Distr.
GENERAL

CES/SEM.40/6
18 September 1998

Original: ENGLISH

STATISTICAL COMMISSION and
ECONOMIC COMMISSION FOR EUROPE

STATISTICAL OFFICE OF THE
EUROPEAN COMMUNITIES (EUROSTAT)

CONFERENCE OF EUROPEAN STATISTICIANS

Joint ECE-EUROSTAT work session
on Population and Housing Censuses[1]
(Dublin, Ireland, 9-11 November 1998)

Study topic 2

WHEN IMAGE MET ASCII

Data Processing in the 1995 Census of Population and Housing in Israel

Invited paper submitted by the Central Bureau of Statistics, Israel[2]

1. The toil of collecting raw data and transforming it to a meaningful statistical knowledge, is the dominant process carried out in statistical bureaus. Changes and improvements are introduced to the process in several realms, among which are the adaptation to the changing needs of the users and the developing of statistical methods to better reflect reality. However, in this knowledge age, the growing needs that put a pressure upon the bureaus to produce a reliable, timely and updated information, may stimulate also technological development, changes and adaptations.

2. Censuses have always been an arrow head of such processes because of their importance in national statistics, their magnitude and consequently, their high costs.

1 The papers which are prepared for this work session will be treated in the same manner as papers that are prepared for seminars.

2 Prepared by Olivia Blum & Eliahu Ben-Moshe.

3. At the beginning of the nineties, the development of windows technology and the improved* optical-reading ability were “drafted” to the census process. Census goals were not altered, but have been re-enunciated in the light of the transpiring changes in the environment of census taking and data processing. The ability to open simultaneously several windows with logical and physical connections between them and the more reliable optical character recognition, opened an opportunity to look at tasks ahead, concerning data processing, in a different manner.

4. During the 1995 Israeli Census, a new approach to processing census data was adopted, while taking advantage of the new technologies available at that time and the international experience accumulated up to that point.

5. In this paper we scan several aspects of the shuttering effect of the changing technological environment on the essential segments of the census. We elaborate on the ideology behind census planning, data processing and process quality control.

Planning for an Optical Data Entry System (ODE)

6. Changing the technological working-environment can be done in two main patterns; by using new tools to perform the very same tasks and processes, as were done before, or by optimizing the use of these new resources while altering the logical principles of the system and its parts.

A: The first tactic, the mere computerization of the data-capture process, means reducing the time needed to perform the process and producing a higher quality census data-file because of the decreased involvement of the human subjective component. Data capture tasks are shifted from the people to the machine, and therefore, processes are uniform and consistent and the end result is more reliable than the less computerized alternatives. Costs might be lower in a computerized system, if technological development is not required.

B: However, the second tactic, where technology is recruited while rethinking the ideology behind the whole census process, has additional merits that are expressed in the organizational aspects as well as the substantive ones:

B1: Organizational Aspects

7. In the technological environment, the order of steps is not stipulated by the logistic of handling paper questionnaires, therefore, it can be altered, avoided as a decision variable altogether, or guided by new logical principles of the system. For example, there is no need to choose whether to first edit data or first code them. Copies of the same images can be sent at the same time to both. Therefore, both tasks can be done simultaneously,

---

* Optical recognition has already been developed several decades earlier, and was used in censuses and census tests since the beginning of the 1970s.

unless a logical principle, like editing the already coded data, dictates differently. Likewise, the new working environment opens the question whether to capture edited and coded data or to capture raw data and perform these tasks later. The issue of order and logistics has a broader effect if one is not confined to the boundaries of data-capture process, but considers processes performed before or after it as well. There is no need to gather the questionnaires in a special order if each page carries identifying variables. They can be put on the scanner's tray in a random order or in any desirable order according to subject-matter considerations. Furthermore, identifying duplicate records, usually done in the macro-editing stage in the central computer, can be performed during the data capture process, while still census records are accumulated.

B2: Substantive Aspects

8. The substantive aspects of optimizing the use of resources in a computerized optical system, are spread all across the census span; the questionnaires have to be designed for optical reading. They have to contain identifying variables for each physical and logical unit, the questions are preferred to be closed ones to facilitate the optical recognition etc.

9. Bringing the computer to the beginning of the data capture process opens the possibility to modify the meaning of all the following tasks; keying becomes the corroboration of optical recognition, micro-editing can be reduced or even avoided entirely, coding can be carried out by way of interactive queries and so on.

10. Moreover, if a file is used for cold deck imputation, performed in the macro-editing stage in the central computer, it can be linked with census data as early as the first character recognition round is accomplished. It doesn't have to be detained until all census data are gathered, this external file can be part of the data capture process, corroborate the optical character recognition, support the identification of duplicate records while still having the images of the questionnaires involved in the system, and it can also serve for imputation of variables, if micro-editing is performed during data-capture.

11. Although census goals remain the same, regardless of whether the working environment is technologically advanced or not, new tools provide flexibility in planning the process and enable the achievement of new intermediate objectives. In a system with optical character recognition, the accessibility of the images of the questionnaires, during and after data capture, eliminates the need to perform editing and coding tasks at the beginning of the process, while the paper questionnaire is in front of the operators. Therefore, planners can strive for a real raw data file, as an intermediate objective.

12. In the past, a raw data file in censuses was not a goal for itself and it was never accomplished. The implication was that there has never been a possibility to find out, after data-capture, what has been written on the questionnaires. There was not a way to separate between errors made in the

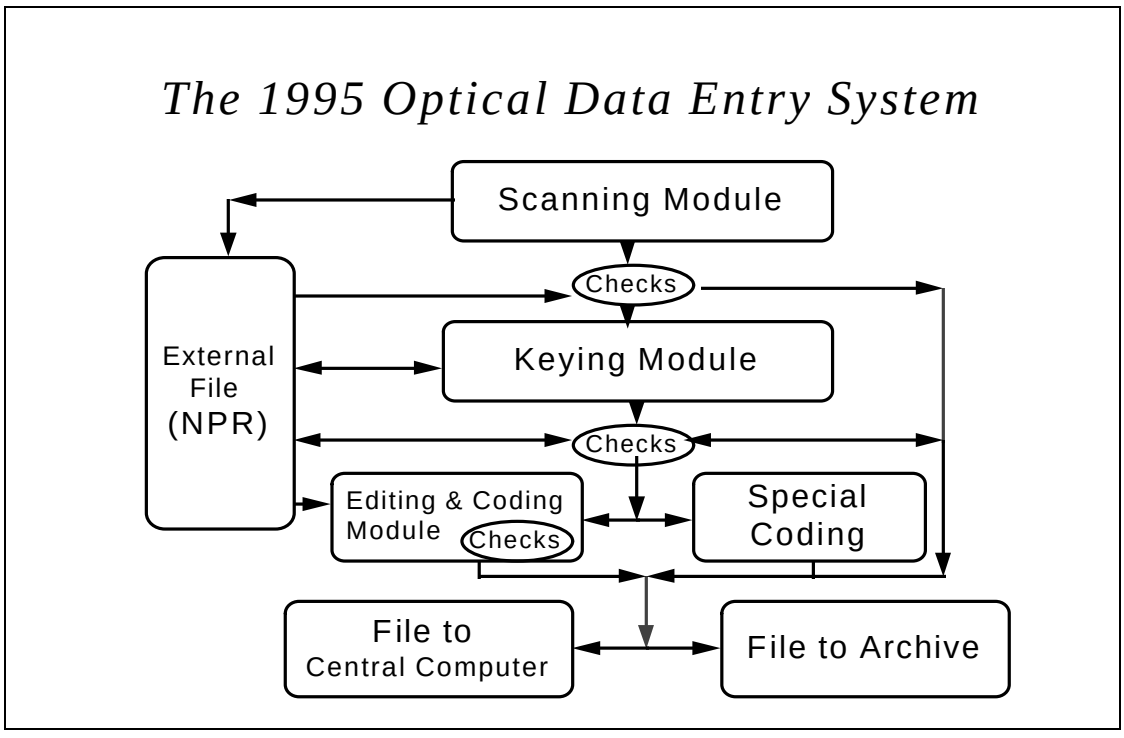
field by the respondents or the enumerators, errors introduced during manual editing or errors added during keying and other data capture processes. Putting a raw file as an objective changes the logical principles as well as the timing of performing data processing tasks.

13. The Israeli 1995 Census of Population and Housing was conducted while using the overall optimization approach. Improved optical character recognition and WINDOWS working environment were the starting conditions of a process in which census planners redefined processes and intermediate objectives, and while doing so, invoked further technological improvements. The next sections are dealing with three issues:

(1) Data processing in the working environment described above;
(2) The use of external files in a computerized system;
(3) The questionnaire and the ODE system.

Data Processing in the Optical System

14. Data processing has several essential modules: data collection, data capture, coding, editing, data analysis and quality control throughout the process. The ODE system of the 1995 Census of Population and Housing in Israel did not include data collection and data analysis, and its structure was as follows:



15. Quality control was an integral part of the system and it was the underlining parameter of all modules and the channeling mechanism between them. The mode in which each of the modules is carried out is a function of two main factors; goals and means.

16. The goal of census data capture is to create a census file with known biases. In the technologically improved system, where images of the questionnaires are available throughout the process, the goal can be refined into getting a structural, raw data file, that reflects the respondents' answers as accurate as possible. This file provides the means to locate errors by their source and to fall back to the original data when one or more of the processes have to be revised. The ability to create several files that have undergone different editing processes enables the evaluation of editing processes using comparative methods, and designated editing of the raw file according to various needs. It contributes not only to the census file, but also as an input to the decision-making process regarding the manner in which large data files should be edited in the future.

17. All tasks carried out during data capture process were attuned to fit the stated goal, while optimizing the use of resources in the provided working environment.

Guiding Principles of Each Module

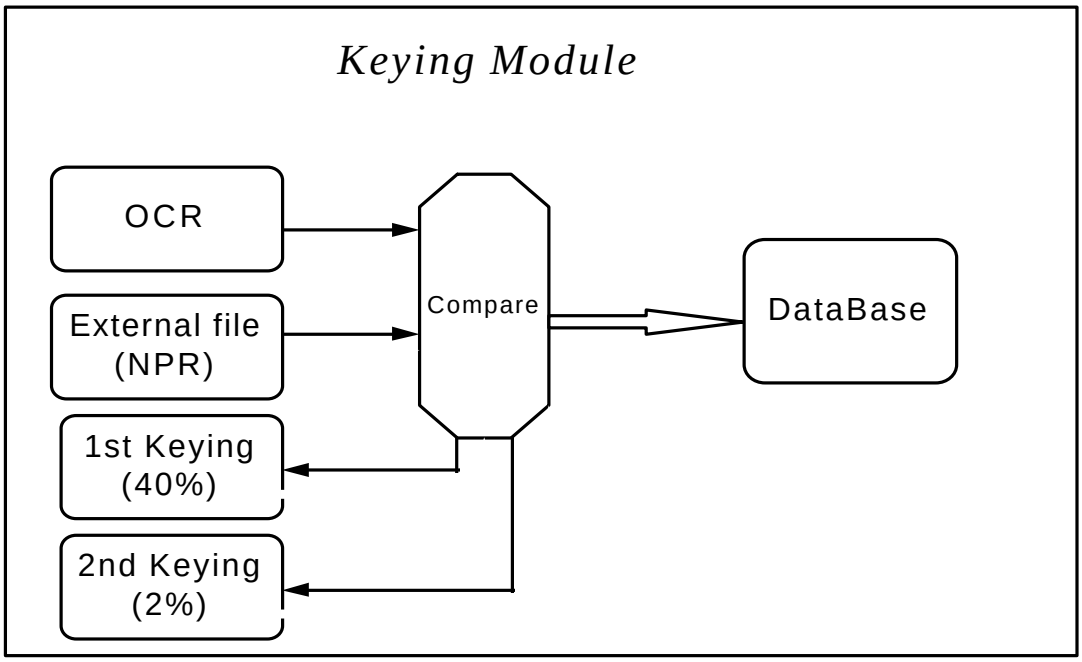
18. The main tasks performed during the **scanning module** are completeness and exclusiveness checks, and Optical mark recognition (OMR) and optical character recognition (OCR).

19. Completeness checks means verification of scanning of both sides of each page and of the expected amount of questionnaires. Exclusiveness checks come to ensure that each enumeration area is scanned only once.

20. OMR & OCR processes assign a value for each field that has been defined as a target for optical identification, and a status of reliability level of the optical recognition. These statuses (Super Sure, Sure, Doubtful and Fail) dictate further treatment of the character, or the field, in the system.

21. For efficient use of computer resources in the scanning module the image of each form is kept only once. The fixed layout of the questionnaires (the template) is dropped out as soon as they are scanned. Furthermore, before inserting the values and accompanying statuses into the database, they are compressed.

22. The **keying module** in the ODE system is defined as a completion and verification of the optically read values. A value of a field is verified once two identification sources indicate the same value. Since the National Population Register (NPR) is linked to the file immediately after scanning, and since there are two keying rounds, comparisons of values are performed in the following pattern:



23. Comparisons are made between OCR (including OMR) and the NPR, identities are sent to the database. The remaining values go through 1st keying and the comparison is made between OCR, NPR and 1st keying. If two out of three are identical, the value is sent to the database. The remaining values go through 2nd keying and the comparison process repeats. In this sequence, only 40% of the characters (excluding alphabetical ones) reached 1st keying and only 2% reached 2nd keying round (see also in the section of “The Use of External Files”). The tremendous saving of human involvement in the keying stage does not end in the above automatic process; keying itself is “smart keying”, using image tailoring technology. There are three types (levels) of keying:

- Verification keying in “carpets” - The operator sees images of 10X10 characters identified as the same number or mark, and has to point at the errors. For example, in a carpet of number “8” all the characters that are not 8 are marked. The system was tuned to have 4-5 wrong characters at most. It means that by clicking 4-5 times, 95-96 characters have got their verification and are not directed to additional handling in this module.

- Correction keying in strings of three characters - The operator sees images of characters from different questionnaires and keys them in, in triplets. The decision to arrange them in triplets is a product of an experiment, and the findings are those three characters is the optimal number to remember and key-in in a relatively high speed.

- Full-field keying - The interface is of images of the same variable in different questionnaire. For example, images of 12 images of country of birth, that have to be fully keyed in. Forming homogeneous keying items are found to be faster and more reliable for data capture.

24. The level of keying is determined by the level of reliability of OCR recognition, by the value captured (whether it falls within a legitimate range or not), and by the corroborating auxiliary information. Basically, the principle is that “super-sure” status means that the value has been corroborated and need no keying at all, “sure” values are brought to verification in carpets, “doubt” to triplets and “failed” to full field keying.

25. Quality control is embedded in the process, using synthetic characters planted in keying items. It enables a continuous follow-up of the quality of keying and quality of keying operators.

26. The objective of receiving a structured raw data file dictates four main **editing** tasks: definition of structural units, modification and completion of the keying stage in those cases where a final value has not been defined, confirmation or modification of captured values in fields which were found to have logical contradictions (without altering the values written on the questionnaire), coding “other” categories (an open category in a closed question).

27. To perform these tasks, a set of preparatory tasks is activated during and immediately after keying stage. It includes automatic coding of two variables (Country of Birth and Relation to Reference Person), detecting written responses in the “other” categories, all physical (enumeration area, questionnaire) and logical (household, individual) units are automatically defined and out-of-range and consistency checks are performed. A failure in a preparatory task creates an editing item that includes all editing problems found in one household.

28. The editing work environment enables seeing simultaneously the image of the full page of the questionnaire, images of all pages belong to the household, as well as images of any household in the enumeration area. It is also possible to see simultaneously written responses on the questionnaire, as recorded by the enumerator and the respondents, together with the ASCII values of the corresponding fields in the database. There is an interactive accessibility to an external file, coding dictionaries and process-information tables, accessibility to virtual boxes, through which problems can be sent to experts, wait for a latter handling or transferred to the right enumeration area batch. Editors can change, confirm or complete data, separate or join images of pages and questionnaires, assign statuses to records or send the households to be handled by specific others.

29. Quality control of editing is performed after each editing activity by way of completeness, out-of-range and consistency checks. Failures in these checks brings back the item or create a new item to be handled by the same editor or by a senior editor. Statistical reports, generated throughout the process, draw an attention to problematic editors.

30. In the **special coding** module the three more complicated coding tasks are performed; geocoding, occupation and economic branch coding. Geocoding is operated in three address fields: the residential address, address 5 years

ago and address of work. All of which undergo keying and automatic coding. A failure in coding results in presenting the coding item to the coder to code it in a computer assisted process.

31. Preparations for occupation and economic branch coding differ from geocoding in the keying stage. Since automatic coding of these fields has not been part of the coding process, there is no need to key-in the verbal descriptions. In a computer assisted coding process the coder can make queries with words seen in the image of the questionnaire (for elaborate explanation see section “The Use of External Files”). Because of this feature, coding items are created when the optical character reader identifies written text in the relevant question fields; i.e. the mere existence of handwritten text is a trigger for creating an item.

32. The interface coder-machine is user friendly. Coders have accessibility, via a query, to a various external files: designated dictionaries, employers’ file and addresses files. They can also see the image of the relevant page of the questionnaire and images of all pages belong to the household.

33. Control of economic coding is not included in the optical data capture system, but does depend on its technology. Coding items are retrieved from the optical archive, sampled and sent for second coding in stand alone PCs. An expert receives a coding item in cases where the first code given by the ODE coder does not match the second code. The expert's code replaces the code generated by the ODE only when it is different from the first code.

34. At the end of the data capture process, a file is prepared to be sent to **the optical archive**. This file includes all data from the questionnaires, administrative data, statistics reports produced and used in the process and all audit trails of each field. This file can be retrieved and already serves us for evaluation purposes.

35. Additional **file** that is sent **to the central computer** includes the complete census information (without images) and a “tail” for each field, which enables the identification of the main characteristics of the data capture process (the status of the linkage to the external file, or a status of cancellation of the record). This tail provides the means to a precise and sensitive macro editing of the data, while preparing the final census file.

The Use of External Files

36. External files are files of different types; organizational files, registers with demographic information, files with socio-economic variables and coding dictionaries. They support diversified tasks in data processing, substitute components, supply information and supplement data. External files can also serve for evaluation purposes, given that they were not used for the process they come to evaluate.

37. Organizational files are lists referring to different units in a census; a list of all enumeration areas nested in their larger geographic units, a list of all households visited by each enumerator (enumerator’s report book),

a list of all authorized workers specified by job type etc. Organizational files can be assigned to administrative roles like coverage checks, transfers completeness checks, exclusiveness of each unit in the system and process follow-up, workers follow-up, data protection and such. They can also be used for improving data-capture quality (see below) and to generate statistical reports of any sort or kind in the system.

38. Registers with demographic data and files with socio-economic data, significantly improve work planning, coverage, data capture, editing and imputation. The Population Register in Israel served, during all censuses, for districting enumeration areas, according to the population estimates. In the 1995 Census it was also used for the production of geographic information layers in the GIS system. Coverage purposes were served by the production of self-adhesive labels carrying names and personal ID numbers, for each enumeration area. It was useful in the field, when the enumerator made sure that for each label there is an explanation whether the person was found in the expected enumeration area or not. It means that registers are very useful even if they are only partially correct. In Israel, about 70% of the population were found in the register's addresses.

39. Population Register also supported the data-capture process. Having preprinted ID numbers of 70% of the records means that 70% of the ID numbers on the questionnaires were identified with 0% substitution rate (zero mistakes in the optical recognition). Furthermore, record linkage between the register and census records, was a fundamental feature of the data capture process. It was used to substitute keying in the following process: census records were linked to the corresponding records in the register, using very rigid criteria. If the link was successful, all shared variables were compared. In a case of identity, the value was accepted as the right one and the field skipped keying completely. In this process, 60% of the characters on the questionnaire (except for alphabetical letters) skipped keying. The contribution of the register for data capture is very important, since abridging the keying step was possible not only for fields identified by the optical reader with high certainty, but also for doubtful recognition. Whenever an OCR value was supported by the register's value, it was accepted as the right one with no further treatment.

40. Data capture was also supported by the register in answers that were written in a faded color or vague handwriting. An interactive query to the register helped identifying the value (but not to substitute it with the register's value). It was the human-machine interface, having the image of the questionnaire, the corresponding ASCII value in the database, and the register's record of the same person, that allowed for such support.

41. The editing phase in the ODE system dealt with, among others, failures in the automatic record linkage between the Census and the Population Register files. By using census data for flexible queries, editors could identify the right register record among all records offered as possible to be linked to the census record. This record linkage, in a computer-assisted mode, enabled the supplement of ID numbers in records that missed this variable or had the wrong number. A verified unique ID number made possible the finding of

duplicate records and their annulment. Moreover, having a verified ID number facilitated record linkage with other files carrying the same ID variable. As a result, the process of editing census file and imputation could lean on external files that included all, or partial, census population and the variables to be edited or imputed. The external files were also used for full record imputation and as a source of information not investigated in the census questionnaire (religion and Social Security allowances).

42. Coding dictionaries are the heart of any coding process, regardless of the pattern it is conducted by. In a computerizes working environment, automatic coding is stipulated by data capture of the descriptions to be coded, whether it is done by optical recognition or by keying in the data.

43. In systems where optical recognition includes alphabetical characters, a first attempt to code can be done on the values offered by the OCR. Second attempt can be done after completing data capture phase, usually by keying. In the Israeli Census, alphabetical characters were not read by the optical reader, meaning that automatic coding could have been operated on keyed-in fields. This was the case with several variables: address 5 years ago, address of work place, country of birth (except for 7 closed categories), and relation to the reference person (except for 9 closed categories). The leading rule was to code automatically unequivocal descriptions of variables that have a finite and known number of categories. All other verbal descriptions were coded in a computer assisted coding process (CAC).

44. The crucial element of CAC process is the query to the coding dictionary. It can be done by using an already keyed-in verbal descriptions. However, images of the questionnaires are an input not to be ignored in an ODE system. Keying in verbal descriptions is error prone and moreover, keyed in description of complicated variables, such as occupation and economic activity, is not effective in CAC process. There is a need to make a query with key words rather than the whole mambo jumbo written in the designated field. Images of the questionnaire provide the mean to do so in a user-friendly interface. Coders can make a flexible query with words that were not keyed in or captured in any other data capture mode, but rather seen as part of the image. The savings in keying terms, done either by keying operator or by a coder, is enormous.

45. To summarize this point, coding dictionaries as external files should be used in an altered perception of coding process, once they are used in a technologically advanced working environment. Optimization of the whole data processing process is the better policy here too, when coding is considered.

46. After the 1995 Census in Israel, a decision was made to use the census as a source of empirical entries to occupations dictionary, to be used for automatic coding. One may think that the decision not to key-in the census entries during the data capture process has been wrong. However, the stored images of the questionnaires in the optical archive and the ASCII values of the relevant codes open the possibility to plan a parsimonious process. A query with a numeric code to the optical archive will result in a presentation of groups of images (using image-tailoring technology) that were

coded by the same code. The following activity should be keying-in unique descriptions, meaning that from all images of verbal descriptions that were coded by the same numeric code, only few will be keyed in. Never keying twice two descriptions that use the same word or combination of words ensures having all unique descriptions with a relatively low keying investment. In this case again, images that can be cut and pasted change the perception toward a selective data capture process.

The Questionnaire and the ODE System

47. The questions to be asked and their order were the only attributes of the questionnaire that were not affected by the decision to have an ODE system. All other features were stipulated by the system's abilities and disabilities.

Paper texture - The way the fibers of the paper are woven, warp and woof or only one way, influence the ink absorption (into a single spot vs. spread in an area) and therefore the readability of the written values.

Paper thickness - Too thin paper causes two main problems; double feeding during scanning and a lot of OCR noise caused by the transparency of the paper. The OCR “sees” the characters that are written on the other side of the page. Too thick paper is too heavy to carry by the enumerator, needs more space in the boxes transferred from the field to the data capture site and; jams in the scanner's feeder.

Ink density - The preprinted characters have to be clearly seen on one side of the page. If they are to be read by the respondents they should be clear to them. If they are to be identified by the OCR, they should be readable by the machine.

Colors - Colors are used for different features in the questionnaire; there are background colors, colors of text characters to be read by the respondents and colors of characters intended to be read by the optical reader.

48. On the one hand, it is preferable to use transparent colors (or non at all) all over the questionnaire, except for the characters and marks intended to be read by either the respondents or the OCR. Computer resources are always limited. If the image of a full page is to be stored, the number of pixels captured may amount to many megabytes, cause storage problems and may complicate the flow of information within the system. On the other hand, colors have to be used since the questions have to be clear and understandable and the questionnaire has to be appealing to the person who fill it in, especially if it is in a self-enumeration process.

Number of pages - Optical scanning of only one side of a page is problematic since double feeding cannot be automatically detected. A control process can be effectively activated if there is a second scanning of the other side of the page i.e. two sides of one page are necessary for scanning control. If there is a need to use several pages for one unit, each side

of a page has to have identifying variables. One should take into consideration the fact that many pages mean a higher potential to err. It prolongs the scanning process because of the number of scanning needed, and may also prolong the editing that comes to identify all structural units of the census file (questionnaire, household etc.).

Organization of the pages - Scanning has to be done on each page separately. It means that the questionnaire has to either include separate pages or to include joined pages that have to be separated before scanning. In the first case, there are problems of loosing pages, giving pages of different questionnaires to one household, using pages of the same questionnaire to different households. In the second case, person power has to be allocated for cutting and rearranging the questionnaires for scanning. The pages have to be cut well in advance in order to avoid dust in the scanners, since dust damages the scanner and also garbles the optical recognition. The margins have to be wide enough to be cut without ruining the written answers.

Graphics - There is a couple of reasons to design the graphics of the census questionnaire; to make it clear to the respondent, to support scanning process and to facilitate the optical character recognition (OCR).

49. Respondents have to identify the beginning and the end of a question, the flow of the questions, the spaces to be filled in, the places allocated for each person in the household.

50. As for scanning support, lines or frames drawn on each page serve for the alignment of the page, which is essential for opening windows for optical recognition in the exact locations on the image of the questionnaire. It is also important to align each page for exact template drop out. Here again, if the preprinted questions and marks are kept even partially, computer resources are wasted.

51. Since OCR is activated in pre defined fields, there is also a need to well define the boundaries of the recognition window, opened for the process. This window has to be flexible (the field + agreeable margins), in case people erase an answer and write above or under it, but it also have to be distinguished from the neighboring field. It means that the spaces between fields on the questionnaire have to be wide enough for a flexible OCR but not too wide in a way that it increases the number of pages needed for each questionnaire.

Closed Questions - It is much easier for an optical reader to identify a mark than a character. It is easier and more reliable to find out that something is written than what is exactly written. When planning for an optical system, it is worthwhile to close as many questions as possible without hampering their exhaustiveness and reliability.

Use of Preprinted Fields - When marks are not suitable, OCR can cope better with pre defined shapes of characters. Preprinted characters in defined font and size present a relatively small probability to err. Therefore,

it is preferred to print any previously known variable. For example, questionnaire number, number of page, number of record. At times, it is preferred to print the values on adhesive labels, since its use is stipulated by unpredictable factors. For example, a number used to combine two questionnaires for one household, should be printed on a label because its use is found out during the enumeration process and not before it. Enumeration area number can be preprinted on the questionnaire but it leads to a waste of questionnaires that were not used in the designated enumeration area. Personal ID number on an adhesive label will be attached to the questionnaire if the person has been found in the expected address, etc.

Adding "System" Variables - The need to identify each physical and logical unit is more intensified in a computerized working environment. In the ODE system, variables added to define one page (for the right drop off and OCR processes), two sides of the same page, all pages of the same questionnaire, all questionnaires of the same household, all households in same enumeration area and all parts of each individual record.

Fill-in Instructions - In a time of growing reluctance to cooperate with census takers, questionnaires have to be clear and fill-in instructions have to be simple and not too many. The burden of adjusting the questionnaire to an optical reading system has to be the census bureaus' responsibility and not the respondents'. If interviews by an enumerator are the way the census is conducted, instructions can be extended. The decision whether to use capital or free-flow letters when filling in the questionnaire is a derivative of the enumerators and respondents cooperation. Capital letters contribute to the optical recognition ability and reliability, but writing capital letters requires more time and self-discipline. Hence, when it is reasonable to expect to receive the answers in the desired capital letters, it is worthwhile to ask for it. When the requirement to fill in the questionnaire with capital letters can further hamper the cooperation with the respondents, planners have to find a different strategy.

52. Cooperation, discipline and the way people are writing are culture dependent, and thus have to be decided upon locally and not as a global decision. In the Israeli case, the expectation of getting answers in capital letters is not realistic, however, the Hebrew language is written in separate letters and can be identified easier than English. In the 1995 Census, alphabetical letters were not identified by the OCR at all, because of a negative cost-benefit analysis, yet, free flow numeric characters were.

Concluding Remarks

53. Our conclusion is that the optimal way to recruit new technologies for census purposes is by rethinking and re-enunciating census processes and intermediate objectives. New technologies have already proved, and will probably prove again in the coming censuses, to be effective and beneficial to census processes, utilization of resources and to the quality of census products.

54. However, non of the new technologies can be automatically adopted, without ongoing quality checks and adjustments to ensure the required quality level. Continuous quality control throughout the process is a must that can be relatively easy to accomplish in an image-ASCII working environment.

55. Computerization of processes allows for a tremendous reduction of human involvement, yet machine is not a full substitute of people. Some of the census tasks are to be handled in a human-machine interface, usually in a computer-assisted procedure.

56. The future may hold the seeds of new technological developments. However, there are still technological improvements to be made in census planning, using existing technologies. The most prominent one is the integration of additional census processes, starting from data collection up to the preparation of census products, in these advanced technological systems. The ODE system, that has been developed for the Israeli 1995 Census, aimed to include data capture tasks. However, its characteristics had a positive influence on census field-work performed before, and on data-processing and data-analysis performed after data capture. A systematic and careful planning of an integrated system will further enhance the possibilities to exploit the potential of the already available technological tools.

57. Further information regarding the ODE System, used for the 1995 Census of Population and Housing in Israel, **see in the following publications**:

Blum, Olivia "Editing Definition and Operation in the Optical Data Entry System (ODE) of the 1995 Census of Population in Israel". Working paper No. 48. SCECE Work Session on Statistical Data Editing. Athens. Nov. 1995.
Blum, Olivia "Evaluation of Data-Editing Using Administrative Records". Working paper No. 16 SCECE Work Session on Statistical Data Editing. Prague. Oct. 1997.
Blum Olivia & Eliahu Ben-Moshe "Automated Record Linkage and Editing: Essential Supporting Components in Data Capture Process". In Statistical Policy Working Paper No. 25. Data Editing Workshop and Exposition. Executive Office of the President of the United States. Washington DC. December 1996.
Nathan Gad & Israel Givol "The ODE (Optical Data Entry) Experience in Israel". Paper presented at InterCasic '96', San Antonio, Texas. December 1996.

The following papers were published in New Technologies for the 2000 Census Round. Euro-Mediterranean Workshop in Israel. March 1997:
Blum, Olivia "The ODE System: Logical Structure and Guiding Principles".
Blum, Olivia " Keying Module".
Blum, Olivia "Editing and Coding Module".
Givol, Israel "The ODE System: Technology and Human Engineering.
Kagan, Oren "The World of Scanning".
Khomenko, Yuri "ODE Telecomputing Systems".
Yitzhaki, Ruhama "The Operational System of the ODE".