such software will be found in the relevant appendices.

The process of developing the multiple subsystems was carried out with varying degrees of difficulty, for a number of reasons. Some subsystems were poorly defined; some were defined sufficiently to begin development, but were subject to frequent change; and some subsystems were developed with virtually no specifications from the user, relying on the data processor's experience and intuition to achieve the desired result.

NOTE: All Arabic-language text, in all Census-related applications, was entered using the Nafitha™ Enhanced character set (code page 711) of the Nafitha™ Arabization software package.

##### A. Data entry

As was mentioned earlier [*see Section III.A.2.b, above*], the data entry function would be carried out by a combination of keying groups provided by local contractors and keying groups hired directly by the Census Project. [The standard keying group was defined as ten keyers and one keying supervisor.] From the beginning, it was assumed that, with the exception of entry of the Summary Forms, for which only one contract keying group would be required, all keying groups would participate in all phases of the Census operation (i.e., Establishment Census and Housing and Population Census), but that participation would be gradually phased in—that is, keying would begin with one group/one shift; then, as the volume of output from the manual editing/coding staff increased, the remaining groups would be phased in to avoid either excessive backlogs or idle keyers. In actual practice, this plan was followed only partially.

##### 1. STAFF SELECTION

The process of selecting staff was different for contract and local-hire. In the case of keyers to be hired from local contractors, the CDPU was responsible only for providing to the CTC a worksheet analyzing the offers received, after which the CTC made the final decision on awarding the contracts. However, with respect to the keyers hired locally, the CDPU was left with complete responsibility for recruiting, evaluating, and selecting the required number of staff.

##### a. Contract

By recruiting keyers through local providers of such services, the Census Project would pass responsibility for recruitment, testing, and selection of appropriate personnel to the contractor. In addition, the contractor would also have permanent responsibility for seeing that his keyers continued to meet performance standards. If any keyer did not meet standards, the contractor would be responsible for finding substitute personnel with no delay to project deadlines. Any delay attributable to a contractor's performance would be punished by financial penalties.

Based on the offers received, the Census Technical Committee decided to award contracts to two vendors: INSYST and OITC. INSYST would provide a total of two groups, one of which was to begin work immediately upon the completion of field work with the keying of the Summary Forms, and the other of which would be added when a sufficient number of Housing and Population Census books became ready for keying (estimated to be late January). The OITC group would also begin at the same time as the second INSYST group. The first INSYST group would be released before the other two groups, in a gradual phase-out of keying activity.

##### b. Local

Because data entry was originally to be accomplished using only contract keyers, the CDPU did not make plans to carry out any recruitment or selection activities with regard to keying personnel. The subsequent last-minute insistence by management that three groups of keyers be recruited locally to supplement the three contract keying groups meant that the CDPU was suddenly responsible for recruiting and testing enough candidates to permit selection of thirty keyers (if possible). The ad-

vertisement which appeared in the local Arabic-language newspapers generated a strong response, with almost two hundred candidates requesting an opportunity to try out for the positions. Because management also insisted that the keyers begin work at the same time as the first contract keying group, it became necessary to defer system development work in favor of the testing and evaluation of the local candidates.

It was decided that testing would be based on the actual data entry application to be used during production keying. This application had been developed using the CENTRY module of IMPS. Since the enumeration had not yet been carried out and completed Census books were not available, the data entry screens were modified to reflect the information found in the books used during the Pilot Census. CDPU staff created batch files, containing basic geographic identification and the Book Cover record, for each of the books to be used, and copied the batch files and keying application to each of the machines to be used for testing.

Initially, only machines not being used by the CDPU staff for programming and system development were to be used for testing, but after the first few days, the large number of candidates and the rapidly-approaching deadline forced the use of programmers' machines, as well, and thus normal development work had to be suspended for much of each day during which testing took place. In addition, it was necessary to schedule testing not only during normal work hours, but in the late afternoon and evening, as well, to accommodate the greatest number of candidates. Even with all this effort, by the time production keying began in mid-December, there still remained candidates who could not be tested, and therefore not considered for employment, because of lack of time.

For greater efficiency, candidates were scheduled for testing in groups. At the first testing session it became clear that, in spite of the stated requirement that candidates be "computer literate," many of them were unfamiliar with computers of any kind, and many had never used a keyboard, even on a typewriter. Therefore, at the beginning of each testing session, the staff member administering the test explained the basic functions of the microcomputer and demonstrated how to invoke the data entry program. Each candidate was then given a book for a practice keying session, which lasted approximately fifteen minutes. Finally, each candidate was asked to key as many books as he/she could in one hour (timed by the person administering the test). The one-hour interval was a compromise between the minimum time needed to obtain a sense of the candidate's likely performance on the job and the maximum time that could be allowed to a single candidate, given the overwhelmingly large number of people to be tested. The CDPU quickly discovered that allowing sufficient time for explanations, practice, and actual testing made it necessary to block at least two hours for each testing session. In practice, sessions often ran even longer, causing a "ripple" effect which extended the staff workday well into the evening.

Following each testing session, CDPU staff key-verified the work produced by the candidates, which ranged from part of a book to two or more books, depending on the skill of the individual tested. After each batch was verified, the statistics for the batch (keying speed and error rate, collected automatically by the CENTRY software) were displayed and recorded, so that candidates could be ranked somewhat objectively. After several weeks of testing, the CDPU staff felt that enough potential keyers had been identified to permit a return to normal working operations. The approaching deadline (completion of Census field work and beginning of production keying) also made it urgent that no more time be spent on unplanned-for activities, since other more important tasks remained uncompleted and management was imposing new demands daily. Those candidates tested were ranked according to a number of factors (performance on the test, previous work experience, willingness to work either shift) and the list of names selected was passed to the administrative section of the MoD to carry out the formalities of hiring.

Once those candidates meeting minimum requirements had been identified and contracted, three keying groups (ten keyers each, as with the contract keyers) were formed. Because of the lack of relevant experience among the new hires, and in order to more closely monitor their work, it was decided to use CDPU staff as supervisors for the local keyer groups. Each of the three staff members who had been carrying out most of the auxiliary tasks (geographic data base entry, verifying keying tests, etc.) was assigned to supervise one of the groups of new keyers. Since none of the new hires had much experience in production data entry, intensive training was needed to make the keyers

comfortable with the equipment, the software, and the routine procedures to be followed. The supervisor of each group was given responsibility for overseeing the training of the keyers in her group. Preparation of the new keyers began shortly after production data entry started, and continued until keyers were needed to work on entry of Establishment Census books.

## 2. SHIFTS

Management's insistence on accomplishing data entry in the shortest possible time was one of the factors that led to the decision to use six workgroups (60 keyers). The original request for equipment had assumed 27 microcomputers would be needed; to accommodate six workgroups, 66 machines (keyers and supervisors) would be needed. In the interests of economies of time, scale, and space, CDPU staff decided to split the six groups into two shifts (7h30-14h30 and 15h00-22h00), which would cut the number of computers needed to 33. Data entry would be run five days a week. (It should be noted that no modification was made to the keying schedule during the month of Ramadhan, although keyers were permitted to substitute weekend days for the Eid holidays.)

When only contract keyers were contemplated, it made little difference which shift a contract group would work, since the contractor agreed to accept any shift assignment. However, once the decision was made to hire local keyers, it soon became apparent that the large number of females among them would make it preferable to segregate the local keyers from the contract keyers as much as possible. It was thus decided that the first contract group would begin keying on the morning shift, but as soon as the second shift was required, all contract keyers would work the afternoon shift and all Omani keyers would work the morning shift. This plan was in effect throughout production keying.

## 3. KEYING ROOM SET-UP

A few months before enumeration, the DPA and one of the Census Project administrative staff had visited all suppliers of computer and office furnishings in the Muscat area and, based on what was available, had made recommendations to the Census Project with respect to keying tables and chairs. Space in what had been the MoD library was allocated to the data entry operation, but delays in construction of the new building (to which the library would eventually move) impeded efforts to clear out the room and set up the keying environment until virtually the end of the enumeration phase. A sufficient number of tables and chairs had originally been purchased, but in the interval between the delivery of the furniture and liberation of space for the keyers, many of the tables and chairs had been diverted to other units of the Census operation, particularly the Central Control room for the field operation. When, the night before keying was to begin, it was finally possible to set up the keying room, tables and chairs were available for only about 20 of the keying machines. This was sufficient for the first group to begin, but it was necessary to scour other offices in the Ministry for unused desks and chairs in order to complete furnishing the keying room. Unfortunately, almost all of the replacement desks were standard office desks, totally unsuitable for data entry. The greater desktop height and the limited leg space available made it difficult for keyers to work comfortably, and may well have contributed to a higher error rate for clerks at those stations.

## 4. STANDARD PROCEDURES DURING PRODUCTION KEYING

From the beginning, the data entry operation functioned on the principle that no keyer should be permitted to verify his/her own work. At the start of operations, when only one (contract) keying group was working on the first Summary Forms, keyer-verifier independence was controlled by the group supervisor, who assigned each batch individually. When the volume of work began to rise and additional groups were added, this independence was raised to the group level and formalized—that is, work assignments were allocated to a keying group, with the proviso that no keying group could be responsible for both keying and verifying the same Census Books. This independence was monitored by the operations control clerks (Data Controllers).

Because the keying room had been used as a library, many built-in shelves were still available. Sections of the shelves near the entrance were marked off and used for temporary storage of materials to be keyed or verified. At the beginning of each shift, the keying supervisor would check the work available in the storage areas and select an assignment for his/her keying group. Selection of work units was done on the basis of two general rules: (1) no group could verify the work it had keyed, and (2) work to be verified

should not be neglected in favor of keying. After the first several weeks of trial and error, the keying group supervisors and the Data Controllers became adept (with the constant assistance of the CDPU staff) at balancing the keying/verifying workload, so that no significant backlogs developed. This general routine was followed beginning with Establishment Census keying and continued through the completion of Population Census work.

## B. Operational Control

There were two major and interconnected elements of operational control used during the processing of the 1993 Census: the Geographic Data Base [GDB] and the Book Control Data Base [BCDB]. These elements are described in the sections which follow.

### 1. GEOGRAPHIC DATA BASE

This collection of data was fundamental to virtually all phases of processing of Census data, and was used as the basis of the BCDB. In its final form, it may be considered to be a complete directory of all known localities in the Sultanate of Oman, and contains, in addition to locality names, housing unit, household, and population totals for each locality.

#### a. System development

Initial development of this system began with an attempt to use the CENTRACK module of IMPS. This utility will generate a data base which can be used for monitoring and controlling post-enumeration operations (editing, coding, data entry, consolidation, etc.). The system requires construction of a data base with a structure which reflects the geographic structure to be used for field enumeration. The data base should also contain one (and only one) record for each unit to be tracked (for example, EA, book, etc.). In the case of the Omani Census, a decision had been made to track each book individually during the post-enumeration processing.

This decision had been mandated by the use, during enumeration, of multiple books within an EA, where each book could have geographical codes different from other books in the same EA. In areas of greater population density, the codes usually differed at the hilla and/or block level, and for those EAs covering multiple small villages, the codes differed at the locality level. However, in a few rare cases, the codes differed at the highest levels (region and/or wilayat). This latter difference was usually the result of post-enumeration adjustments, but any of these differences in codes meant that each book had to be tracked separately, in order to maintain proper control and produce proper results.

In the original system design, the only function of the Geographic Data Base [GDB] was to be the foundation of an operational control system for the data entry, verification, and editing phases of data processing. To fulfill this function, the GDB needed to contain only the numeric codes assigned to each locality, to the lowest level needed for tracking production. In the case of the Census books, this would require all levels of codes (region, wilayat, locality, hilla, block, EA, and book number), but would not require any alphanumeric information (that is, names).

However, management soon made it clear that the GDB must contain names, both Arabic and English. The GDB was then expanded in design to include names for all geographic entities at the levels of region, wilayat and locality. In addition, where hilla names were available, they were also included. The data base was initially designed to include only the codes and names, but eventually was modified to include count fields (i.e., numbers of housing units, households, and population), for ease in generating reports with totals by locality, hilla, and block.

In addition, the GDB's function as a checklist of processing units for purposes of operational control also underwent modification. In the original conception, the GDB would have been composed of one record for each unit to be monitored—that is, the Enumeration Area [EA]—on the assumption that all geographic codes would remain the same within any given EA. However, the experience of field work made it evident that this assumption was false, because one EA often included different localities and/or hillas and/or blocks. Since geographic codes were uniform **only for any**

**given Census book**, the book was defined as the basic processing unit and the control system was designed to permit tracking individual books as well as EAs.

The system to manage this GDB was developed in phases, as the users' needs were modified and refined, and could not be considered complete until almost one year after enumeration. The data base itself underwent constant modification in response to information from users both within and outside the Ministry, and from the Census data themselves. The GDB data files were developed using the Clipper Version 5.2 software. Executable modules were compiled using the same version of Clipper. Subsequently, in response to user and CDPU requirements, the dBase IV software package (Release 1.5) and FoxPro were also used to produce lists and routines for checking, viewing, and browsing the data files.

b. Data base structure

*[Section VI.A, below, includes a description of the structure of the GDB.]*

c. Entry of codes and names

A dBase application was generated to permit the CDPU support staff to enter codes and names from the lists prepared by the cartographers and the field operations staff. These lists were usually organized at the wilayat level, and contained an entry for each unique combination of locality, hilla, block, and EA codes. (Of course, the number of books that would later be used during enumeration was not yet known, but it was assumed to not exceed 99 in any given EA.) Each locality's name was indicated, in Arabic and English, although many differences were encountered in the transliteration from Arabic to English. Where the hilla had a name, it was also indicated. The Nafitha™ Arabization software permitted entry of text in both languages. Data for each wilayat were keyed into a separate set of data files to permit ease of modification and checking and to reduce the risk of file damage or corruption.

d. Operational control

The creation of the GDB was carried out in a non-systematic manner—that is, the worksheets were delivered to the CDPU on a random basis and at unpredictable intervals, and it required continuous urging by the CDPU staff to obtain all worksheets needed to complete the data base.

Because the number of localities was not known before the GDB was created, there was no way to verify that the number of entries in the GDB was accurate. Control of this operation within the CDPU therefore was limited to a manual check that at least one worksheet had been received for each of the 59 wilayats in the Sultanate. Whether the total number of localities entered into the GDB for the wilayat was correct or not was discoverable only after field operations were carried out, and even after enumeration there continued to be many questions about the accuracy of locality names and codes.

e. Quality control

After each entry session, a list of the information entered (i.e., the new entries in the data base) was generated and returned to the user for comparison with the original worksheets. Any errors or omissions were indicated on the printout, which was returned to the CDPU for eventual correction. This cycle was repeated until the information in the data base was to the user's satisfaction. Once wilayat-level data had been accepted by the user, region-level data files were created and listings, with names in Arabic and English and all codes, produced for further verification. Eventually, a Sultanate-level GDB was created, from which printouts, in order by EA with all locality names and codes, were produced for distribution to field Supervisors, Assistant Supervisors, and Crew Leaders for use during enumeration.

## 2. CENSUS BOOK CONTAINER STORAGE (BOOK CONTROL DATA BASE)

In the original plans for the Census operations, the CTC had approved a manual system (designed and implemented by the Census Advisor) for receipt and check-in of Census Books after enumeration. All the necessary forms were designed and made ready, and office staff were trained in the procedures to be followed when Census Books arrived from the field after enumeration. After enumeration was completed, MoD management reversed the decision of the CTC, and insisted on a computer-based system. Because of the last-minute nature of this decision, no planning was possible, nor did the CDPU have staff available to develop such a system. It was therefore necessary to call in an outside contractor to create the desired system. Specifications for the system were established by management, in coordination with CDPU staff, and development was carried on at the contractor's work site. The contractor based his system on the GDB developed by the CDPU staff, using FoxPro software (Version 2), and the modules were eventually copied to the Census Project microcomputers to be accessed and used by the BCDB. The two data bases were linked to create a single file, with information on each Census Book used. This combined file provided listings of localities or higher-level entities, with approximate totals of HUs, HHs, and population; it also helped in checking consistency of identification between Census Books, the GDB, and the BCDB.

The purpose of the BCDB system was to track the containers of Census Books as they moved through the phases of processing, from pre-entry editing and coding, through key data entry and verification, and finally through the manual editing and correction process. The process was controllable at the sub-container level—that is, an EA in the sample could be removed from its container and sent through processing without causing problems of control for the remaining EAs in the container.

Because this system had not been planned for, no equipment had been allocated to this function, and it was necessary to make an emergency procurement of microcomputers and printers to implement the system. In addition, staff who had been trained in the use of the manual system now had to be re-trained in use of the computer-based system; many had difficulties adjusting to the use of this equipment, and the software itself went through a number of revisions in an effort to remove the "bugs" and conform to the specifications. [*Further information may be found in Section VI.A, below.*]

### C. Pre-enumeration control counts

In the pre-enumeration period during which the Crew Leaders (CLs) were numbering buildings, field Supervisors were instructed to transmit to the Central Office daily summaries of estimates of buildings, households, population, etc. in each CL area. Forms were forwarded to the CDPU for data capture and processing, and after processing, were returned from the CDPU to the Census Project management for permanent storage.

#### 1. SYSTEM DEVELOPMENT

When the pre-enumeration field operation began, those in charge realized they needed certain information from this operation, but that they had neglected to specify to the CDPU that such reports would be required. As a result, a system to provide management with the desired reports was programmed and implemented at the last minute. (At the same time, the Computer Department of the MoD developed a system, running on the HP9000 minicomputer, to provide similar information on field operations during the actual enumeration.) The system developed by CDPU used daily production reports from the field and generated tables tallying numbers of housing units, households, and persons from estimates made by the field workers during the process of numbering buildings. In addition, at the request of the Ministry of Communication, the tables showed the number of households reporting good and bad TV and radio reception, as reported to the Census field staff. The results for each successive level of the field hierarchy, from Crew Leader through Supervisor, were calculated, with subtotals by wilayat within region. After each day's production was tabulated, cumulative totals were produced by combining current and previous data files.

This system was developed using IMPS software. A CENTRY module was created to permit CDPU staff to enter the data, and information submitted to Census Project management each day was transmitted to the CDPU, where it was keyed and verified. Following the data entry, the data were input to a CENTS tabulation program. Since no edit specifications were provided by the user, the tabulation program had a "trap" for data records with information missing (such as number of housing units, number of population, etc.).

## 2. FILE STRUCTURE

The file contains one record type; the data dictionary [BOOKSUM.DD] describes two record types (the actual data record and a dummy record) to permit the use of a record type code in the data.

## 3. ENTRY OF INFORMATION

As forms were received in the CDPU, they were grouped by region and wilayat, and the number of forms for each region/wilayat combination was recorded on the control form. All forms for a given region were grouped into one batch for keying and verification. CDPU support staff carried out the data capture operation. After verification, the number of records in the batch was compared to the number of forms; any differences were investigated and corrected.

Operational control was made more difficult by the way in which forms were submitted and by the structure of the field operation. Because the Supervisors in the field depended on fax transmission (which occasionally presented technical problems) to report daily count summaries, and because of the modifications made to the field structure right up through the period of enumeration, there was no way to be certain that all CL areas had reported. In fact, it is now clear from the final data that all CL areas did not report, and of those that did, many reports were incomplete and/or were identified with incorrect geocodes.

## 4. TABULATION

No editing was carried out, since the user specified no checks on the data. The keyed and verified data file was sorted and then input to the table-generating programs. Management had requested two separate table formats (one containing calculations of average number of persons per household and average occupancy rate, and one with simple totals). It was also requested that totals be presented for each CL area within each Supervisor area, and for the same control levels within each region and wilayat. Because control levels are a function of the AREA-STRUCTURE statement (in the CENTS subsystem), different TAB programs, CON parameters, and area name files were needed to produce tables at the different levels requested.

### D. Summary Forms

The purpose of the Summary Forms was to permit Census Project management to announce preliminary totals within a few days of completion of enumeration. The first page of each Census Book was a replica of the Book cover, and was designed to be detachable. On completion of the Book, the Enumerator was instructed to tally up the number of housing units and households enumerated, and write these numbers into their respective boxes on the Book cover. The enumerator was then to tally up, for each household, the number of Omani males and females, and non-Omani males and females, and place these numbers in their respective boxes on the page where the listing for the household began. Once all household totals had been derived, they were to be summed and the results placed in the appropriate boxes on the Book cover. Once the Book cover was complete (geographic codes and summary counts), the numbers were to be transferred to the first inside page (the replica of the Book cover), which was called the "Summary Form."

At the end of the official enumeration period (1–10 December), after a final review of his/her books, the Enumerator was to detach the Summary Form page inside each book and keep it separate from the Books themselves. Books and Summary Forms were to be delivered to the immediate supervisor (the Crew Leader), for delivery to the Assistant Supervisor. Once the Assistant Supervisor had received Summary Forms from all Crew Leaders in his area, he was to dispatch them immediately to the Census Project central office at the MoD in Muscat. (Census Books followed a separate, and slower, path to the Central Office.)

In spite of all precautions taken by the CDPU, the receipt of the Summary Forms from the field offices was not without difficulties. The DPA and the CDPU staff together had devised procedures to ensure that all forms would be counted on arrival and that appropriate control forms would be filled in to monitor the progress of each batch of forms through the keying and verification process. However, last-minute demands imposed on the CDPU by MoD management nullified much of the planning and invalidated many of the controls which had been established. This created serious and time-consuming problems for the CDPU staff in their attempt to process the Summary Forms before the prescribed deadline.

During the processing and tabulation of these Summary Forms it quickly became clear that the majority of the forms had errors in arithmetic (that is, the male and female totals did not add properly, nor did the Omani and non-Omani totals) and/or errors in the geographic codes reported (including missing information). Tabulation for preliminary results was thus done using only male and female counts for the Omani and non-Omani populations.

As a result of the haste in processing which had affected the first results, Census Project management decided to re-create and re-process the Summary Forms after the Census Books themselves had arrived at the central office. Unfortunately, due most probably to lack of attention on the part of the clerks assigned to the task, there were almost as many errors in the second set of forms as in the first. During the second round of processing, the pressure to produce results by a specific deadline was absent, so the CDPU was able to "edit" the data and ensure that, at a minimum, the numbers on each form were consistent, even if they were not absolutely accurate with regard to the information contained in the Census Book from which the Summary Form came.

## 1. SYSTEM DEVELOPMENT

The Summary Forms were included in data processing system design from the beginning. However, as can be seen from the description below of the detailed procedure for this subsystem, the execution was marked by unplanned-for activity.

This subsystem was developed using the IMPS software. Separate CENTRY modules were created to permit key entry, key verification, and key modification of data from the forms. As with the pre-enumeration counts, no edit specifications were provided, so the first CENTS tabulation program had to include data-screening functions that would normally be covered in an edit program. As will be seen from the descriptions [in Sections *IV.D.3* and *IV.D.4*, below], the very large number of errors in the forms made evident the need for a separate edit program. For the second round of Summary Form processing, the DPA developed a pre-tabulation CONCOR edit program, based on "common-sense" specifications, to flag errors (geographic and arithmetic) for manual correction. Tabulation provided totals of population (Omani and non-Omani) by sex (male and female) and administrative division (region and wilayat). The tabulation system also permitted summation at levels below wilayat (for example, locality or EA), if necessary. [Details may be found in Section *IV.D*, below.]

## 2. FILE STRUCTURE

The data file contained one record type [SUMMARY], defined in the Data Dictionary used for the Housing and Population Census Books [CENSUS.DD].

## 3. ENTRY OF FORMS (I)

As mentioned in an earlier section, the timetable established by management required that the entry of data from the forms be accomplished in a very few days. It was expected that enumeration would be concluded on 10 December 1993, and that it would require two or three days for the Summary Forms to be detached from the books, grouped into bundles according to Crew Leader and then Assistant Supervisor area, and transmitted to the Census office in Muscat. In fact, the first groups of Summary Forms began to arrive on 12 December, and the first group of contract keyers, which had been on duty since the last day of enumeration, was able to begin entry of the data without delay.

### a. Operational control

Well before Census enumeration began, the CDPU was aware of the importance of proper control procedures to the success of the Summary Forms processing operation. For this reason, the CDPU staff designed a detailed plan for receiving the forms and controlling their flow through the process. The procedure required that both the field staff person delivering the forms and the CDPU staff member count the number of forms in each batch received, and that detailed control forms be filled in, showing the geographic identification (CL areas) and the number of forms received for each area.

Once the forms were accepted for input, the CDPU staff member would assign a unique batch name for use by the data entry staff. The name would be derived, in a systematic fashion, from the geo-

graphic codes appearing in the batch of forms. Once the control form was completed and batch name assigned, the forms would be sent to the data entry operation. After keying and full verification, the forms would be filed for future reference, and the data copied from the diskette to the hard drive of the main processing computer. Batches of data would later be consolidated into regional files for tabulation.

During the enumeration period, the DPA explained to the CDPU staff members the procedures to be followed and the forms to be used. During the first two days in which forms began to arrive from the field, the procedures were carried out precisely according to plan and to schedule. However, on the third day, management became impatient with progress and assigned non-CDPU staff to the task of checking in and batching forms. Unfortunately, the later arrivals did not clearly understand the procedures to be followed, and a number of mistakes were made in processing the forms and in assigning batch names. Correcting these mistakes created additional work for the CDPU staff during a period when there was already a very tight time schedule. This situation could have been avoided if management had permitted the CDPU to carry out its procedures as originally designed and without unnecessary modifications.

b. Data entry

Entry of data from the Summary Forms was controlled by using the CENTRY application SUMM.AP. Data were verified using the same application, and the resulting batch files were modified, when necessary, using the application MSUMM.AP. Each batch of forms was assigned to a specific keyer, and key verification was carried out by a different operator. The keying supervisor noted the identification code of both keying and verifying operators on the control form which accompanied the Summary Forms. Upon completion of keying and verification, all forms were returned to the CDPU along with the diskette containing the keyed data.

c. Tabulation

The subject-matter staff had defined three basic table types for displaying results of the Summary Form processing. One table showed totals for each region and for the Sultanate; another table showed totals for each region, and within the region for each wilayat; and the third table showed results for all urban localities within each region and in the Sultanate. (It should be noted that Census Project management had earlier identified the specific localities to be considered "urban.")

#### 4. RE-CREATION AND ENTRY OF SUMMARY FORMS (II)

As was mentioned earlier in this section, management decided to re-create and re-process the Summary Forms. [For details of the manual review process, see Section II.F.4, below.] When a sufficiently large number of forms had been accumulated for a given area (to the level of CL or AS, at least), they were transmitted to the CDPU for processing.

a. Operational control

To monitor completeness of processing—that is, to ensure that every Summary Form was processed once and only once—a control form was developed on which the numbers of forms for each CL area were recorded. The control form provided a means of tracking each batch of Summary Forms, at the CL level, from beginning to end of its journey: from the Central Stores, where the forms were completed, through all the phases of processing, and back to the Central Stores after the data were accepted. At each step of the process, the person responsible for the section involved was required to acknowledge receipt of the forms by signing and dating the control form attached to the batch of Summary Forms.

Even though the control form provided for the maximum number of CL areas permitted in one AS area, it was not necessary, nor in many cases was it possible, for all CL areas in the same AS area to be processed simultaneously, using the same control form. The Data Controller (in the CDPU) was responsible for maintaining the file of control forms, to ensure completeness and integrity of processing. Daily reports from each section permitted monitoring progress, and by checking the

record of Summary Forms submitted against the Book Control Data Base, it was possible to identify discrepancies in numbers of Forms.

b. Keying

Data entry was carried out under the same CENTRY applications used in the first round of Summary Forms keying. The same procedures for guaranteeing independence of keying and verification were followed, with the additional restriction that keying and verification of batches should be carried out by different groups, and not just by different keyers in the same group. For this reason, the Data Controller assigned the keying of all CL areas listed on a single control form to a particular keying group. The supervisor of that keying group assigned the various batches to individual keyers. When keying of all batches was completed, the supervisor placed diskettes, forms, and control form on the "Keying Completed" shelf, and the supervisor of another keying group retrieved the materials for verification. When all batches listed on the control form were verified, all materials were returned to the "Verification Completed" shelf, and the Data Controller retrieved the materials for the next step (data editing), which was carried out in the CDPU.

c. Quality control (editing)

During the first round, the exigencies of time did not permit formal editing of the data, even though many of the geographic codes were invalid or missing, and many of the numbers (housing units, households, persons) were either missing or inconsistent with each other—that is, the numbers of males and females reported did not add up to the total reported, and/or the numbers of Omanis and non-Omanis reported did not add up to the total reported. Since there was no time to research the errors and determine which, if any, of the numbers were correct, first-round tabulation used only the basic male and female Omani and non-Omani numbers, and generated all totals by calculation. This ensured that the final numbers were consistent with each other, but could in no way guarantee the validity of the numbers themselves. Numbers of housing units and households were accepted as reported, since there was no way to check their validity apart from the Census Books themselves, which were not available at that time.

During the second round, editing was mandatory, since the sole purpose of repeating the effort was to provide more accurate numbers against which to control the subsequent processing of Census Books. In the absence of formal specifications prepared by the subject-matter staff, the DPA devised a simple program to cross-check the various numbers (including the geographic identification) for validity, reasonability and consistency. The program was executed for each batch of Summary Forms, producing a list of errors on a form-by-form basis. The error listing and the Summary Forms from which the data were keyed were returned to one of the clerks in the forms preparation section. The clerk was expected to review the errors and, where necessary, to return to the Census Book to determine which numbers were correct. Any changes were to be indicated on both the error listing and the Summary Form itself, and after all corrections had been marked, forms and listing were returned to the CDPU. Programming staff in the CDPU made the required changes to the batch, and the edit program was again executed. If errors persisted, the cycle was repeated until no more errors were found.

In only a very small number of batches (perhaps five percent of the total number) did the first execution of the edit program detect no errors; in approximately sixty percent of the batches, errors were detected and corrected in one round. However, in the remaining approximately thirty-five percent of the batches, resolving all errors required at least two iterations of the correction cycle, and in some cases three or even four error listings had to be produced to eliminate all errors. The high error rate was almost entirely due to carelessness—particularly in cross-checking arithmetic calculations—and lack of attention to detail on the part of the clerks who reviewed the Census Books and filled in the Summary Forms. When all batches had been corrected to the satisfaction of the subject-matter staff, they were consolidated into region-level files and then into a Sultanate-level file.

#### d. Tabulation

Tabulation of the edited Summary Forms data file was carried out in exactly the same manner as during the first round of processing, using the same table formats (.FMT file) and tabulation program (.TAB file) to produce the results as had been used earlier.

### 5. USE OF SUMMARY FORMS IN LATER PROCESSING

The Sultanate-level file of re-created Summary Forms was used during the post-data entry manual edit to validate data files keyed from the Census Books with respect to geographic identification and numbers of records [*see Section II.F.7, below*]. It was also used in calculating weights to be assigned to sample EAs. In addition, it was linked to the GDB to provide quick listings of localities or higher-level entities with approximate totals of housing units, households, and population. Much later, after the data file for the Sultanate was accepted as complete, a new Summary Forms file was generated from the data file itself. This permitted the use of the GDB to create tables including both names of geographic and administrative entities (in Arabic and English) and totals of HUs, HHs, and population for each entity. In addition, the generated Summary Forms were used to check consistency between the Census data, GDB, and BCDB.

### E. Establishment Census data

Collecting data on establishments was a major sub-operation during the period in which the Census workers were in the field. An **establishment** was defined as any structure which was not wholly residential in nature, and thus included not simply business establishments, but also mosques, schools, and buildings of mixed use. It was not necessary that an establishment be currently in operation to be included in the count of establishments. The information was collected during the first two weeks of November 1993, during the operation in which the master listings of housing units and households were created. In preparing these lists, each Crew Leader canvassed the area under his supervision, EA by EA, and used the operation not only to collect the data on establishments, but also as a final check on, and opportunity to update, the maps which had been provided to his enumerators. [*Attachment II.E.1 shows a sample set of pages; Attachment II.E.2 is a descriptive list of the data collected.*] The completed books were accumulated to the level of Assistant Supervisor and were sent to the Central Office before beginning the HPC enumeration. Once the books arrived at the Central Office, they followed a path separate from that of the Housing and Population Census books in both manual and computer processing.

#### 1. SYSTEM DEVELOPMENT

During the pre-enumeration listing operation, Crew Leaders [CLs] were required to collect specific information on all establishments, which were defined to include any and all structures used for purposes other than exclusively residential. "Establishments" thus included mosques, schools, and other public and private buildings in which activities other than those of private households were carried out. The Establishment Census [EC] attempted to be as comprehensive as possible, including along with active entities those establishments which were identifiable but not currently in operation.

The programs in this subsystem, with the exception of entry of establishment names, were developed using the IMPS software. To ensure standardization, the DPA developed a data dictionary containing descriptors for the record types (book cover and individual establishment) derived from EC books. The descriptor for each record type included descriptions of all individual data items and permitted values for each item. For selected items, group definitions were made available to facilitate editing and tabulation. A decision was made to maintain this dictionary separate from that used for the Housing and Population Census, since there is no commonality of data between the two operations other than geography (and that only to the level of CL area). This dictionary was subsequently used as the basis for all other IMPS modules used to process EC data. For data capture, the CDPU staff developed separate CENTRY programs for key entry, verification, and modification of EC data.

Edit specifications were prepared by the Census Advisor, but the specifications provided only for manual correction. Thus, the DPA developed a post-verification CONCOR edit program which would generate a listing flagging errors of validity and consistency. These listings, and the EC books from which the data were keyed, would be used by editors to determine the corrections to be applied. Corrections would be

marked on the listings by the editors, and books and listings re-cycled to the data entry operation, where corrections would be applied. In theory, all errors would be caught and corrected in this phase, and the data would be ready for tabulation; in practice, however, many errors persisted, and the DPA was obliged to create another CONCOR edit program, to be run on the data before tabulation. The purpose of this program was to catch all errors which could adversely affect the tabulation process, and permit the computer program to determine the best correction.

General tabulation specifications had been delivered to the CDPU earlier, but many critical details were missing because the level of detail of activity classifications was not determined until the post-enumeration period, when the editors began to review the data collected before sending the books to data entry. In addition, the descriptions of activity classifications, in Arabic and English, exceeded system capacity when used as stubs in a table. For this reason, all tables using activity [ISIC] classifications as stub descriptors were subdivided into multiple parts for purposes of table generation. Tables were produced for individual regions and for the Sultanate as a whole. A number of *ad hoc* tabulations were produced, as well, in response to demands from various users. All tables were generated using CENTS programs.

In late summer (August 1994), after the required tabulations had been generated and users had thoroughly reviewed the data, names (in Arabic only) of active commercial establishments were added to the data file. The program for this operation was developed in dBase IV, and the subsequent data file became the master file for Establishment Census data.

## 2. FILE STRUCTURE AND CONTENT

The information collected on establishments depended on whether or not the establishment was currently active. For those establishments not currently active, the minimum of information (name and location) was collected; for those in operation at the time of the Census, the Crew Leader attempted to obtain much more detailed information (sector, type of activity, number of employees, etc.).

When defining the file structure for this information, the CDPU attempted to maintain compatibility with the HPC data file to the extent possible. Of course, given the difference in content, the only point of compatibility possible was in the geographic identification. However, this minimal level of compatibility was maintained, so that the Establishment Census data file and the HPC data file can be linked, if necessary, to the level of Crew Leader area (the first four digits of the EA code). In addition, the record type codes in the EC data file are distinct from those used in the HPC data file, so that if the files from the two Censuses were to be combined for any reason, the data for each Census would remain distinguishable.

## 3. PRE-ENTRY MANUAL PROCESSING

Upon arrival at the Central Office, the EC books were checked in manually, and were passed to the clerks in the Manual Edit and Coding Section for review and preparation for data entry. In reviewing the books, clerks were requested to check first for completeness of the geocoding information on the cover of the book, and to correct and/or complete the data required to identify the establishments geographically. The clerk next checked the listings of individual establishments for completeness of information, and where information was missing (for an active establishment), were instructed to attempt to reach the establishment (using either the telephone number already recorded in the Census book or using directory services to find the telephone number) and complete the missing information. Unfortunately, these instructions were not always carried out in a consistent fashion, with the result that in the final data file, a number of data items have a high incidence of cases in the 'Not Reported' category.

Once the clerk was satisfied that information for an establishment was as complete as possible, s/he passed to the next phase of "translating" the information on the establishment's economic activity (primary and secondary), as recorded by the Crew Leader, into a numeric code. (The coding scheme adopted was the International Standard Industrial Classification [ISIC]). The subject-matter analysts had decided to capture this information in as much detail as possible, and as a result, the ISIC codes were used at the four-digit level.

After reviewing the information for each establishment, the clerk counted the number of establishments recorded in the book and verified that the number recorded on the cover was correct. The book then passed to data entry.

#### 4. DATA ENTRY AND VERIFICATION

The CDPU developed a set of data entry screens to control the entry of information from the EC books. After entering the geographic identification codes from the cover, the keyer entered the number of establishments in the book on a separate screen, after which the information for each establishment was entered, using the second data entry screen in the same program. On the assumption that establishments in a given book would be numbered from 1 to  $n$ , the establishment data screen had been designed to enter the sequence number automatically. This caused some problems during keying, as the manual editing of this number had not been carried out in a consistent manner, and there were often gaps or overlaps in the sequence of numbers in a given book. In addition, on occasion an editor would staple together several books sharing the same geographic codes and treat them as if they were one book. This would also cause discrepancies in the sequence checking of numbers by the data entry program. These discrepancies did not result in fatal errors, but they did create some difficulties in verifying the correct number of establishments in a book, and complicated the process of manual editing which followed the data entry operation, as well as the much-later process of adding names to establishment data [see Sections *IV.E.6* and *IV.E.10*, below, for details]. Entry of EC data began 1 January, and was completed within approximately one month.

#### 5. OPERATIONAL CONTROL

The basic unit of control for data entry and verification and for manual editing and correction was the Crew Leader [CL] area. The form which was used to monitor the movement of books and estimate progress allowed for the maximum possible number of CL areas within a given AS area. If all CL areas within the AS area were not received at the same time, a separate form would be used for each set of books from a given AS area. Over the course of processing, the forms would accumulate in the control book, and would eventually show the AS area as complete (or not, as the case might be—in which case the Data Controller would need to track down the missing CL area(s) before the AS area could be considered completed). Progress was monitored on a daily basis: at the end of his shift each day, the Data Controller was required to present summary statistics to the CDPU Manager, who then prepared the daily report for Census Project management. Progress was measured by number of books keyed, verified, edited, and corrected.

#### 6. VALIDATION OF DATA (MANUAL)

The subject-matter staff had prepared specifications regarding the checks to be carried out on the data. In transforming these specifications into a computer program, the DPA added some validations that were needed to ensure consistency in the final data file. In theory, the validity tests would produce a listing of error messages, which an editor would review. The editor would then mark the appropriate correction on the listing, and a computer clerk would apply the corrections to the data file, and run the edit program once again. If the corrections had been properly specified, the second run of the edit program would produce a listing with no errors.

In practice, however, many errors persisted in the data files even after two or three rounds of correction, simply because individual editors chose to accept the information specified (even though the data were in conflict with the specifications) and because the individuals supervising this editing activity did not consistently review the work of the persons under their responsibility. Had they done so, many of the errors which were "let go" by the editors would not have escaped correction. In addition, a number of the edit specifications were modified over the period of processing Establishment Census data, with the result that some cases [combinations of data] were accepted as correct if processed earlier in time, while the same cases were considered incorrect when processed later in time. All of these factors led to some inconsistencies in the data and affected the overall quality of the final data file. This operation began in mid-January and required approximately six weeks to complete.

#### 7. FINAL PRE-TABULATION COMPUTER VALIDATION

Once all the batches of data had been passed through the manual validation process, the files were consolidated into wilayat- or region-level files, depending on size. A final consistency edit program was executed against each of these files. The program was developed from the specifications for the manual edit, and was designed to ensure that the data file would produce tabulations as consistent as possible, given the nature and quality of the data collected. The specifications, as prepared originally by the subject-matter staff, were concerned only with **detecting** invalid or inconsistent responses, but left the matter of correction up to the individual editor. The final computer edit program could not depend on manual checking, so corrections were built in to the program logic. As far as possible (number of employees, for example), the original responses were used to determine the correction, but in the majority of cases, invalid or inconsistent responses were replaced with a "not reported" code, for lack of information that could provide a more meaningful alternative. The various iterations of the computer edit, and the creation of back-up copies of the files, were completed by the end of February.

## 8. TABULATION

Preliminary specifications for tabulations had been prepared in late 1993, but for the most part, these specifications were lacking certain details which were critical to the preparation of the computer programs for generating these tables. These details had been omitted largely because the subject-matter staff were unsure of (a) what quality of data would be collected and (b) what they actually wanted to see in the tables. However, to avoid too great a delay in producing the tables, CDPU staff had proceeded, as far as possible, to create the programs and table format files needed, although the missing details meant that programs and format files could not be developed beyond the skeleton stage.

Once the data had been keyed and edited, the subject-matter staff began deciding on the level of detail in the tabulations, with the help of the CDPU staff, which provided multiple alternative versions of table layouts among which to choose. Once a table layout had been determined, even provisionally, the CDPU staff moved to complete the tabulation program and format file. However, further complications arose from the decision to classify all active establishments by their four-digit ISIC codes. The use of codes to such a level of detail implied a number of problems. One was the sheer volume of text (in English for the left-hand stubs, in Arabic for the right-hand stubs) required to describe activities in enough detail to differentiate one from another in the same major group; another was the number of cells required for a given table, when cross-classifying establishments by one or more other variables. The combination of these two factors caused most of the publication tables to exceed the capacity of the table-generating software (the CENTS module of IMPS), and made it necessary to split most of the tables into four logical parts for purposes of processing. The split was transparent to the user, but made the programming of the tables more complicated. Eventually the programs and format files were completed, and the tabulations generated. Most of the tabulations were produced and sent to the user by early April. Subject-matter staff review and revision made it necessary to re-generate the tables a number of times, but final versions were completed by June.

## 9. SUBFILE EXTRACTION OF ACTIVE ESTABLISHMENTS

During the pre-enumeration listing activity, when information on establishments was captured, each entity listed was classified as either active, temporarily closed, or inactive. The complete data file included all establishments, regardless of activity status, and many non-commercial entities, as well. The subject-matter staff were interested in only active commercial establishments, so a sub-file of such establishments was drawn from the master file and used in all subsequent processing.

## 10. ADDITION OF NAMES TO SUBFILE OF ACTIVE ESTABLISHMENTS

Once the sub-file of active commercial establishments had been generated, it was converted to data base format, with one subfile for each wilayat. A program was then created (using the dBase IV software package) to permit adding the name (in Arabic) to the record for each establishment in the file. Clerical staff from the Ministry were assigned to the task of entering the information from the Establishment Census books. The program displayed on screen the geographic codes, the line number, and other key information for each establishment, and the clerk would verify that the information on the screen matched the information in the book at the line number specified. When the clerk was satisfied that the correct establishment had been identified, s/he keyed in the name of the establishment (in Arabic). The name then became part

of the information in the data base. Once all names had been entered, the file was written out in ASCII format, so that it could be used with the IMPS software when necessary.

#### 11. MERGE OF MODIFIED SUBFILES

Once the names had been inserted into the wilayat subfile, wilayat files were recombined into their respective region-level files. When all region-level files had been reconstituted, they were merged into a single file for the Sultanate. From this file, a number of subfiles were generated to meet specific user requests.

#### 12. CREATING SECTOR FIELD FOR ACTIVE ESTABLISHMENTS

The information on economic activity (primary and, where available, secondary) for each establishment was recorded as a four-digit code. The tabulation scheme used the numeric codes at the four-, three-, and two-digit levels, but when totals were required at the one-digit level (major economic activity groups), alphabetic letter codes (e.g., A, B, C, etc.) were specified instead. This letter code was not part of the original data keyed, and had to be re-calculated for each tabulation. To eliminate the need for constant re-calculation, CDPU staff wrote a program to insert the letter code of the major group associated with the **primary** economic activity (secondary economic activity, where available, was disregarded). Thus, the final EC data file was comprised of all data originally keyed, plus the name and the letter code for major economic activity group, for all economically active establishments in the Sultanate.

### F. Housing and Population Census data

The major activity undertaken by the CDPU during its existence was, of course, the processing of data from the HPC. The data collected during the ten days (1–10 December 1993) of enumeration were recorded in books of forms. In any book, each set of facing pages contained blocks of columns and rows in which information could be recorded about a particular housing unit [HU], household [HH], or person. [*Attachment W.F1 shows a sample set of pages; Attachment W.F2 is a descriptive list of the data collected.*]

The top part of the right-hand page of each set contained blocks for noting the type and occupancy characteristics of a given HU. If the HU was unoccupied, the remainder of the two pages was blank. If the HU was occupied, data were collected on the characteristics of the household (top of left-hand page) and of the individual(s) in the household (lower section of both pages). In addition, for certain types of HUs, characteristics of the HU were also recorded.

The cover of each book, and the first inside page, also contained critical information. The cover was designed to show clearly all geographic codes, from region through EA, the book number and total number of books used in the EA, and the numbers of HUs, HHs, and persons, by sex and nationality, recorded in the book. The first inside page was identical in format to the cover, and was perforated for ease of removal.

#### 1. SYSTEM DEVELOPMENT

All modules for this subsystem were developed using the IMPS software. To ensure standardization, the DPA created a data dictionary containing descriptors for all record types (summary form, book cover, housing unit, household, and population) associated with these data. The descriptor for each record type included detailed descriptions of all individual data items and permitted specification of valid values for each item. For selected items, group definitions were made available to facilitate editing and tabulation.

##### a. Data entry

Using the data dictionary, the CDPU staff developed separate CENTRY modules for use by keying group supervisors, by data entry keyers, and by key verifiers. The module for the supervisor permitted entry of geographic identification codes to initiate the batch (book cover record). The module for data entry permitted keying of data from housing unit, household, and population information only. The module for verification permitted access to all record types (except summary form data, which was not relevant to this operation).

b. Editing

The edit specifications prepared by the Census Advisor contained a number of checks which were to be resolved by editor review (that is, manual rather than automatic correction). These checks, along with others specified by the DPA, were combined in one CONCOR edit program which was to be executed against each keyed and verified batch. Corrections were to be applied manually on individual records.

A second CONCOR edit program was developed from the specifications, and included all checks specified in the program for manual correction, as well as all checks scheduled for correction by computer. This program went through several iterations, as the specifications were modified and refined to arrive at the final version. A number of smaller CONCOR programs were written to meet requirements for specific processing situations or problems involving subsets of data. No formal specifications were provided for these programs, since they were usually written to solve a particular problem which required immediate attention.

c. Tabulation

Detailed tabulation of the data was unquestionably the biggest single task faced by the CDPU professional staff. In theory, the steps required to produce a table for publication are:

- (a) The subject-matter staff must prepare, on paper or in a computer-readable medium, a drawing or layout of each table, showing clearly the title, column headings, row descriptors, footnotes, and any other text or design elements required (vertical/horizontal rules, typeface [italic, bold], etc.). The layout must also indicate, for each cell in the table, the conditions for tallying into the cell or the calculation required to place a value in the cell.
- (b) The programmer must then reproduce, using the software selected, the layout of the table, with all text in its proper form and place, and with the display format (i.e., with or without decimals or separators) for each cell specified. In addition, the programmer must determine the number and kind of calculations (e.g., row and column totals and/or subtotals, percentages, etc.) to produce the results required by the subject-matter staff. The programmer must then submit the layout to the subject-matter staff for their approval. If changes other than "cosmetic" are required, the programmer must modify the layout and re-submit for approval before going on to the next step. If changes are required only to the physical appearance of the table, the programmer may begin the next step while awaiting approval of any modifications made.
- (c) Once the subject-matter staff have approved the structure of the table, the programmer must then write the computer instructions to generate the proper values in each cell of the table. This includes all instructions for selection of universe (the population to be tabulated—e.g., all Omanis, all non-Omanis, all ever-married females, all persons 15 years of age and older, etc.); all recoding (rescaling or regrouping) of variables to determine appropriate row/column coordinates; and the proper sequence of instructions for generating the table correctly at the geographical levels desired (e.g., region, wilayat, urban/rural, etc.).
- (d) Once the program is complete, the programmer must then test the selection and tabulation logic, as well as the calculations required to arrive at the final result. The numbers in each cell must be checked, as far as possible, against control totals, and tables with similar characteristics must be checked against each other to ensure consistency and correctness of results. Any discrepancies or errors noted must be investigated and resolved. When the programmer is satisfied that the numbers are valid, the table must be sent to the subject-matter staff for final approval. Any modifications or corrections requested by the user must be incorporated into the table layout and/or the tabulation program, and the cycle of testing and approval begins again.

The subject-matter staff prepared the index of tables to be generated and made it available to the CDPU by mid-1993, but many of the individual table layouts required were not available for many months. With very few exceptions, CDPU staff developed all programs for generating the publication tabulations, with assistance from the DPA when required, using the CENTS subsystem. (The

exceptions were generated from the Geographic Data Base using the FoxPro software.) Because of the unusually large number of tabulations specified by the users, the DPA attempted to make the task of producing these tables easier to control by creating table groups—that is, taking into account system limitations and table subject matter to combine tables into "clusters" which would tax the capabilities of neither the system nor the programmer to whom the task of creating the programs was assigned. Each cluster (group of tables) was assigned a specific and unique name, following the rules for identifying table groups.

Once the tables in each cluster had been determined, the programmer responsible for the cluster created the layout (with formatted cells, calculations, and English and Arabic text) for each table in the group. Many tables had to be subdivided into two or more parts, for purposes of tabulation, because of the large number of columns and/or rows involved. [All such layout files are identified with the extension .FMT; the name reflects the subject group—housing, education, fertility, etc.—of the majority of tables included in the cluster.]

After the .FMT file had been established, defining the number of rows and columns in each table, the programmer then wrote the instructions to generate the cell values and the appropriate geographic area control breaks from the Census data file. [All such programs use the same name as the .FMT file and the extension .TAB.] Needless to say, both .FMT and .TAB files went through many iterations as the user modified layouts, tabulation universe, etc.

Two types of auxiliary files were also used to generate the tables: with extension .CON and with extension .ARA. The former contained parameters indicating to the CENTS subsystem the levels (Sultanate, urban/rural, region, wilayat, wilayat centre, etc.) at which totals would be required. Since the same file could be used with many table clusters and only a few .CON files were required, the name assigned to a .CON file did not link it with a specific .FMT/.TAB combination. Similarly, the latter [.ARA files] contained the names (in Arabic and English) associated with various levels of publication geography, and, like the .CON files, were also not linked, logically or by name, to specific .FMT/.TAB files. [When the text and code values of an .ARA file had been stabilized, the CENTS subsystem generated a file with extension .ANF to be used in the printing of the final tabulations. Such files have an "internal" format and cannot be edited or viewed. Any modifications to area names or codes must be carried out on the .ARA file, after which the .ANF file must be regenerated.]

## 2. HPC DATA FILE STRUCTURE

The first requirement, fundamental to any further system development, was the definition of file structures for HPC data. The IMPS software used required that the Census data file be non-hierarchical in nature—i.e., it imposed a "flat," or sequential ASCII, file format. The conceptual approach to structuring the data file required that all items of information relating to a particular Census data element (i.e., HU, HH, Summary Form, etc.) be recorded in separate record types. To permit eventual joining of different data files (should it ever be desirable), each record type was given a different type code, and the placement of geographic identification fields was standardized over all record types. (It should be noted that Establishment Census [EC] data are excepted from this rule: while the geographic identification codes, from region down to EA, occupy the same positions in the data record as they do in the other data files, the record type code is in a different position, because the EC data file was designed with reference to an early prototype of the HPC data file, and went into production before HPC data file structure was made final. However, if necessary, the EC data may be manually converted to conform to the HPC data file structure.)

The keyed HPC Census data file would contain the following records:

- Book Cover record (one for each Census Book)
- Housing Unit data record (one for each HU)
- Household data record (one for each HH)
- Population data record (one for each person)

### 3. SAMPLE SELECTION AND WEIGHT CALCULATION

As mentioned earlier [see Section III.B.6.b above], MoD management required the CDPU to produce tabulations based on a sample of the data collected. The DPA felt that a total population size of approximately two million was much too small to justify the additional effort required to process a sample, but management insisted that it needed estimates at the Sultanate and regional levels before the date at which the numbers would become available through normal processing. Once it became clear that sample processing would be required, the most immediate and critical task became the selection of the sample. Since priority in **all** phases of processing would have to be given to data from sample units, these units would have to be identified before any manual or computer processing could begin.

At the time the DPA was required to define the sample, there were only two possible sources of information on which to draw: the Geographic Data Base [GDB] and the Summary Forms (which were known to contain many errors in identification codes and in other fields). For this reason, the GDB was chosen as the universe from which the sample would be drawn. Since EAs had been designed approximately equal in size, they could serve as the sampling unit. A subfile (one for each region) containing one record for each EA was generated from the GDB. Each record contained the region, wilayat, and EA codes; within each region file, the records were ordered by EA within each wilayat. The DPA wrote a program to pick every tenth EA, beginning from a random start (generated by the program). This program was used on all regions in the Sultanate **except** the Governorate of Musandam and Al-Wusta Region. As mentioned earlier, the total population size of each of these was so small that sampling would save very little time and would undoubtedly introduce distortions.

In five of the six regions for which a sample was generated, every wilayat was represented by at least one EA. In the Governorate of Dhofar, however, the random start generated by the program left two wilayats (of nine in the Governorate) unrepresented in the final sample. One EA was selected at random from each of these two wilayats so that the final sample would contain data from each wilayat in the Sultanate. In this fashion, the "ten percent sample" actually contained slightly more than ten percent of the total data, but the weights were adjusted appropriately.

Ideally, weights would be derived with greatest precision using sample data and full data. However, full data would be available only at the end of the manual processing cycle, which was much too late to be used for calculating sample weights. The next best alternative was the Summary Forms, **after** being corrected, re-keyed, and re-tabulated, and this was the option chosen. Since weights would be required only at the time of tabulation of the sample data (estimated to be mid-May), it was possible to begin manual processing of the sample data without having completed the correction of the Summary Forms.

Once the sample data had passed the final edit and correction phase, weights were calculated by first generating from the corrected Summary Forms the totals, by region and wilayat, of HUs, HHs, and persons by nationality (Omani/non-Omani) and sex. Next, the sample data files were tabulated in precisely the same fashion, and the totals from the Summary Forms were divided by the totals from the sample data to provide a weight, calculated to three decimal places, for each sample element (i.e., HU, HH, or person) within each wilayat. The appropriate weight (in the case of Al-Wusta Region and the Governorate of Musandam, 1.000) was then assigned to each record in the sample data file, which was then ready for tabulation. A subset of tables from the full tabulation list was produced from these data, and was submitted to the subject-matter staff for analysis and publication. The Census Advisor prepared an analytical report, but neither the report nor the tables were released to the public.

### 4. PRE-ENTRY MANUAL PROCESSING

From the field, Census Books were returned through the field hierarchy—that is, each Enumerator returned the completed Books for the EA to his/her Crew Leader, who in turn sent all Books for EAs in his/her jurisdiction to the appropriate Supervisor. At this point, books were packed (grouped by CL area) in custom-made safeboxes, with padlocks to ensure security. When all CL areas had reported in, the safeboxes were locked and sealed, and brought in to the Central Office for temporary storage. In the Central Office, the Books were prepared for the data entry operation in phases, as follows:

a. Initial check-in and review

Each safebox was unsealed in the presence of the Supervisor and Central Office staff. Books were reviewed by CL area—that is, all books for EAs within one CL area were reviewed before passing on to another CL area. When one CL area had been completed, the clerks passed to another in the same Supervisor area. Review consisted of checking the following:

- the geographic codes (particularly the locality codes), which were verified using a list generated from the Geographic Data Base as it existed at that time; and
- the total number of books for the EA, which included verifying that all books for an EA were present, that no book number was duplicated within an EA, and that the "Book Number" and "Total Number of Books" fields on each Book Cover were accurate.

As CL areas were verified, they were passed to the Central Stores area, where clerks began the process of registering the Books into the Book Control Data Base (BCDB). However, because the time required to enter information into the BCDB was much greater than the time required for visual verification of Book covers, the two operations did not operate precisely in parallel—there was often either a backlog of Books to be entered into the data base, or the BCDB staff waited for another supply of Books from the check-in clerks.

b. Re-creation of Summary Forms

Once the Books for an EA had been checked for correct codes and totals, the clerk was required to re-create the Summary Form. As was mentioned earlier [*see Section II.D, above*], errors in the original Summary Forms caused problems during the processing. For this reason, clerks assigned to re-create the Forms were instructed to carry out the following steps:

- re-verify the geographic codes (particularly the locality codes), using the list generated from the latest version of the Geographic Data Base;
- re-verify the total number of books for the EA, which included verifying that all books for an EA were present, that no book number was duplicated within an EA, that all book numbers were in an uninterrupted sequence from *n* to the "Total Number of Books," and that the "Book Number" and "Total Number of Books" fields on each Book Cover were accurate;
- verify the total counts as noted on the Book Cover [total number of HUs, total number of HHs, total number of persons (M/F/T), total number of Omanis (M/F/T), and total number of non-Omanis (M/F/T)]. To verify these counts, each clerk was required to:
  - (a) re-count persons listed in each HH by sex and nationality, compare the totals with the totals listed for that HH, and correct the listed totals if necessary;
  - (b) re-count the number of HHs listed in the Book;
  - (c) re-count the numbers of HUs listed in the Book;
  - (d) add population counts for each HH to find the total Book population count;
  - (e) compare the re-counted totals with the totals recorded on the Book Cover; and
  - (f) investigate any differences and correct the Book Cover totals, if necessary.

However, the rate of error in the re-created Summary Forms was almost as high the second time as the first, which, on the evidence of the types of errors found (invalid geographic codes and simple arithmetic mistakes), was due more to haste and carelessness on the part of the clerks who prepared the second round of Forms than to any unreliability of the data themselves.

c. Registration in Book Control Data Base

After completing work on the Summary Forms, clerks filed the Census Books in special containers, sized to permit storage of all Books for one CL area in the same container. There were, of course, CL areas which required two or more containers, but the majority of CL areas occupied one container only. Each container was marked with a unique number, which identified its location in the storage room by unit, shelf, and relative position on the shelf.

As each container was filled, it was checked into the Book Control Data Base. Using a program written by the contractor who created the system, the clerk entered the container's locator number, and then the geographic codes from the cover of each Book in the container. When entry was completed for a given container, the system generated an adhesive-backed label listing the geographic codes of all the Books in the container. This label was sized to fit on the narrow face of the container, so that it could be read when the container was stored on the shelf. The label was then applied to the container, and the container placed on the shelf in its correct position. When a sufficient number of CL areas had been checked in, the manual editing and coding began.

d. Manual editing and coding

During the phases of processing which followed registration of the Census Books in the data base, the basic unit for work assignment and operational control was the CL area or container. The subject-matter staff had created teams of editors and coders, who would review the Census Books for processability and would carry out the translation of written responses (e.g., nationality, occupation, educational level, etc.) into numeric codes acceptable to the computer. Some of the editors/coders had been enumerators during the field operation, but most were hired for this specific task.

The editing and coding phases were separate—that is, first an editor carried out the preliminary check of all the Books in a container, after which the container was passed to a coder to complete the preparation of Books for data entry. "Editing" involved reviewing all information recorded by the Enumerator, checking that the numbering of HUs, HHs, and of persons within HHs was sequential, checking that the numbers of persons in each HH and in each Book was reflected on the Book Cover, checking that occupancy status of a HU was accurate, etc. "Coding" involved translating responses to questions on nationality, residence, education, and employment into numeric codes which could be entered into and processed by the computer.

The objective of this activity was to make the Census Books ready, in a uniform fashion, for the data entry operation. Unfortunately, this objective was not always met, due largely to differences of understanding among the editors and coders with respect to procedures. Manual procedures are **always** subject to inconsistency in execution, simply because no two persons understand or interpret instructions in precisely the same fashion, no two persons will carry out the same set of instructions with the same degree of diligence and attention to detail, and even the same individual will carry out the same instructions differently, depending on his/her physical and mental condition at the time. Once the data entry and manual editing operations began, these differences in the level of quality of the work done by different editors and coders became very clear, as will be explained in the sections below.

Processing of Census data, in all phases, was carried out in parallel and overlapping activities, on a "flow" basis. That is, manual editing and coding was begun with the EAs from the first region in the sample; when all sample EAs from the first region had been sent for keying, the editing/coding staff began work on sample EAs from the second region to be processed, etc. When all sample EAs had been prepared for keying, work began on EAs in the first non-sample region, and so on.

**5. ENTRY OF CENSUS BOOKS**

Keying of the Census Books (which included 100% verification, or full re-keying) began during the last week of February 1994, immediately upon completion of entry of Establishment Census books and the revised Summary Forms. In the same fashion as the manual editing/coding, the keying activity began with the sample EAs from the first region, and when those Books had been completely keyed and verified, the

data entry staff moved on to the EAs from the second region, and so on. As work was completed on the sample data (including full data for Musandam and Al-Wusta), the remaining Population Census Books entered the workstream. All keying and verification were completed by the end of July 1995.

## **6. OPERATIONAL CONTROL**

The task of monitoring the flow of documents into and out of the keying phase, and of monitoring the data files as they passed from keying through manual editing and correction, was the responsibility of the Data Controllers. Two separate control forms were developed for this phase of the processing, and were employed (a) to track work in and out of the CDPU and (b) to provide management with daily production totals. In spite of occasional lapses in carrying out the procedures, the Data Controllers were critical to the success of the Census DP operation. When different activities for different phases were being carried out simultaneously, the operational control staff were able to keep the separate workstreams flowing and production on schedule.

## **7. VALIDATION OF DATA (MANUAL)**

The same procedures were followed in the edit with manual correction: editing staff concentrated on completing work on the sample EAs in the first region before moving on to the EAs in the second region. All operations up to this point were carried out on individual book files. Based on the section of the edit specifications (prepared by the Census Advisor) labeled "Check Only," the DPA prepared a CONCOR program [BATCHK.CN] which generated error messages for each condition of inconsistency identified in the data file. The program was run on each keyed and verified batch file (i.e., each Census Book), and the error listing and file were passed to an editor, who was required to review the listing, determine the cause of error, and mark the correction on the listing. The listing and data file would then be passed to a keyer, who would make the corrections and re-run the edit program. If errors persisted, the cycle would begin again. Most of the serious problems with the data, discovered during the tabulation phase, can be traced to the failure of editors to correct all problems identified in this phase.

## **8. CONSOLIDATION TO HIGHER-LEVEL DATA FILES**

Once an individual batch file had completed the manual process and been accepted as error-free, the eventual consolidation into higher-level files (first to EA level; then EAs to CL level; and last, CL areas to wilayat level) before the final edit was controlled and carried out by CDPU professional staff. Separate forms were used for the various levels of consolidation.

## **9. FINAL PRE-TABULATION COMPUTER VALIDATION**

When the file for a given wilayat was complete (that is, all books in all EAs in the wilayat had been edited, accepted, and consolidated to wilayat level), the file was run through the final consistency edit, and corrections were applied by the computer according to rules previously established by the subject-matter specialists. The DPA wrote another CONCOR program [BOOKEDIT.CN] based on the section of the edit specifications (prepared by the Census Advisor) labeled "Check and Correct." Like the first edit program, the second program generated error messages for each condition of inconsistency identified in the data file, but instead of requiring manual intervention, the program itself made corrections based on the specifications and automatically generated a corrected output file. When the file for each wilayat had been edited and corrected, the output files were consolidated into region-level files for preliminary tabulations.

## **10. USER ACCEPTANCE OF DATA**

Once the preliminary tabulations had been carried out on edited data for all regions, the results were presented to Ministry management, who requested further verification of the data before final acceptance. After management accepted the data, the various region-level files were consolidated into a single file for the Sultanate.

## 11. TABULATION

Once the data had been edited and accepted, the CDPU staff generated control tabulations to the level of wilayat, using the automatic tabulation facility provided by the QUICKTAB subsystem of IMPS. These control tabulations provided numbers against which to check the publication tabulations, and thus were used continuously during the final phase of processing.

The formal program of tabulations was carried out over several months, as the subject-matter staff reviewed and revised the tables generated by the CDPU staff. The first four tables to be produced were based on the GDB, which had been modified to contain summary totals (of HUs, HHs, and population), as well as the codes and names of geographic entities. These tables provided summary counts of localities:

- regions in descending order by population size, and wilayats within each region in descending order by population size;
- ☒ regions and wilayats in ascending order by geographical code, and localities within each wilayat in descending order by population size;
- ☒ localities within the Sultanate, in ascending (alphabetic) order by Arabic name; and
- ☒ localities within the Sultanate, in descending order of population size.

Immediately following completion of the first four tabulations, the CDPU staff began to generate the remaining tabulations for publication. *[For a list of table titles (in English) of the tabulations requested, see Attachment B.F.11, below.]* These tables (several hundred in number, when geographic variations are considered) were produced over a period of several months, as the subject-matter staff reviewed each table and made revisions in table text and/or content. As of the date of this publication [July 1995], table production continues.

## 12. UNSCHEDULED REQUESTS FOR DATA AND TABULATIONS

From the moment the summary results were announced at the end of 1993, MoD officials were inundated with requests for data files and/or tabulations. Such requests had to be postponed until the Census data themselves were available (September 1994), but from that date on, CDPU staff were kept busy satisfying the demand for information from users both within and outside the MoD.

## V. CONSIDERATIONS FOR FUTURE CENSUSES

The experience of the 1993 Census of Population, Housing, and Establishment offered multiple object lessons in what to do and what not to do in future operations of this type. Many of the decisions made in matters not obviously related to data processing had consequences which ultimately affected the processing operations, often in a negative manner. The following sections discuss some of the ways in which the post-enumeration activities might be improved, from the viewpoint of the CDPU.

### A. Planning for processing

It is important to distinguish between the kinds of processing which are usually required in a Census operation. First (and most important) is the processing of the data collected during the Census enumeration. These data will form the basis of the publications, data bases, reports, etc., which are usually developed as products of the enumeration. Second, and equally important to the overall success and accuracy of the Census, is processing of what may be called "auxiliary" data—the various information systems which permit Census project management to monitor and control the flow of operations, both before and after the enumeration itself.

#### 1. PROCESSING OF DATA COLLECTED DURING ENUMERATION

The systems which are developed to process the Census data themselves are relatively standard, although the media used may vary. Since a Census by definition involves the entire population, and since the ultimate objective of the Census data is to provide information on which government plans and policies may be reliably based, the processing system must be designed to keep errors to a minimum while maximizing efficiency in handling the huge volumes of data collected. First, the data must be transferred from the forms in which they were collected into a computer-readable medium. (In the case of the Omani Census, this was done in the most traditional manner: key data entry and key verification. Other options include use of optical mark readers, or even direct entry by the enumerator using a portable microcomputer.)

Next, the data must be checked for validity and consistency, and errors must be corrected. When the users (i.e., the subject-matter specialists) are satisfied that the data are of acceptable quality, tabulations are produced. While some operations might be combined (e.g., validity checks with data entry), and methodology may vary, Census data processing will invariably involve the same basic functions: data preparation and conversion, data validation and correction, and data tabulation. Once the hardware and software have been specified, and the questionnaire has been made final, the analysts and programmers may begin design and development of the modules which will be used for processing.

With respect to the Omani Census, one of the problems which affected the CDPU's operations was the delay in approving the Census questionnaire in its final form. Until the form was accepted by the CTC, the subject-matter specialists made no effort to define the codes to be used for the questions with "open-ended" responses (i.e., those questions where the response is written on the form by the enumerator and must later be "translated" into a numeric code during the manual processing phase prior to data entry). Without a list of the specific codes to be permitted, it was not possible to complete development of the data dictionary, the data entry module, the edit module, or any tabulation modules which used those variables.

For the next census, it will be critical for the planners to arrive at a final questionnaire design and to define validation procedures in a timely fashion, to avoid **unnecessary** last-minute pressures on the CDPU and on the subject-matter specialists. Computer systems which are developed "under the gun" (i.e., at the last minute) are never as efficient or as trouble-free as systems which are produced with sufficient time for good design and thorough testing before they are implemented.

#### 2. SYSTEMS FOR MANAGEMENT CONTROL

Equally important to the ultimate goal of producing reliable information from the Census data in a timely manner are the systems which provide management with information about the Census operation. These systems can be manual, computerized, or both, but the principal characteristic is that they permit management to monitor progress on a periodic basis. These systems will vary more in kind and quality than systems for processing Census data, but in general they will permit management to track the following:

- pre-enumeration operations (e.g., listing of households or buildings in each EA)
- enumeration (e.g., daily output from enumerators)
- receipt of Census questionnaires at central office
- progress of Census questionnaires through pre-keying phases
- progress of Census questionnaires through keying and verification
- consolidation of data from lower geographic levels to higher levels, prior to editing
- editing and correction of data, and review of edit error listings
- consolidation of data to tabulation-level files
- tabulation of data at all required levels
- progress in any of the above areas, as measured against a time schedule approved prior to the start of operations and accepted by all parties

In the case of Oman's Census operation, systems to provide information on post-enumeration operations were defined, and developed, prior to the enumeration. The Census Advisor devised a manual system for receipt and check-in of Census books returning from the field to the central office, and the staff was instructed in procedures to be followed and use of the appropriate control forms. However, at literally the last minute (the eve of enumeration), a computerized system was decided on, and an outside contractor was called in to provide such a system. In order to provide the contractor with the information needed to create the system, key personnel had to be diverted from more important tasks during the most critical phase of the project. Inevitably, the system which was developed in such haste was bug-ridden when first installed, and required many iterations of repair and re-installation before it performed to specifications. In addition, in contrast with the manual system developed earlier, personnel could not be trained in use of the system until it was fully installed and functioning, which implied a further delay in registering the receipt of the Census Books. On occasion, "bugs" in the system caused information to be incorrectly registered in the data base, forcing check-in staff to spend valuable time re-entering the information—an unnecessary duplication of effort and therefore a reduction in productivity.

As for the pre-enumeration activities, and the enumeration itself, no indication had been given that any sort of computerized information system would be required. However, just as with the system to receive Census Books, a decision was made, only days before beginning field work, to request a system to track the progress of the Crew Leaders in numbering and listing buildings and households, and a system to track the progress of Enumerators in covering their assigned areas, both on a daily basis. By the time this decision was made, time for implementation was extremely limited, so responsibility for the systems was divided between the CDPU (pre-enumeration) and the Computer Department of the MoD (enumeration). Once again, resources which were needed for other tasks had to be diverted to meet a last-minute demand which could reasonably have been foreseen earlier. Only the exceptional dedication and skill of the staff involved in developing these systems made it possible to implement them within the time frame allowed, given the very minimal specifications on data to be received and entered, and the total lack of specification of how the data were to be tabulated and used. In addition, no apparent thought had been given to integrating these data into the overall system nor of using them to validate subsequent processing operations.

However, if planning is properly done, with sufficient anticipation (including time to train users in system operation), and with careful thought as to the eventual utility of the data, staff will not be required to make superhuman efforts to complete, by an arbitrary deadline, a system which has only a very limited utility.

**For the next census**, it is recommended that Census Project management define information requirements well ahead of the date at which they will be needed, to permit both developers and users time to ensure that the correct information will be captured, generated, and used in the proper fashion.

### 3. STAFFING

As with the information systems, the data entry staffing operation was also subject to last-minute decisions. After having stated that all keyers would be hired through local third-party contractors (who were free to employ expatriates), management reversed the decision and decreed that half of the keyers must be hired directly by the Census Project and must be Omani nationals. This decision placed an enormous additional burden on the CDPU staff, at a time when all resources were concentrated on preparing the systems needed for initial processing of the Census data. In addition, neither space nor furniture for the keying

operation had yet been allocated, so the only machines available on which candidates could be tested were those computers in use by the CDPU staff itself. Since the number of candidates to be tested was extremely large (almost 200), all other work in the CDPU was effectively suspended during the weeks in which candidates were tested. Once again, only extraordinary efforts on the part of the CDPU staff permitted them to overcome this obstacle and continue system development more or less on schedule.

It is difficult to overstate the disruption and difficulties caused by this particular decision: the Census project was fortunate to have data processing staff with the skill and determination to make things work in spite of negative circumstances. In the future, circumstances may be different, and a similarly hasty and unconsidered decision could have disastrous results on the Census timetable.

**For the next Census**, it is strongly recommended that all decisions relative to the hiring of personnel be made early enough to permit normal recruitment, evaluation, and hiring procedures to be carried out without demanding extraordinary efforts on the part of the staff or the interruption of other scheduled work.

#### **4. DATA ENTRY ENVIRONMENT**

Some weeks before keying began, management decided that the data entry operation would be carried out in the space occupied by the library. It was assumed that existing library staff and equipment would move to the new building in time for the space to be modified for use by keyers. However, this did not occur: delays in completing construction of the new building had a ripple effect on all activities related to the library space (including installation of new electrical circuits to permit large numbers of microcomputers to work safely), with the result that very late on the night before data entry was to begin, CDPU staff were still trying to set up enough tables and chairs to permit keyers to work effectively.

This last-minute rush, and the attendant stress, was unnecessary: if there was any doubt that the space would be available at least one week before start of keying, Census Project management should have had a contingency plan (i.e., an alternative keying site), which could be adopted if the library was still unusable by the deadline. This would have permitted the installation of equipment and software with time for testing and adjustment before entering into full-scale production.

**For the next Census**, it is recommended that the physical space and furnishings needed for each operation be made available early enough that adjustments and modifications can be made and still meet the deadline for beginning operations. A sufficient number of stressful situations inevitably arise in even a well-run Census operation; all concerned should avoid creating any more through negligence or poor decision-making.

### **B. Processing**

Once the pre-enumeration obstacles had been overcome, and the first data (Summary Forms) moved through the system, processing became a series of routine operations. Census Books were checked into the data base, then assigned to the manual editing and coding operations, following which they were keyed and verified. Each Census Book passed through a computer edit with manual corrections, after which all batches were ultimately consolidated into wilayat-level files for the final edit. Finally, data were tabulated according to the specifications established by the subject-matter specialists. The problems which occurred can be generally attributed to human error, inadequate specifications, and/or poor decisions.

#### **1. CHECK-IN OF CENSUS BOOKS AND VALIDATION OF GEOGRAPHIC CODES**

The difficulties and confusion caused by incorrect locality codes, and even the occasional wrong region and/or wilayat codes, were the single most pervasive error in the entire post-enumeration processing phase. The errors in region and wilayat codes were basically attributable to carelessness in checking the lists provided on the back of each Census Book. The errors in locality codes can be attributed to mistakes made, singly and in combination, during any or all of the various geography-related operations, from initial mapping to final tabulations. For example,

- (a) the locality may have been incorrectly sited (i. e., physically located) on the map by the cartographer, during field mapping;

- (b) the locality may have been incorrectly sited on the map by the draftsman in the office, after field mapping;
- (c) the locality may have been incorrectly identified (i.e., given the wrong name);
- (d) the Enumerator may have recorded the wrong locality name and/or the wrong locality code on the Census Book cover;
- (e) the locality name and/or code may have been incorrectly modified during the post-enumeration check-in operation at the Central Office;
- (f) the locality name and/or code may have been modified on the Census Book cover, but not in the GDB;
- (g) the locality name and/or code may have been modified in the GDB but not in the data file; etc.

In the rush to complete mapping operations in time to permit demarcation of the Sultanate into EAs, a number of questionable locality identifications were left unverified; during the field operation, when enumerators were supposed to re-verify (and correct where necessary) the names and relative positions of all localities on the map of their EA, many errors remained uncorrected because of the pressures of time or the enumerator's carelessness; during the post-enumeration check-in of Census Books, many of the corrections to locality names specified by enumerators were ignored or lost in the shuffle of Books and papers; during the entire pre- and post-enumeration period, the Census Project received from the local governments (wilayats) and from the Ministry of the Interior a steady stream of "corrections" to locality names and positions, which were unevenly implemented; and corrections made to a locality name and/or code at any step in the process were not always applied uniformly to all required sources. These mistakes accumulated to cause serious problems—still in the process of resolution—with geographical attribution in the final data set.

**For the next Census**, the Census Project should establish a firm rule that no EA be checked in or otherwise processed unless and until the geographic names and codes on the Census Books for that EA, the names and codes on the map for that EA, and the codes in all computerized data base(s) which refer to that EA are in full agreement. This may slow down initial processing, as some EAs are held aside for resolution of differences, but in the long run much time and effort will be saved by guaranteeing a minimal uniformity among all sources of information referring to locality names and codes.

## 2. SAMPLE PROCESSING

Management's decision to require, by mid-June 1994, results based on a sample of the HPC data added a great deal of unnecessary additional work for the CDPU staff. The sample EAs had to be selected, weights had to be calculated, care had to be taken that sample EAs were given priority in all phases of processing, the tabulation programs had to be modified to use weighted data—all these tasks would not have been necessary for normal (non-sample) processing. All of the extra work would not have mattered if the tabulations had finally been used for some purpose; however, after extraordinary efforts to produce the tabulations by the deadline, they were not, in fact, published or used in any meaningful way.

**For the next census**, the use of sample processing should not even be considered. Even with projected increases in population by the year 2003, the volume of data to be expected in the next Census will not justify the additional effort and procedural complications involved in processing a sample. The experience of this first Census has shown that normal (non-sample) processing can be accomplished quickly enough to satisfy user demands for data, particularly if processing is done on a flow basis (i.e., wilayat by wilayat, one region at a time). Sample processing offers no real benefit to any country where normal data entry can be completed within one year, as was the case with Oman.

## 3. MANUAL EDITING

Large numbers of editors were employed to review the information recorded in the Census Books (both HPC and Establishment) and ensure that the information was minimally usable—that is, that it would cause no problems during the data entry operation. The advisors who supervised this operation had devised a set of editing instructions for each type of book, and prior to beginning actual production the editors were trained in following the instructions. However, there was much inconsistency among the editors, with respect to both an individual editor's output from day to day and the work of one editor compared to another; and while the supervisors were always present to resolve doubtful situations, many of the edi-

tors apparently did not bother to ask for advice when a question arose—they simply made a decision based on their "best guess," or they ignored the matter entirely, leaving either blanks or incorrect information for the item in question.

In addition, after the first few weeks, the supervisors apparently did not systematically review the work done by the editors. This was evidenced by the continual appearance, over the entire period of processing, of EC and HPC Books with errors attributable to poor editing, most commonly errors in sequence numbers (of establishments, or housing units and/or households). While certain manual operations can be reduced or eliminated entirely, it is not practical to eliminate the pre-entry phase of manual editing, since it can, **when properly done**, reduce the rate of error in later stages of processing.

**For the next Census**, it is strongly recommended that greater care be taken in the supervision and verification of editors' work. Verification must be systematic, and while it is not required that all work be 100% verified, each editor's work should be subject to some review on a periodic basis. For example, each editor's work might be 100% verified for the first two weeks, or until the editor's error rate drops below a pre-determined level. At that point, the editor's work should be subject to partial verification, on a regular basis, until the error rate rises above the threshold. When this occurs, the editor's work should be once again subject to full verification, until the error rate once again reaches an acceptable level.

Of course, this procedure requires that standards be established (i.e., what constitutes an error, how many errors per unit are acceptable, etc.), and that the work of each editor be continuously monitored. The verification of an editor's work should be independent—that is, it should be done by someone other than the editor her/himself who is familiar with the work, or by persons whose sole function is to verify the work of editors. Forms recording numbers of errors per work unit should be prepared, and the manual editing staff should include one or more quality control clerks to maintain accurate records of the performance of each editor. Editors showing consistently high error rates must be re-trained or dismissed. Many of the errors which slowed the computer editing and tabulation phases were manual editing errors which were not corrected before moving to the next phase.

#### 4. CODING

As mentioned earlier in the document, coding is the translation of written responses into a numeric code, prior to the data capture operation. Virtually everything that was said in the preceding section about the manual editing process could be repeated about coding, with one exception: the coding activity can be eliminated, if only pre-coded questions are used. This is often done when optical mark technology is used, or when questionnaire size is limited, or simply to maximize efficiency in enumeration and processing. However, if "open-ended" questions (i.e., questions requiring coding before data entry) cannot be eliminated from the questionnaire, they should at least be kept to a minimum. In addition, the users must balance the degree of detail required (i.e., how many digits in each code) with the degree of detail captured by the enumerator. In other words, if the enumerator was in the habit of recording only minimal information for occupation, translating minimal information into a four-digit, or even a three-digit, code is a waste of time and effort, because the information available in the questionnaire is not sufficient to discriminate one occupation from another below the two-digit level.

The other factor to be considered when determining the degree of detail in coding schemes is to evaluate the frequency of occurrence of certain categories (i.e., occupations, educational degrees, etc.) in the population. For example, if 90% of the population engaged in agricultural activities is involved in "subsistence fishing," there is no need for a code below the level which separates "farming" from "fishing" in that category, because the effort of identifying the relatively few numbers of people involved in other kinds of "fishing" would not be compensated by the information obtained.

However, when codes with a greater level of detail are used, the coders must be very thoroughly trained to interpret and classify the information recorded by the enumerator, and to recognize the subtle differences between one category and another. Some of the distinctions between categories, particularly in the economic activity and occupation code groups, are very finely nuanced, and usually require much more information than was probably recorded on the Census forms. This can be seen from the tabulations at the three-digit level for HPC occupation and industry codes, and at the three- and four-digit levels of activity code in the EC tabulations, where "bunching" of values occurs at the lower levels of a given category.

In coding the forms (EC and HPC) for the 1993 Census, the most common error was inconsistency in the use of codes. For example, the same coder would use two or more codes for what was apparently the same occupation description (or economic activity or educational certificate), sometimes for persons in the same HH. Even more common was a difference of interpretation from one coder to another: the same written information might have been coded quite differently, depending on the coder. Many of these errors showed up in the preliminary edit/manual correction phase, but it is probable that even more of them went undetected. Without any control records from the coding operation, there is no way of knowing how seriously the quality of the data were affected by these inconsistencies.

The geographic items presented other types of difficulties, particularly where the respondents offered less-detailed information. The coding scheme was complicated by management's last-minute decision to indicate the locality, rather than a simple urban/rural designator, in the final code. The forms had been designed to hold only a one-digit code, and coders were obliged to squeeze three digits into the space (for two separate questions). The resulting illegibility of the numbers often caused problems during the keying operation, but the biggest problem was the possibility of error in assigning locality codes. Following management instructions, the CDPU alphabetized the data base of localities in each wilayat, and assigned a unique three-digit code to each locality according to its position in the alphabetized list. The task of the coder was to read the information recorded by the enumerator and match it to a name on the list. For wilayat centres, and some of the more populous localities, this was not usually difficult. However, for smaller and/or lesser-known localities, there was no guarantee that the name recorded by the enumerator was the name recorded in the data base. The difficulties, encountered by the cartographers during the mapping operation and later by enumerators during field operations, in determining the precise name of a locality made it very likely that localities were incorrectly identified during the coding operation. Data users must be aware that while information on residence may be accurate to the level of wilayat, below that level it may be of doubtful quality.

**For the next Census**, it is strongly recommended that the number of individual items requiring coding be kept to a minimum. For those items absolutely requiring coding, it is recommended that the number of digits used be commensurate with the likely incidence of the characteristic in the population. Regardless, however, of which items are coded and the scheme used, coders must be intensively trained in interpreting the written information to identify the correct code to be used. In addition, as with the manual edit, strict control of quality must be maintained, following a full/partial verification scheme similar to that outlined in the previous section. Coded data in which the code bears only an occasional relationship to the written information are of no use to the data analyst or any other user.

## 5. DATA ENTRY

In spite of difficult beginnings, the data entry operation ran fairly smoothly once a routine was established. In the coming decade, when Oman will be considering its next Census, the options for collecting data and rendering it processable by computers will likely be very different, and the next Census Project may not have to be concerned with conventional keying and verification. However, if this option *is* chosen for the next Census, it is imperative that those responsible make an early decision on staffing the data entry operation. Based on the experience of this Census, there seems no good reason not to use local (i.e., Omani nationals) as data entry staff. If population projections are accurate, there should be a plentiful supply of secondary and technical school leavers from which to select the number of staff necessary to carry out traditional data entry.

**For the next Census**, however, it is strongly recommended that the process of recruiting and selecting potential staff be assigned to the appropriate administrative office. The CDPU staff of the future will have more than enough substantive work to be carried out, and should not be burdened with administrative tasks for which other Ministry or Census Project units are better qualified. The CDPU can work with the administrative section(s) to establish minimum standards for selection, and should have authority to review and reject or accept those candidates initially selected by the administrative unit, but the bulk of the burden should be borne by staff whose principal function is to select personnel for Ministry or Census operations.

## 6. DATA EDITING AND CORRECTION

In the opinion of the DPA, no single issue caused greater difficulties to the CDPU than the insistence of the subject-matter staff that some data (in both Censuses) be corrected manually (rather than by computer). This opinion is supported by the number of post-edit corrections and adjustments which had to be applied to certain data items (e.g., school attendance, educational attainment, occupation, etc.) **even after the final computer edit** simply because the manual corrections were not carried out in a consistent fashion. In the field of statistical data processing, it is accepted that operations carried out by human beings are subject to error in a way in which computer operations are not. That is, for a given series of steps to be carried out, a computer (properly programmed, of course) will never deviate from the program and will always follow the same path to reach the same conclusions. In contrast to this precision of operation, it is virtually guaranteed that any two human beings, required to follow the same instructions, will perform them differently and will reach different conclusions from the same premises. In addition, the same human being will, at different times or under different circumstances, reach different conclusions from the same premises.

The subject-matter staff argued that to use only the computer for editing and correction would require greater knowledge of the population and its characteristics than was available prior to the Census itself. For this reason, they felt that the intervention of editors was required to resolve certain errors and inconsistencies in the data on a case-by-case basis, and the edit specifications prepared by the subject-matter staff were marked accordingly: "Check only" (for cases requiring manual intervention) or "Check and correct" (for cases where the computer could make a correction).

In spite of the known tendency of human beings to perform inconsistently, the manual correction operation **could have produced data with fewer errors, if the editors making the corrections had been properly supervised and their work reviewed on a regular basis.** Unfortunately, editors assigned to the manual correction task were not given enough specific training in resolving errors (they were trained only in use of the error listings and the computer), and their work was not subject to any supervisory oversight. This led to great variations, from one editor to another, in responding to similar errors, and ultimately obliged the CDPU to re-edit the data before acceptable tabulations could be generated.

**For the next Census,** the subject-matter staff should eliminate entirely any editing tests which require human intervention to determine the correction. In the years between this Census and the next, the Sultanate will accumulate, by means of periodic surveys and other sources, a body of information [about the characteristics of its population] on which rules for automatic correction may be based. Manual editing and correction, while it can be justified for a survey (with its much smaller number of respondents), should have no place in a processing system for the huge volumes of data associated with a census.

## 7. DATA TABULATION

The difficulties encountered during tabulation varied slightly from one Census to the other. Each will be discussed separately.

### a. Establishment Census

In the Establishment Census, the problem which caused most difficulties was the requirement to show economic activity at the four-digit level. Because of the extremely large number of individual codes at this level of detail, all tabulations using economic activity as one of the variables (i.e., the majority of the tabulations requested) had to be subdivided into segments (usually four), because of system limitations on the number of cells permitted in a single logical table. In addition, the use of complete, detailed descriptions, in both English and Arabic, for each one-, two-, three-, and four-digit code made the process much more cumbersome than it might otherwise have been. Add to these problems the fact that table layouts were not properly specified to begin with, and it is easy to see why so much time was lost in checking and re-checking the tables.

**For the next census,** data users should seriously consider eliminating the use of verbal descriptions of economic activities in the tabulations, or at least restricting such descriptions to one- or two-digit codes only. Wherever codes of three- or four-digits are used in a tabulation, descriptions should

not be included in the table itself, but may be attached as a supplement. In this way, one copy of the full descriptions, in English and Arabic, can be used for all tables produced at any of the levels of detail of economic activity. This will make the tables much more readable, and will permit the programmer to group codes visually, without uneven spacing because of excess text. It will also make it possible to print tables using less paper overall. It will also greatly smooth and speed up the task of the programmer, by eliminating the need for manipulating massive quantities of text, and all the problems attendant on this task.

b. Housing and Population Census

In the HPC, there were a number of problems related to the generation of tabulations. The biggest single problem was the lack of precise specifications with respect to table layout and text (table title, heading, stub and column descriptions, footnotes, etc.). In spite of an early attempt by the CDPU to establish (in conjunction with the subject-matter specialists) formatting standards for the tables, the requirements changed several times over the period of months from first design to final table generation. In many cases, tables were modified several times as successive users each specified different changes (e.g., table centered vertically and/or horizontally on the page or moved to one margin; different type faces and font sizes specified; different text and table number styles, etc.). The number and frequency of changes, particularly of those which effectively canceled each other out, unnecessarily burdened the programming staff.

In addition, there were many modifications made to table universes—that is, the population tabulated (e.g., all Omanis, all persons 10 years and older, all ever-married Omani females, etc.). Such changes have implications not only for the physical layout of the table (headings and text), but also for the logic of the program which generates the cell values. And, in spite of all the time and effort on the part of the CDPU to satisfy the user requirements, the tabulations produced by the CDPU were not published as generated, but were used merely as input to other software. The number of person-hours, wasted in trying to meet specifications that were later nullified, is incalculable; had the CDPU known ahead of time that layout was unimportant, the table-generation process might have been accomplished months earlier.

**For the next census**, the subject-matter staff should first decide whether it wants the CDPU to produce publication-ready tables, or simply to produce the numbers for input into other formatting software. If publication-ready tables are desired, then the user must establish precise specifications for all aspects of table presentation: position on the page, heading text (upper- and lower-case, or upper-case only; heading centered or right- or left-justified, etc.), table numbering conventions, continuation of tables that exceed one page, etc. Clear and unambiguous standards must be established for all of these items before programmers begin the work of designing tables on the computer. In addition, the precise content of each table (population tabulated, variables used, value grouping of variables, etc.) must be defined and explicitly specified at the same time the physical layout is prepared. These preliminary steps will save a great deal of time and effort on the part of both subject-matter staff and data processors, and will help ensure the ultimate value of the tabulations.

## 8. OPERATIONAL PROCEDURES

With the exception of instructions for the data entry operators, none of the procedures required to prepare data for the final edit were documented. This made it difficult to ensure that all persons involved in the manual edit/correction phase (for example) carried out the task in the same manner, and, in fact, the results show clearly that they did not. This lack of written instructions for the various tasks was also felt in the area of operational control, as well, where decisions were often made without regard to proper procedure, simply because the procedures were not documented. Part of the reason for this lack of documentation was the "last-minute" nature of many of the procedures—that is, they were defined as the need for them arose during the post-enumeration processing, rather than having been defined prior to enumeration. However, even such hastily-devised procedures can, and should, be documented.

**For the next census**, Census Project management should ensure that post-enumeration procedures are defined, in detail, soon enough before the procedures must be implemented to allow for writing them down in a clear and unambiguous manner. These documents (i.e., written descriptions of procedures) should

## Data Collected

**Geographic Classification**

Region (1:8)	nn	
Wilayat (according to Region)	nn	
Locality (numbered within Wilayat)		nnnnn
Hilla (for areas with Hilla)	nnn	
Census Block (1:n)	nn	
Enumeration Area (in 15 Supervisor Areas)		nnnnn

**Response Unit Identification**

Census Book serial number (1:n within each EA)	nn
Housing Unit serial number (1:n within each Book)	nn
Household number (1:n within each Census Book)	nn
Household number (1:n within each Housing Unit)	n
Person number (1:n within each Household)	nn

**Housing Unit Information**

Type of unit (1:4 = conventional, 5:6 = rural, 7 = collective, 8 = other)

Occupancy (occupied, closed, vacant)

Reason for vacancy (if not occupied)

*For occupied Housing Units:*

Tenure

Source of water supply

Lighting

Air conditioning

Toilet

Bath

Kitchen

Cooking fuel

Living rooms

Number of households

**Household Information**

Number of:

Television

VCR

Stove, oven

Refrigerator

Freezer

Electric washing machine

Telephone

Private cars

Household size

## Population Information

### Basic demographic data

Relationship to Head of Household  
Sex  
Age at last birthday (00:98+)  
Nationality (3-digit)  
Religion (no sect)

### Residence and migration data (Omani nationals only)

Where enumerated (inside Oman, outside Oman)  
Usual residence (wilayat/locality, location outside Sultanate)  
Length of residence (years)  
Previous residence (wilayat/locality, location outside Sultanate)  
Wilayat of origin

### Education data (6-40 years of age)

School attendance  
Level and grade (currently attending: current grade/level;  
previously attended: highest grade/level achieved)

### Education data (10 years of age and over)

Educational attainment (3-digit)

### Economic data (10 years of age and over)

Activity status (working, seeking work, housewife, student, etc.)  
*For currently or previously employed:*  
Occupation (3-digit)  
Industry (3-digit)  
Employment status  
Sector  
Place of work (wilayat, location outside Sultanate)

### Nuptiality/fertility data (10 years of age and over)

Marital status  
*For ever-married Omani nationals only:*  
Age at first marriage  
*For ever-married female Omani nationals only:*  
Number of children born alive (M/F)  
Number of children still living (M/F)  
Number of children born in 11 months preceding Census  
Number of children born in 1992 (12 months)  
*For currently-married male Omani nationals only:*  
Number of wives

### Disability data (for Omani nationals only)

Type of disability  
Cause of disability

## Tabulations

**NOTE:** **Administrative division** refers to region and/or wilayat, depending on the individual table.

**Economically active/inactive:** any reference to **type of activity** automatically implies population 10 years and older.

**Marital and fertility:** any reference to **marital status** or **fertility** categories automatically implies population 10 years and older.

**Education:** any reference to **educational attainment** automatically implies population 10 years and older; **school attendance** implies population 6-40 years of age.

**Urban/rural [U/R]:** all localities with population  $\geq 2500$ , and all wilayat centres regardless of population size, are designated as **urban**; all other localities are **rural**.

Basic Demographic: data in these tables shown by administrative division [region/wilayat]

- Localities by population size
- Population [Urban/Rural] by sex
- Omani/non-Omani households by size of household
- Total/Omani population by sex and relation to Head of Household
- Total/Omani population by sex and age group
- Total/Omani population by sex and religion
- Total/Omani population by sex and educational attainment
- Total/Omani population by sex and marital status
- Total/Omani population by sex and type of activity
- Economically active total/Omani population by sex and occupation
- Economically active total/Omani population by sex and industry
- Economically active total/Omani population by sex and employment status

Household composition

- Heads of Households by age and sex and other members of household by age and relationship to Head
- Omani Heads of Households [U/R] by age and sex and other members of household by age and relationship to Head
- Population 10+ by relationship to Head of Household, age, and sex
- Omani population [U/R] by relationship to Head of Household, marital status, and sex
- Omani Heads of Households [U/R] by type of activity, employment status, and sex and other members of household by relationship to Head and type of activity
- Total/Omani/non-Omani households [U/R] by size of household and number of economically active members

Residence and migration: Omani population only

- Population [U/R] by place of enumeration, place of usual residence, and sex
- Population [U/R] by place of usual residence, duration of residence, age, and sex
- Population [U/R] by place of usual residence, duration of residence, place of previous residence, and sex
- Population changing residence within the 10 years preceding the Census by age and sex
- Population changing residence within the 10 years preceding the Census by educational attainment, occupation, and sex

## Attachment IV.F.11 - 2 of 4

- Population changing residence within the 10 years preceding the Census by educational attainment, industry, and sex
- Population by wilayat of origin, wilayat of usual residence, and sex
- Employed population by wilayat of usual residence, wilayat of workplace, and sex
- Employed population whose wilayat of usual residence differs from wilayat of workplace, by occupation and sex
- Employed population whose wilayat of usual residence differs from wilayat of workplace, by industry and sex

### Nationality

- Population by nationality, sex, and administrative division
- Population by nationality, sex, and age
- Population by nationality, sex, and educational attainment
- Population by nationality, sex, and marital status
- Population by nationality, sex, and activity status
- Economically active population by nationality, sex, and occupation
- Economically active population by nationality, sex, and industry
- Economically active population by nationality, sex, and employment status

### Secondary demographic (including marital/nuptiality)

- Total/Omani population [U/R] by single year of age and sex
- Total/Omani/non-Omani population [U/R] by age and sex
- Total/Omani population by age, marital status, and sex
- Omani population [U/R] by age, marital status, and sex
- Non-Omani population by age, marital status, and sex
- Ever-married Omani population [U/R] by age, age at first marriage, and sex
- Currently-married Omani males [U/R] by age and number of wives
- Currently-married Omani males [U/R] by educational attainment and number of wives
- Currently-married Omani males [U/R] by type of activity and number of wives
- Currently-married Omani males [U/R] by age at first marriage and number of wives

### Education

- Total/non-Omani population currently attending school by grade, level, and sex
- Omani population [U/R] currently attending school by current grade, level, and sex
- Omani population [U/R] who attended school in the past by highest grade, level, and sex
- Total/non-Omani population by educational attainment, age, and sex
- Omani population [U/R] by educational attainment, age, and sex

### Economic activity

- Total/non-Omani population by age, sex, and type of activity
- Omani population [U/R] by age, sex, and type of activity
- Economically active/inactive Omani population by age, marital status, and sex
- Economically active Omani/non-Omani population by sex and sector of employment
- Economically active total/non-Omani population by age, sex, and occupation
- Economically active Omani population [U/R] by age, sex, and occupation
- Economically active total/non-Omani population by age, sex, and industry
- Economically active Omani population [U/R] by age, sex, and industry

Attachment IV.F.11 - 3 of 4

- Economically active total/non-Omani population by age, sex, and employment status
- Economically active Omani population [U/R] by age, sex, and employment status
- Economically active total/non-Omani population by educational attainment, sex, and occupation
- Economically active Omani population [U/R] by educational attainment, sex, and occupation
- Economically active total/non-Omani population by educational attainment, sex, and industry
- Economically active Omani population [U/R] by educational attainment, sex, and industry
- Economically active total/non-Omani population by employment status, sex, and occupation
- Economically active Omani population [U/R] by employment status, sex, and occupation
- Economically active total/non-Omani population by employment status, sex, and industry
- Economically active Omani population [U/R] by employment status, sex, and industry
- Economically active total/non-Omani population by industry, occupation, and sex
- Economically active Omani population [U/R] by industry, occupation, and sex
- Economically active total/Omani/non-Omani population by occupation, industry, and sex

Fertility: Ever-married (10 years and older) Omani female population

- Population [U/R] by age and number of live births, and total number of live births by sex of child
- Population [U/R] by age and number of surviving children, and total number of surviving children by sex of child
- Population by age and educational attainment, and total number of live births by sex of child
- Population by age and educational attainment, and total number of surviving children by sex of child
- Population [U/R] by age and educational attainment, and total number of live births in 1992
- Population [U/R] by age and educational attainment, and total number of live births in 1993 [Jan.-Nov.]

Disability: Omani nationals (any age) declaring disability

- Population by type of disability, reason for disability, and administrative division
- Population [U/R] by type of disability, reason for disability, age, and sex
- Population [U/R] by type of disability, reason for disability, activity status, and sex
- Population [U/R] by type of disability, reason for disability, educational attainment, and sex

Living Quarters I: data in these tables shown by administrative division [region/wilayat]

- Living quarters, households, and persons by type of quarters
- Living quarters, households, and persons by occupancy status
- Vacant living quarters by reason for vacancy
- Occupied living quarters by type of tenure
- Occupied living quarters by water supply
- Occupied living quarters by type of lighting
- Occupied living quarters by existence and type of air conditioning
- Occupied living quarters by existence and type of toilet
- Occupied living quarters by existence and type of bathing facilities
- Occupied living quarters by existence and type of cooking facilities
- Occupied living quarters by type of cooking fuel
- Occupied living quarters by number of living rooms
- Average number of persons per household and per room in occupied living quarters

Living Quarters II: data in these tables shown for urban/rural divisions

- Living quarters by type and occupancy status
- Occupied living quarters, households, and persons by type of quarters and tenure
- Occupied living quarters, households, and persons by type of quarters and water supply
- Occupied living quarters, households, and persons by type of quarters and lighting
- Occupied living quarters, households, and persons by type of quarters and air conditioning
- Occupied living quarters, households, and persons by type of quarters and toilet
- Occupied living quarters, households, and persons by type of quarters and bath
- Occupied living quarters, households, and persons by type of quarters and cooking facilities
- Occupied living quarters, households, and persons by type of quarters and cooking fuel
- Occupied living quarters, households, and persons by type of quarters and number of living rooms
- Average number of persons per household and per room in occupied living quarters

Household

- Occupied living quarters, households, and persons by nationality of Head of Household and administrative division
- Occupied living quarters, households, and persons by nationality of Head of Household and type of quarters
- Occupied living quarters, households, and persons by nationality of Head of Household and number of living rooms
- Occupied living quarters, households, and persons by nationality of Head of Household and existence and type of air conditioning

Domestic conveniences

- Number of households and number and type of domestic conveniences by administrative division
- Number of households and number and type of domestic conveniences by nationality of Head of Household and U/R

be used in training the persons who will carry out the procedures, and management should be prepared to revise—and rewrite—the procedures whenever circumstances require. Revisions should also be communicated to the staff in both written and oral forms, and until such changes in procedure become "second nature," supervisors should oversee the work of their staff more closely to be sure that the new procedures are being followed properly. Written procedures provide both a standard against which performance can be measured and assurance that all persons involved in a task operate from a common ground.

## 9. TIME SCHEDULE

In a census processing operation, there are multiple sub-operations being performed simultaneously, each one of which is usually linked to others in a complex web of interdependency. For example, the data entry operation is dependent on the flow of work through the manual editing and coding operations; data editing cannot begin until a sufficient number of work units (i.e., Census Books) have been checked in and registered in the control data base; computer editing cannot be done until a sufficient number of work units have been keyed and verified, etc. Each phase of post-enumeration processing attempts to follow a pre-established timetable, so that the timetables for all subsequent phases can be effective.

When events force a change to any one timetable, others are likely to be affected, as well. In the best of circumstances, the only effect will be the buildup of a backlog of work for the next phase(s), which usually can be tolerated, but it is more likely that the effect will be a drop in productivity, so that subsequent phases (which depend on output from the affected activity) will suffer a similar decrease in productivity. And when the change in timetable is imposed by management, it usually involves a requirement for increased output in one or more phases. Such an increase can usually be achieved only by increasing the number of staff and/or the number of hours worked, and will have consequences that must be carefully examined before adopting the new timetable.

**For the next census**, it is strongly recommended that all post-enumeration timetables be carefully studied before being accepted as targets for production. It is further recommended that, in the production environment, when management desires to make changes to any prevailing timetable, approved activity, and/or procedure, the proposal for such a change be accompanied by full details on the resources which will be required to implement the change and consideration of the effect of the change on other activities. Only after review of this information, and approval of the change by the persons responsible for all affected activities, should such a modification be accepted. This will reduce the likelihood of unilateral changes which impose unreachable goals or which unnecessarily create difficulties for the persons responsible for processing activities, and will eliminate questions about why target dates are not being met.

## 10. POST-ENUMERATION SURVEY (PES)

During the planning for the 1993 Census, the CTC considered the need for a formal post-enumeration survey, and concluded that it could effectively substitute certain field review procedures to guarantee completeness of coverage. Thus, a PES was deemed unnecessary, and was not carried out. Unfortunately, the field procedures which could have provided information for post-enumeration evaluation were not carried out in a uniform fashion, and there was no formal evaluation of either completeness of coverage or accuracy of data collection. Such an evaluation would be useful to both Census Project management and to the users of the data. Management would have information indicating where field operations and staff training need to be reinforced; users of the data would know how well the Census operation covered the territory of the Sultanate, and whether there was under- or over-enumeration (by region, age group, sex, etc.). It would also help users (and Census Project management) understand how precisely the data (on age, sex, nationality, etc.) were recorded by the enumerators, and could have provided valuable guidance for future statistical activities, particularly with respect to enumerator training.

**For the next census**, it is strongly urged that the plans include a Post-Enumeration Survey. Properly executed, a PES will require a separate organization to guarantee independence of operation (and statistical validity). Because the PES will be carried out only in a sample of areas throughout the Sultanate, the numbers of persons required for the field procedures will not be as great as those required for normal enumeration. Since the PES is also carried out after the close of the normal enumeration, it is possible to re-use staff from the regular enumeration for the PES, but great care must be taken to ensure that no staff are assigned to the same physical areas. If a PES is properly planned for and conducted, and the results carefully analyzed, it will add greatly to the ultimate value of the data collected in the next Census.

## VI. APPENDICES

In this section will be found a description of the major data bases and of the software used in processing.

### A. Data bases

During the development of the basic systems used for processing the Census data, as well as during production processing, a number of data bases—that is, data files in a formal data base structure—were used. These data bases were developed using the dBase IV software package; entry of names in Arabic and English was achieved through use of the Nafitha™ Arabization software.

#### 1. OCCUPATION

This data base consisted of one record for each of the one-, two-, three-, and four-digit occupation code groups defined for use in the Census. Each record contained the following information:

- o Occupation code (one-, two-, three-, or four-digit, depending on level)
- o Description of occupation code group in English (200 characters)
- o Description of occupation code group in Arabic (200 characters)

#### 2. INDUSTRY [ECONOMIC ACTIVITY]

This data base consisted of one record for the one-character industry [economic activity] code groups and one record for each of the two-, three-, and four-digit code groups defined for use in the Census. Each record contained the following information:

- o Industry code (one-character, two-, three-, or four-digit)
- o Description of industry code group in English (200 characters)
- o Description of industry code group in Arabic (200 characters)

#### 3. GEOGRAPHIC

This data base consisted of multiple files, with one file for each level of the hierarchy, as follows:

- o for regions (names and code), with one record for each region;
- o for wilayats (names and code), with one record for each wilayat within each region;
- o for localities (names and code), with one record for each locality within each wilayat;
- o for hillas (names, if any, and code), with one record for each hilla within each locality;
- o for blocks (code only), with one record for each block within each hilla;
- o for EAs, with one record for each EA; and
- o for Census Books, with one record for each Census Book used in each EA.

Each field was defined as follows:

- o Region code (2-digit)
- o Region name in Arabic (40 characters)
- o Region name in English (40 characters)
- o Wilayat code (2-digit)
- o Wilayat name in Arabic (40 characters)
- o Wilayat name in English (40 characters)
- o Locality code (5-digit)
- o Locality name in Arabic (40 characters)
- o Locality name in English (40 characters)
- o Hilla code (3-digit)
- o Hilla name in Arabic (40 characters)
- o Hilla name in English (40 characters)
- o Block code (2-digit)
- o EA code (5-digit)

- Census Book number (2–digits)
- Number of Census Books used in EA (2–digits)

#### 4. **BOOK CONTROL**

This data base contained the same types of files and records and the same information as the GDB, although the number of records was different, due to the post-enumeration inclusion of additional Census books not previously entered.

### **B. Abbreviations and Definitions**

AS	Assistant Supervisor (of enumeration)
BCDB	Book Control Data Base
BuCen	U.S. Bureau of the Census
CDPU	Census Data Processing Unit
CL	Crew Leader (of enumeration)
CTC	Census Technical Committee
DP	Data Processing
DPA	Data Processing Advisor
EA	Enumeration Area
EC	Establishment Census
GDB	Geography Data Base
HH	Household
HPC	Housing and Population Census
HU	Housing Unit
IMPS	Integrated Microcomputer Processing System
MoD	Ministry of Development
MoH	Ministry of Health
OAJC	Omani-American Joint Commission

### C. IMPS (Integrated Microcomputer Processing System)

IMPS is a collection of modules for performing the most common tasks associated with processing of survey and census data. It was developed by the U.S. Bureau of the Census under financing by AID/W. The modules have been designed and developed specifically to function as an integrated system, although each of the elements can be used without reference to any of the others except, in some cases, the Data Dictionary, which is required for all modules which refer to data items (including geography). The component subsystems in IMPS are:

#### 1. Data dictionary

The data dictionary subsystem permits the user to create a reference file describing the structure and content of the file(s) to be created from the census/survey data. Dictionary files can also be created for already existing data files, as well as for auxiliary data files, as long as the file in question exists in a standard ASCII flat-file format. The IMPS subsystem cannot work with data files in non-standard formats (such as data base or spreadsheet files, for example). In files which contain multiple record structures, each separate structure must also contain a unique record type code identifying the structure. This type code must be in the same physical position in each of the different structures. In addition, any identifying information (such as geography, survey class, etc.) which is required for controlling processing must also be included in each structure in the same absolute position(s).

#### 2. Data entry [CENTRY]

Once the data dictionary has been defined, the user can create one or more modules for controlling the entry of data from census/survey forms or other documents. The screen control program will automatically contain valid value checks (if valid values have been specified in the dictionary), and other checks may be programmed by the user. The CENTRY subsystem also permits the user to control the entry, verification, and modification of data as separate operations, and the software collects statistics (keying speed, error rate, elapsed time) on each operator for each session and type of activity. These statistics may be accumulated by the user for analysis of operator effectiveness and productivity.

#### 3. Data edit and imputation [CONCOR]

Once the data have been entered, verified, and corrected, the CONCOR subsystem may be used to further validate and correct the data. [A CONCOR program may also be used in conjunction with the actual entry of data, but the CDPU elected not to use this option during Census processing so keying output could be maximized.] A CONCOR program is developed from a set of edit specifications, usually prepared by a subject-matter specialist, and when final, can detect and correct errors without human intervention. Edit programs written in the CONCOR language can generate various types of error statistics for post-edit analysis.

#### 4. Publication tabulation [CENTS]

Cross-tabulations of data from the census/survey operation may be generated by using the facilities of the CENTS subsystem. Table layouts can be designed with all text (mono- or bi-lingual) in place and with cell locations indicated by using the *formatting* subsystem; a program can be generated to provide the correct values for each cell by using the *tabulating* subsystem; and, when necessary, tabulations can be generated at up to five separate levels of a hierarchy (geographic or other) by using the *area structure* subsystem. Tables may be created in "publication-ready" format, for use in presentation documents, or they may be generated with minimal text for internal or preliminary use.





## Establishment Census Data Collected

**Geographic Classification**

Region (1:8)	nn
Wilayat (according to Region)	nn
Locality (numbered within Wilayat)	nnnnn
Hilla (for areas with Hilla)	nnn
Census Block (1:n)	nn
Enumeration Area	nnnnn
or	
Crew Leader Area	nnnn

**Response Unit Identification**

Establishment serial number (1:nnn within each EA)	nnn
--	-----

**Establishment Information**

For all establishments:

Census building number	nn
Street name or number [not recorded in keyed data]	
Number of sikka or zaqaq	nnnn
Municipal building number	nnnn
Establishment status [operational, temporarily closed, other]	n

For operational establishments only:

Sector of activity [private, public, other]	n
Principal activity	nnnn
Secondary activity	nnnn
Establishment nationality [Omani, mixed, non-Omani]	n

For operational commercial establishments only:

• Commercial registration number [12 characters]	
Establishment structure [single, main office, branch]	n
Legal status of establishment [single-owner, partnership, etc.]	n
Number of unpaid employees	nn
Number of paid Omani employees	nnnn
Number of paid non-Omani employees	nnnn
Total number of employees, paid and unpaid	nnnn
Telephone number of establishment	nnnnnnnn
Post Office Box number of establishment	nnnn
Postal Code of establishment	nnn
Establishment name [50 characters]	
Major economic activity group [alphabetic]	x

For branch offices only:

Telephone number of main office	nnnnnnnn
---------------------------------	----------



