



Economic and Social Council

Distr.: General
26 March 2012

English only

Economic Commission for Europe

Conference of European Statisticians

Group of Experts on Population and Housing Censuses

Fourteenth Meeting

Geneva, 24-25 May 2012

Item 4 of the provisional agenda

Address/dwelling listing

Dwelling and Building Register Based on Municipal Taxation List – Quality and Distinctiveness

Note by the Central Bureau of Statistics, Israel¹

I. Back ground

1. Israel conducted six censuses since its establishment in May 1948. The first census which was carried out in November 1948 was conducted under curfew and was used as the base for the Central Population Register (CPR). The four census that followed were classic censuses in which the whole country's area was surveyed, all possible dwelling facilities were enumerated and a demographic short questionnaire was completed by the interviewer for 100% of residents. A short socio-economic questionnaire was completed for one fifth of the households. All five first censuses were conducted using a face to face paper questionnaire.

2. The sixth census, in the round of the 2010 censuses, conducted in December 27, 2008 was a innovative census – Integrative Census (IC). It was a register based census complimented by two sample surveys which were aimed to evaluate the quality of the register and to supply parameters for calculating coefficients to correct improper registration in CPR.

http://www.cbs.gov.il/census/census/pnimi_sub_page_e.html?id_topic=2&id_subtopic=1

3. There were numerous incentives that initiated the change in the methodology of the IC.

¹ Prepared by Pnina Zadka.

- a) Moving toward a computerized census -. A computerized census was expected to speed up the production of census results. The 1995 census although using OCR techniques, still required a long time to produce demographic results and even more for the socio-economic results to be produced and published. This lag was not acceptable by users, at time of extensive use of computerized data collection techniques available at the onset of the third millennium.
 - b) Reduce census cost - conducting a traditional census using computerized interview was expected to be such that would not be approved by the budget authorities
 - c) Register based census - the integrated census was expected to be used as a first step toward a full register based census. The need to obtain, explore and evaluate the registers available in the country for the Integrated Census were expected to make possible the conduction of register based census in the future
 - d) Reduce response burden on the public - the increase in the number of surveys conducted by the Central Bureau of Statistics (CBS) as well as by other governmental, non-profit and for-profit organization induced response burden on the public, which was expected to increase refusals to census response as well as to other CBS surveys.
4. The execution of the 2008 IC as well as the exploration and evaluation of the available registers, revealed, that the quality of some of the registers in the country was not yet sufficient for a full register based census. Therefore surveys were needed to produce coefficients to correct the registers.
5. It also became obvious that the IC, although reducing the cost substantively, still required extensive organizational, logistic and human resources for the sample surveys implementation.
6. These outcomes of the 2008 IC, initiated the search for ways to improve the IC methodology to overcome these deficiencies in the upcoming censuses.
7. For the 2020 round of censuses Israel is modified the method of the IC into the **Integrated Rolling Census** (IRC). The first round of the updated methodology was launched in January 2012.
8. The IRC methodology is a modification of the IC methodology. A register based census with a **sample of statistical areas** (census tracts) within localities. The sample size was reduced based on statistical tests that showed that the variances were improved insignificantly with the samples being over 10% of the households (instead of 17% in 2008). The two sample surveys; the under-coverage census survey and over-coverage census survey combined with current household surveys such as the Labor Force Survey. The current household survey will be added for the estimation at the locality level.
9. As the proportion of households sampled to the various surveys conducted by the CBS annually has increase a reached almost 5% of the households, the chance of a household being sampled to more than of the surveys was unacceptable and called for the harmonization of the sampling frames for all current surveys. The route chosen for harmonization was the construction of the Dwelling and Building Register. A dwelling is perceived as a good proxy for a household.

II. Construction of the Dwellings and Buildings Register

10. The objectives of the Dwelling and Building Register were defined in vision of the various potential uses within the CBS as well as for research by external organization. The main objectives and uses of the register are:

- a) Sampling frame for household surveys;
- b) Improving the construction of administrative household;
- c) Creating the infrastructure for register based censuses in the future;
- d) Linking with other registers such as the Population register, The business register, personal income tax register, real-estate purchase tax file, etc.

11. Different administrative sources were investigated as potential data sources for the construction of the Dwelling and Building Register. Municipal taxation list were found to supply the most reliable and near complete data source.

12. Municipal taxation list were used for over 40 years as the sampling frames for the labor Force Survey and for the Household Expenditure Survey. They were used as separate lists and the samples were drawn from each municipality separately. For the construction of the Dwelling and Building Register the taxation lists of **all** municipalities in Israel will be pooled, standardized and harmonized.

13. Currently 200 urban municipalities (about 90% of the population) are included in the register. About 1000 rural localities grouped into 56 Regional Councils (rural municipalities), which were not use in previous years, are in the process of being incorporated into the dwelling and Building Register

14. The register is considered a "statistical register" meaning it would serve statistical purposes including research but would not serve as a source for providing services by government or municipal authorities. Hence, insignificant accuracy is permitted as long as it has no statistical effects on distributions and statistical parameters.

15. All real-estate properties were included in the register primarily, as the objective of the register was to expand the use beyond sampling frame. Such uses are linking the register with various data sources and secondary as a precaution for misclassification by municipalities of properties into dwelling or non-dwelling purposes.

16. The first steps in the construction of the Dwelling and Building Register were re-evaluation of the quality of variables in the taxation list that were provided in the past. The evaluations process that was used, involved extensive acceptance tests which were performed in two stages: Global macro-level tests and Individual municipality micro-level tests.

- a) Completeness – Completeness was checked on two levels, first on the macro-level and then on a micro-level;
 - i. Each locality's list was checked for completeness by comparing it to the taxation lists received in previous years. The assumption was that as there are very few demolitions of dwellings, number of dwellings and other buildings in a given year should not be less than that of the previous year;
 - ii. The next macro- level check for completeness was against the number of households estimated in the 2008 census. There were less than 0.5% of multi-households dwellings estimated in the 2008 census. Hence, the number of dwellings in the 2010 taxation list

should not be less than the total number of dwellings found in the census in each locality;

- iii. The micro-level checks for completeness were performed by checking the property identifiers. The taxation lists include a property identifier which is expected to be unique within each locality. These identifiers for two consecutive years were compared and checked.
- b) Data availability – All variables that were flagged as essential variables were a pre-requisition to be considered, if the list missed one of these variables, the municipality was asked to complete the missing variable. The values of each variable were then checked to assure that all values meaning were in accordance with the codebook provided with the file.
- c) Validity – The addresses and property type codes were checked against the standard address list maintained by the CBS, including synonymous names for streets. Property types were checked for minimum square meters of the property, e.g. the number of rooms and the size of the property, minimum floor size to be considered a dwelling according to the regulation requirements. Uniqueness of property identifier - no duplications within the same locality.
- d) Reliability – the reliability was mainly checked against other data sources such as census, surveys and administrative data obtained from other sources, such as billing list etc.

17. As each locality is free to maintain its taxation list to meet the specific operational needs of the locality, there is no central regulation on the structure and content of the individual lists. The next phase was therefore, transformation to a standard database structure and a standard codebook.

18. As mentioned previously the decision taken was that the Dwelling and Building Register will be a statistical register, to comply with that definition several procedures followed:

- a) Editing – transforming values to the standard set for each variable, including the transformation of un-acceptable values into 'unknown'. Locality name, street names and house numbers were divided into separate variables. And street names were modified to the standard name structure.
- b) Imputation – imputing missing information according to a set of rules and adding information from other sources such as the censuses and other available registers.
- c) Geo-coding – The municipalities do not geo-code the taxation list. The CBS is the government authority responsible for the national street codes for maintaining and updating the street codes for the government databases. The codes contain synonymous street names and all eligible street names. Each street name has a unique code. Municipalities are not obliged to have street names for all the streets in the locality. The taxation list may have also informal names or signs such as letters and numbers. The first step in the geo-coding process is assigning the street code for each address. In order to geo-code a dwelling it should have a street code and a house number. Nevertheless formal street names were not a pre-requisition in the acceptance process. The geo-coding procedure was performed in a hierarchical resolution mode: buildings, "Statistical Areas" (census tract) and municipalities/localities. Addresses that complied with the standard full addresses (formal street name and house number) were automatically geo-coded, and assigned the XY coordinates of the building center in the GIS computerized

buildings layer as well as the unique ID number of the building. "Locations" such as a building name or a central point with no street names or street number were geo-coded manually. Addresses with an eligible street name but with no house number or without eligible street number were geo-coded to a point in the center of the street or to a point close to an eligible house number. Addresses with no street code were geo-coded to the center point of the locality. Wherever available sub-blocks were geo-coded to the geometric center of the sub-block or block if the sub-block was unavailable. For the final geo-code for a dwelling we constructed a preference rule by each locality. In fact geo-coding of the taxation list was one of the main procedures in the detection of flaws and inaccuracies in the lists.

- d) Standardization – standardizing the dictionaries and formats of all the variables for all municipalities.
- e) Harmonization – harmonization of the variable formats, names and values with the standards used in other data sources at the CBS and/or government organizations.

III. Census estimates vs. Labor Force estimates

19. The Labor Force is a current household survey based on a sample of dwellings drawn from municipal taxation lists. This sample frame was used for several decades. Traditionally the estimated labor force number of households was about 10% less than the number of households estimated by the censuses. This was also the case in the 2008 census.

Number of households and average size of household

	Census 2008	Labor Force survey 2008
Number of households (thousands)	2, 314.1	2, 087.4
Average household size	3.2	3.3

20. This discrepancy was always considered a flaw of the census, blamed on compensation methods and on inexperience of census enumerators. One other discrepancy that was found was the average household size, larger household size in the labor force survey. Comparing the distribution by household size, as expected, there were fewer households of size 1 and 2 in the labor force survey as compared to the census.

21. With the decision to base the IRC surveys on samples drawn from the taxation list a special investigation was carried out in 2010. The data from the labor force survey was compared on locality and "statistical are" basis, although the investigation was not fully completed yet, intermediate results hint to some deficiencies in the municipal taxation list relating to under- registration of small dwellings in the taxation lists,

22. The first step to estimate the magnitude of the problem was to compare the Dwelling and Building Register with the 2008 census. The comparison of the two databases was aimed to expose possible deficiencies, discrepancies and duplications in the dwelling and buildings register.

IV. Pitfalls in a Dwelling and Houses register based on municipal taxation list

23. The comparison revealed possible pitfalls in the register which can be classified into the following main groups:

- a) Addresses: some of the addresses were modified over the years by the municipality, but were not updated in the taxation list. Some addresses were mailing addresses such as post office boxes numbers, or the address of the dwelling owner not residing in the apartments rather than dwelling addresses. Inconsistency between the address and the cadastre blocks and sub-blocks. As mentioned the cadastre blocks and sub-blocks (the sub-blocks is expected to be unique for each building) were geo-coded based on the cadastre layer. When the geo-codes of the buildings by address were compared to the geo-codes by the blocks, discrepancies were detected. This called for a decision algorithm for preference of the correct geo-code. The geo-code is essential for the classification of addresses into statistical areas. Since it is a pre-requisition for a dwelling to be in the sampling frame for IRC, is, to have the correct statistical area assignment, as the statistical area is the primary sampling unit. This is also the case for the Labor force survey sample which should be evenly distributed over the locality area, representing the different neighborhoods of the locality. Some dwellings were registered before street names were assigned and were not modified since, specifically dwellings which did not change owners for an extended period. In some localities there are blocks and sub-blocks and there are no street names at all or they are missing in some quarters of the city.
- b) Split dwellings and merged dwellings not updated: the rise in dwellings cost as well as the cost of rent in recent years, initiated the phenomenon of splitting large apartment or houses into smaller dwelling units without notifying the local authorities. A similar phenomenon which expanded in the last decade is the transformation of basement floors into separate dwelling units. On the other side is the merging of small dwellings into larger dwelling
- c) Un-registered dwelling units – this refers to two types of dwellings. Very old buildings in which the division into apartments was performed at later stages and was not updated in the municipality. On the other hand are recently completed apartment not yet registered at the time data was obtained. In general dwelling that are exempted from municipal taxes are potential for under-registration
- d) Inaccurate classification of buildings and flats: Another phenomenon which follows the recent fashion of transforming industrial and commercial facilities into dwellings. These transformations are sometimes not updated in the register, causing distortions in the register. This is specifically occurring in the rural localities, where agriculture is no longer the main income source and agricultural buildings are modified into housing these are not registered as dwelling units and are omitted from the sampling process. The shortage of office space in the centers of cities caused the modification of dwelling units into offices, without updating the use in the municipality.
- e) Not updated ownership and occupancy – In cases where the address information on the owner is used to find the location of the dwelling un-updated ownership results in missing dwellings.
- f) Unique identification of dwelling units -The register is intended to be annually updated, for these to be possible the identification of dwelling units should be unique and constant over years. This is not always the case currently in some municipalities. One of the incentives used for the municipalities to comply with the requirement of the register was to offer them receiving the inaccuracies found in their without revealing the incorrect records.

V. Future planes

24. To validate the data in the register as well as validating the enumeration of the correct sampled dwelling. Interviewers are to be equipped with a GPS which will be used to obtain spatial XY coordinates for the building that was enumerated. The XY coordinates will be compared to those based on the original location by the address or sub-block in the register. Discrepancies between the original and the interviewers XY of over 30 meters will be re-checked by a qualified GIS person. This procedure is expected to improve the geo-coding in the Register. The information will then be fed back to municipalities to encourage them to improve the list. An experiment of this kind was started in the second quarter of 2012.

25. The register will also be used for research on taxation and dwellings cost and value, research will most probably reveal more breaches in the register which were not identified in the validation process.

26. Municipalities will be offered to form partnership schemes where the CBS will feed back individual cases with improved data which will be integrated into the municipality taxation file. Hence the data obtained in the future will need less editing and imputations.
