# UNITED NATIONS ECONOMIC COMMISSION FOR EUROPE CONFERENCE OF EUROPEAN STATISTICIANS

#### STATISTICAL OFFICE OF THE EUROPEAN UNION

Joint UNECE/Eurostat Expert Group Meeting on Register-Based Censuses (The Hague, The Netherlands, 10-11 May 2010)

Working paper 3 28 April 2010

Topic (i ) of the provisional agenda

# GOOD PRACTICES AND TAXONOMY

Using Population Registers as a Population List: Some Results form the Pilot Survey 2009

Note by the National Institute of Statistics, Italy\*

#### I. TOWARD TO THE NEW POPULATION CENSUS 2011

- As in many European countries, also in Italy, the National Institute of Statistics (ISTAT) will manage the transition from the *traditional census* to the *census assisted by list* in the forthcoming Census of Population and Housing. Even if traditional objective information still holds, the next census will be made with a completely new strategy. Compared to traditional census, the new strategy will maintain the characteristics of exhaustiveness and simultaneousness of collection throughout the national territory, but it will also introduce the following general innovations:
  - i) the transition to a census based in significant measure on the use list of individuals, households and cohabitations from the local population register (named *anagrafe*) for mailing questionnaires in order to reduce the number of enumerators and to simplify the work of front office of the Municipal Offices of Census;
  - ii) the questionnaires collection by a plurality of channels and according to the preferences of the respondents (web, mail, municipal centres of collection);
  - iii) the creation of a web system which will support the Municipalities for surveying and the involved territorial offices in monitoring;
  - iv) the recovering of units non collected in the Census throughout a sample survey contextual to the Census survey and to the comparison Census-*anagrafe*;
  - v) major and immediate coherence between population register data and census data.
- 2 The new Census will be based, therefore, on a number of innovations in methods and techniques that will be combined together in order to obtain high quality in results, contain the burden of Municipalities, reduce

<sup>\*</sup> Paper prepared by Anna Pezone, Population and Housing Census, Division of Census and Emanuela Scavalli Methods, Techniques, Organization of Census, Division of Census.

the statistical burden on respondents and reduce the time of release of the data as required by the new European regulation. The main innovative element, which will be the focus of this paper, is the introduction of the Municipal *anagrafe* list (MAL) as informative basis for ISTAT at single Municipalities. On the basis of MALs ISTAT will attend to:

- i) acquire and archive data in protected databases
- ii) make controls of coherence and correctness of variables and, whenever possible, rectify data
- iii) normalise and geocodify the addresses assigning the enumeration area codes
- iv) sample the units-households (33% of total) to which submit the long form questionnaire; this will involve only the census area where two types of questionnaire (short and long form) will be used
- v) define the length of the questionnaire on the basis of the number of components; as in 2011, in fact, two different household size will be prepared for the two types of questionnaires
- vi) assign a univocal code to each household which will be pre-printed on each questionnaire
- vii) generate passwords for compiling on line the questionnaire
- viii) arrange the single municipal files for the printing and mailing of questionnaires
- ix) load data on the System for Survey Management which will be used during the Census by different actors for monitoring and operational management.

# II. USE OF ADMINISTRATIVE DATA FOR STATISTICAL PURPOSES

- 3 The main innovation of the Census is, therefore, based on the introduction of an administrative source, the MAL, in the statistical process of the Census. In general, the use of administrative sources for statistical purposes leads advantages and disadvantages which must been adequately evaluated in order to have a more efficient process.
- 4 The administrative sources, generally, are organised to meet the informative and informational needs of the administrations to which belongs: for this reason, different administration could decide to collect and organise the same data with different modalities. Some administrations collect all population data of those which applies to them, creating in this way a non exhaustive list of the Italian population. Some variables, which apparently seem to be equal in different sources, often are based on different definitions, different classifications or different temporal reference. The integration some time is difficult also because sources are not computerised but on paper, or they could use different software for storing data; therefore, it is necessary to develop *ad hoc* programs which read different formats of sources and transform them in the standard format defined universally.
- 5 For these reasons, the comparison of data among different sources frequently is difficult, also for the lack of a univocal key variable among units. Make a statistical use of administrative sources means, therefore, harmonise definitions, classifications and temporal reference of the data of the different sources in order to make them integrable and comparable among them. For making this, it is necessary to define a standard of definitions and classifications and, therefore, to adapt all sources to these standard.
- Obviously, the use of administrative sources has many advantages, as the immediate availability of data already existing, the usability for checking results coming from other sources, the savings in terms of costs compared to conduct sample surveys or ad hoc censuses, the reduction of statistical burden on respondents. On the other hand, disadvantages in using of administrative sources are due to the fact that data are often incomplete or inaccurate, because not of primary interest for the administration, there are different levels of aggregation not compatible with those necessary for statistical purposes, and problems of confidentiality for the access to some typology of data could exist. Frequently, administrative data are used in statistics because they provide the reference population for sampling surveys, they allow to develop, update and spread statistical series in socio-economic sectors of economy, they furnish a support for policy/decision making, they help in monitoring the progress in public policies, they are useful in research.

# III. THE MUNICIPAL ANAGRAFE LISTS (MALs)

- 7 The *Regolamento Anagrafico*<sup>1</sup> provides that citizens who take up residence in a Municipality must communicate it to the *anagrafe* office, indicating the Municipality of origin from which the individual has to be cancelled. The Italian *anagrafe* are organised at level of Municipality. Being present in the Italian territory about 8,100 Municipalities, there are as many archives locally managed often with standards different from each other. Although the *Regolamento Anagrafico* sets standards for such archives, in practice, applying it is still very far from reality, mainly because the process of computerization of these archives was long and, then, their subsequent harmonization to the regulation is still being implemented.
- 8 As far as the research on territory of resident individuals, it has been noted that in Italy, as probably in other countries, there is the problem of residences *de iure* vs. *de facto*. Some times, dwellings which results occupied in reality are not as, for example, they allow to pay less taxes or get tax relief. This implies an overestimate of the number of households and an increasing of the risk of double compilations of questionnaires.
- 9 Even if the use o administrative source for statistical purpose poses some problems, in Italy these are further compounded by the fact that, unlike other European countries, a single centralized archive of the population does not exist but there are about 8,100 archives. Therefore, the new census system foreseen that ISTAT acquires, well in advance of the date of the Census, the households and cohabitation cards contained in the *anagrafe* of each Municipality, posing all the problems of integration of many different administrative sources.
- 10 At present, however, ISTAT does not have, for all Italian registers, detailed information on the quality, comprehensiveness and accuracy of demographic information, technical and computerised equipment and of the ability to adapt them to given standards.
- 11 This differentiation and fragmentation has also strongly influenced the procedures being followed for the household surveys conducted by ISTAT. In fact, by failing the availability of a single store but having 8,100 databases managed separately by the respective Municipalities, the samples of the major household surveys have a two-stage sampling design, where Municipalities are the units of the first stage: this involves a loss of efficiency of the estimates, which, as they are structured now registers Italy, can be overcome only when a single national archive will be available and from which ISTAT will proceed directly to the selection of households.
- 12 To conduct the census operations the MALs sent by the Municipalities should include all the same information and data should be provided according to technical standards defined by ISTAT. ISTAT, therefore, will have to make operations of standardization and normalization of data so that they can be used for printing and mailing questionnaires and for populating the database helping the Municipalities in the fieldwork and the involved regional bodies for monitoring. The respect of standards from Municipalities will represent the operation core for acquiring data.
- 13 The complexity of operations of acquisition and processing of MALs requires planning well in advance the activities related to them as these operations are not free from various kinds of problems. For these reasons, the date by which the Municipalities must send the MALs to ISTAT (01/01/2011) was chosen so as

<sup>&</sup>lt;sup>1</sup> The Italian law on population registers. (D.P.R. n.223 del 30 maggio 1989), Istat, Metodi e Norme serie B – n.29 1992.

to harmonize the technical requirements and objectives of all actors and in order not to be too much prior to the date of the Census (to prevent the high number of demographic events which imply population variations - inter-and intra-municipal transfers and deaths), nor too close to it (which makes it difficult to carry out preparatory operations listed above).

14 The experiences gained so far suggest, before Census, additional preliminary checks on quality, completeness and accuracy of information contained in the MALs without leaving the adaptability of local demographic services to given standards. Basically it is necessary to know in advance which will be the possible problems to be encountered in the integration and harmonization in order to prepare in advance the appropriate software and solutions. ISTAT with the Pilot Survey in 2009 has tried to extend its information assets by testing all aspects of the new census strategy. In this work the focus will be on what is arisen on the problems associated with the use of MALs acquired from a sample of 31 Municipalities of different sizes and located throughout the Italian territory. Solutions and tools that will be taken to improve the process of acquiring data will also highlight.

# IV. THE MALS ACQUISITION IN THE PILOT SURVEY 2009

- 15 Through the Pilot Survey in 2009 ISTAT has began to test a process of acquisition of MALs. It has been requested to 31 sampled Municipalities to do the following preliminary operations:
  - a) appoint and communicate by fax data on the technical responsible for the survey;
  - b) send to ISTAT until 30 may 2009 the MALs on order to allow:
    - i) the standardisation of the variable address;
    - *ii*) the selection of a sample of households to which submit the different questionnaires to be tested (short/medium/long form);
    - iii) the preparation of the plan of mailing;
    - iv) the printing of the households questionnaires.
- 16 In the Pilot Survey the reference unit for the selection of data from the MALs has been the set of individuals which live in household or cohabitation which at 31/12/2008 have the residence in the municipal territory. Even if the individuals living in cohabitations were not been object of the Pilot Survey, it was considered appropriate to include them in order to simulate as fully the operations of data extraction for the 2011 Census and to better understand the ways in which the Municipalities record information on their residents.
- 17 The Municipalities were provided with all technical specifications necessary for the preparation and dispatch of the file. The electronic data should be sent through a secure protocol and according predefined standards relating to:
  - i) name, structure and format of data files (.csv separated by ";")
  - ii) features, format and length of variables
  - iii) name, structure and format of files for decoding variables "relationship" and "marital status" if recorded with numeric or alphanumeric codes and not with alphabetic descriptions.
- 18 The requested record format contained 27 variables (see table 1), each of which responds to different purposes: the identification of the collection unit of analysis (family variables and individual variables), the management of printing and mailing of the questionnaires, the verification of the quality of the data and the computerized control of the files received by ISTAT.

Table 1. Variables of the record format for purpose

| Variables   | Family<br>Var. | Individual Var. | Var. for survey | Var. for internal check | Var. for file<br>management |
|---|----------------|-----------------|-----------------|-------------------------|-----------------------------|
| Province Code                                       |                |                 |                 |                         |                             |
| Municipality Code                                   |                |                 |                 |                         |                             |
| Residence in household or cohabitation              |                |                 |                 |                         |                             |
| Household Code                                      |                |                 |                 |                         |                             |
| Cohabitation Code                                   |                |                 |                 |                         |                             |
| Individual identifier                               |                |                 |                 |                         |                             |
| Surname   |                |                 |                 |                         |                             |
| Name  |                |                 |                 |                         |                             |
| Tax code  |                |                 |                 |                         |                             |
| Sex   |                |                 |                 |                         |                             |
| Date of birth                                       |                |                 |                 |                         |                             |
| Province Code of birth                              |                |                 |                 |                         |                             |
| Municipality Code of birth                          |                |                 |                 |                         |                             |
| Code of foreign country of birth                    |                |                 |                 |                         |                             |
| Citizenship   |                |                 |                 |                         |                             |
| N° components of household/cohabitation             |                |                 |                 |                         |                             |
| Relationship  |                |                 |                 |                         |                             |
| Marital status                                      |                |                 |                 |                         |                             |
| Date of registration in anagrafe                    |                |                 |                 |                         |                             |
| Toponym Identifier                                  |                |                 |                 |                         |                             |
| Type of residence address (street, square, avenue,) |                |                 |                 |                         |                             |
| Denomination of address of residence                |                |                 |                 |                         |                             |
| Civic number of residence                           |                |                 |                 |                         |                             |
| Exponent of civic number                            |                |                 |                 |                         |                             |
| Internal number of residence                        |                |                 |                 |                         |                             |
| Zip code of residence                               |                |                 |                 |                         |                             |
| Enumeration area code                               |                |                 |                 |                         |                             |

- 19 The *tax code* is the univocal code attributed by the Ministry of Finance to each individual combining letters and numbers on the basis of surname, name, sex, date and place of birth. It is generated through an algorithm of which the Ministry is owner and which having the complete archive, can identify the coincidence of codes and manage the particular cases. The tax code is undoubtedly the key variable of the MAL because it allows to make different operations as:
  - i) make crosscheck, at individual level, of furnished demographic data on variables which compose it
  - ii) fill in the on-line questionnaire because it represents the *userid* for accessing to the web compilation
  - iii) make post-census checks because it represent the primary key of linkage between data contained in MAL and data surveyed for the reference person of the family card.
- 20 The *type of residence* (household or cohabitation) and *citizenship* represent two very important variables because the first allows to assign different types of questionnaires to the households while the second will

enable to furnish the language manuals for helping the filling of questionnaires for foreign citizens present in the Italian territory and for increasing response rates of this important component of the population.

- 21 The *relationship*, variable necessary for identifying the reference person of the household or cohabitation, and the *marital status* has been requested also for probing the modalities coded by Municipalities imposing them to give in any case the relative decode. These variables have enabled to assess that Municipalities are following their own criteria for storage, alphabetic description or numerical and alphanumerical codes each of then with a different number of modalities.
- The set of information relative to the localisation on territory of individuals (identifier of toponym, type of residence address, residence address Denomination, Civic number of residence, Exponent of civic number, ZIP code of residence and enumeration area code) was required to allow the mailing of the questionnaires. The addresses have been submitted to processes of standardization and geocoding for the allocation of the enumeration area code; that information was still required to Municipalities as auxiliary variables, where the software fails the operation.
- 23 In summary, all the variables included in the record format required to the Municipalities for the extraction of MAL, have played many features to help to achieve the following objectives:
  - i) the execution of the Pilot Survey 2009;
  - ii) the acquisition of an informative heritage essential to better design the method of operation of the next Census of Population and Housing;
  - iii) the assessment of levels of quality and accuracy of data contained in the MAL and of criteria for registration and classification used by Municipalities for the required variables.
- To send ISTAT the MALs it has been used the official site for acquisition on which there is a procedure developed to allow file sharing between agencies or outside companies and ISTAT safely and reliably in line with legislation for the protection of sensitive data. Since only 31 Municipalities were involved, the procedure did not make any automatic control on compliance with technical standards. All inspections were performed after the transfer and, in case of more relevant and not remediable problems, the Municipality has been invited to undertake a new and correct transmission of data. These limits will be exceeded later by using a different application, already developed, that will allow to control the standard at the moment of loading MAL by the Municipalities.

# V. FIRST RESULTS: PROBLEMS AND QUALITY OF DATA

- 25 From the analysis conducted on MALs sent from the 31 Municipalities involved in the Pilot Survey in 2009 several errors were found. These can be divided into two macro-types:
  - 1. Formal Errors due to the non respect of the requested technical standards;
  - 2. Substantial Errors due to recording errors.

It is interest of principal actors of the census operation that both types of errors are minimized as well the compliance with *Regolamento Anagrafico* have to be maximize: from **ISTAT** in quality of owner, coordinator and guarantor of developments in the Census; from **Municipalities** in quality of actors active on territory and for which the good quality and a proper timing of data collection ensures a product of instant usability and suitable to standards of the informative demand.

Another problem of acquiring MAL for Census is the standardisation of some variables (sex, marital status, relationship and type of residence) for which each Municipality gives a proper classification because no standard code is foreseen from the Regolamento Anagrafico. It will be necessary to monitor and verify the code of variables foreign country of birth and citizenship which Municipalities do not respect even if they are

constrained to the ISTAT code for the *Regolamento Anagrafico*. The first general analysis that was conducted on data concerns the comparison between the number of individuals and households extracted from the MALs on 31/12/2008 and ISTAT demographic balance at the same date.

Table 2 – Individuals and households in MALs and in the ISTAT demographic balance at 31/12/2008

in the Municipalities of the Pilot Survey

| •                             |                    | М  | AT              |  | Demogran            | hic halanca       | A beolute o          | lifforences         |
|-------------------------------|--------------------|--|-----------------|--|---------------------|-------------------|----------------------|---------------------|
|                               | MAL                |  |                 |  | Demographic balance |                   | Absolute differences |                     |
| Municipality                  | Individuals<br>(A) | Reference<br>person<br>of family<br>card | Household codes | Absolute<br>difference<br>(ref. person<br>and<br>households) | Individuals<br>(D)  | Households<br>(E) | Individuals<br>(A-D) | Households<br>(C-E) |
|                               |                    | (B)                                      | (C)             | (B-C)  |                     |                   |                      |                     |
| Abbiategrasso                 | 31.145             | 13.206                                   | 13.207          | -1   | 31.146              | 13.206            | -1                   | 1                   |
| Alba                          | 30.993             | 13.641                                   | 13.642          | -1   | 30.994              | 13.633            | -1                   | 9                   |
| Alghero                       | 43.836             | 19.710                                   | 19.660          | 50   | 40.887              | 19.659            | 2.949                | 1                   |
| Appiano sulla strada del vino | 13.892             | 5.786                                    | 5.787           | -1   | 13.892              | 5.792             | 0                    | -5                  |
| Aprilia                       | 67.949             | 26.968                                   | 27.012          | -44  | 68.587              | 27.045            | -638                 | -33                 |
| Belmonte Mezzagno             | 10.948             | 3.576                                    | 3.558           | 18   | 10.872              | 3.561             | 76                   | -3                  |
| Brindisi                      | 89.767             | 34.736                                   | 34.702          | 34   | 89.691              | 34.676            | 76                   | 26                  |
| Cantalupo nel Sannio          | 766                | 372                                      | 373             | 1  | 765                 | 373               | 1                    | 0                   |
| Challand-Saint-Anselme        | 733                | 355                                      | 355             | 0  | 733                 | 354               | 0                    | 1                   |
| Città di Castello             | 40.303             | 16.277                                   | 16.277          | 0  | 40.303              | 16.288            | 0                    | -11                 |
| Genova                        | 611.204            | 300.704                                  | 300.705         | -1   | 611.171             | 300.708           | 33                   | -3                  |
| Latina                        | 128.795            | 51.130                                   | 51.242          | -112   | 117.149             | 52.553            | 11.646               | -1.311              |
| Marano Principato             | 3.014              | 1.036                                    | 1.036           | 0  | 3.005               | 1.002             | 9                    | 34                  |
| Messina                       | 247.216            | 101.848                                  | 101.849         | -1   | 243.381             | 101.851           | 3.835                | -2                  |
| Naples                        | 1.016.553          | 376.765                                  | 374.520         | 2.245  | 963.661             | 372.142           | 52.892               | 2.378               |
| Nimis                         | 2.842              | 1.221                                    | 1.222           | -1   | 2.860               | 1.217             | -18                  | 5                   |
| Novara                        | 103.037            | 46.408                                   | 46.369          | 39   | 103.602             | 46.305            | -565                 | 64                  |
| Palermo                       | 702.307            | 262.868                                  | 262.877         | -9   | 659.433             | 255.358           | 42.874               | 7.519               |
| Parma                         | 182.389            | 87.275                                   | 87.261          | 14   | 182.389             | 87.275            | 0                    | -14                 |
| Passignano sul Trasimeno      | 5.699              | 2.358                                    | 2.355           | 3  | 5.673               | 2.353             | 26                   | 2                   |
| Perugia                       | 166.005            | 69.904                                   | 69.837          | 67   | 165.207             | 69.665            | 798                  | 172                 |
| Porto Sant'Elpidio            | 24.790             | 9.853                                    | 9.860           | -7   | 25.118              | 9.789             | -328                 | 71                  |
| Prato                         | 186.608            | 74.459                                   | 74.459          | 0  | 185.091             | 73.948            | 1.517                | 511                 |
| Rome                          | 2.838.811          | 1.326.361                                | 1.323.555       | 2.806  | 2.724.347           | 1.099.000         | 114.464              | 224.555             |
| Rovereto                      | 37.053             | 16.191                                   | 16.191          | 0  | 37.071              | 16.190            | -18                  | 1                   |
| Rovigo                        | 51.868             | 23.238                                   | 23.249          | -11  | 51.872              | 23.195            | -4                   | 54                  |
| Scandicci                     | 49.764             | 20.464                                   | 20.489          | -25  | 50.031              | 20.514            | -267                 | -25                 |
| Tolentino                     | 20.401             | 8.080                                    | 8.038           | 42   | 20.390              | 8.013             | 11                   | 25                  |
| Trezzano sul Naviglio         | 19.761             | 8.165                                    | 8.171           | -6   | 19.137              | 7.476             | 624                  | 695                 |
| Tricarico                     | 5.848              | 2.258                                    | 2.309           | -51  | 5.904               | 2.274             | -56                  | 35                  |
| Villaricca                    | 29.777             | 9.361                                    | 9.307           | 54   | 29.732              | 9.302             | 45                   | 5                   |
| Total                         | 6.764.074          | 2.934.574                                | 2.929.474       | 5.102  | 6.534.094           | 2.694.717         | 229.980              | 234.757             |

<sup>27</sup> Table 2 clearly shows that among data of the two sources there is a gap, both for the number of households that for the number of individuals. These differences are particularly pronounced in the Municipalities of medium-large dimension (Rome, Naples, Palermo and Latina). Moreover, some investigations have been conducted to find the reasons for differences in MAL of each Municipality between

the number of reference persons of family cards (column B) and the number of household codes (column C). It is showed that it is not always associated to a household only one reference person. In fact, overall 774 households have been detected without reference person and 5,452 households have more than one reference person<sup>2</sup>. Although this matter involves only 0.2% of households, it will still be a factor to be checked carefully before the next Census.

28 Even if all technical standards and protocol for data transmission have been specified in detail to all Municipalities, the simultaneous application of all standards, however, was found only for 8 Municipalities (26.7%). This percentage, if reproduced for the totality of Municipalities, would amount to 2,160 Municipalities (out of approximately 8,100) on which no action should be computationally done before the next Census. In table 3 the more relevant errors in respect of standard for which an ad hoc intervention have been necessary, for the number of Municipalities and the projection of the percentage on all the 8,100 Municipalities are reported. Obviously, in the table, the belonging to either type of error is not mutually exclusive.

Table 3 – Type of error, number of Municipalities , percentage on total Municipalities of the pilot survey 2009 and projection of percentage in 8100 Municipalities in 2011

| Type of error  | N° of<br>Municipalities in<br>the Pilot Survey<br>2009 | % on<br>Municipalities in<br>the Pilot Survey<br>2009 | Projection of % on 8,100<br>Municipalities in 2011 |
|--|--|---|--|
| Format (not .csv) and filename                         | 4  | 13.3  | 1,080  |
| Record format  | 6  | 20.0  | 1,620  |
| Out of indata protocol                                 | 2  | 6.7   | 540  |
| Birth date and/or registration date in <i>anagrafe</i> | 5  | 16.7  | 1,350  |
| Place of birth (Municipality or foreign country)       | 7  | 23.3  | 1,890  |

In 10 Municipalities of 31 (2,613 of 8,100 at 2011) it has been necessary to send the file of decoding of the variables *relationship* and *marital status*; on these variables numerous errors than required were found. Looking at the decoding file, moreover, it was noted for the variable of *relationship* an anomaly of considerable magnitude compared to data. With surprise, it has allowed to emerge in the archives of 3 Municipalities the presence of dead and not cancelled individuals. Indeed, in some records the modality of *relationship* is registered in the family card with conditions such as "*dead*", "*ex reference person now deceased*" and even "*died of the reference person*"; these modalities, although present to a negligible number of cases, indicate that besides the problem of "*non-cancellation for death*", the municipal registers have a specific modality to archive but not delete dead persons. In the future this aspect will be closely monitored to avoid, where possible, in addition to the problem of the interval between extraction of MAL and census, to send questionnaires to died reference persons.

30 Several checks on MAL files were made in order to verify that, for each record, both completeness of data and inter-and intra variable congruency. Of the total number of records supplied (6,764,074 individuals) the 17.52% of them have at least one error or missing data. The distribution of cases, however, is highly concentrated in the large towns (Palermo, Parma, Rome and Naples), covering 87% of encountered errors. Moreover, the weight percentage of errors and missing data varies widely between Municipalities. However, it has been considered that for Municipalities with high incidence of records with at least one error or missing data the cause of high prevalence consists of systematic errors on format and coding of variables. It was found, moreover, that the variables strictly necessary for mailing the questionnaires (*surname*, *name*, *family code*) are present in almost all records. This is a very encouraging signal for use of MAL in the census.

\_

<sup>&</sup>lt;sup>2</sup> The difference between the two totals, like 5.102, is not equal to the difference between records without reference person and records with more than one reference person because they are distributed in different manner in the Municipalities.

- 31 Another element of reflection is relative to the *date of registration in the anagrafe*. For this variable it has been noted that almost 26,000 records presents an inconsistency with the *date of birth* (date of registration earlier than date of birth), this type of error was significant for two medium-sized towns (20,000 to 50,000 inhabitants) where the incidence is around 20% -22% of the sent records.
- 32 Other elements that will be monitored in future are the presence of duplication of individuals within and between archives, cases relating to missing tax codes that determine the a priori impossibility to complete the questionnaire *online* as well as a minimum set, poorly relevant, of problems identified during the management of free phone customer support number which citizens could contact during the survey.

# VI. CONCLUSIONS AND FUTURE PROSPECTS

- The first significant experience on data contained in the MAL and test performed on the transmission/acquisition of data system are very positive because they allowed the emergence of problems and strengths of different nature. The definition of standards, modalities and transmission constraints of data and all aspects related to the definition of record format are key elements to be developed in the most categorical as possible and unambiguous mode before the acquisition of MAL for the Census of Population and Housing. The experience of the Pilot Survey 2009 has meanwhile highlighted the need to change the record format of two types of variables; in particular the format of the variables *date of birth* and *date of registration in anagrafe* and the insertion of a separator (/) between day, month and year can significantly reduce the number of errors due to misalignment of data. Moreover, the fieldwork has revealed the need to inquire information inside the building (floor, building or stairs) whenever these variables are held by the MAL office.
- 34 It was also confirmed the strong contribution that in several Municipalities the Software Houses that manage various aspects related to technical and computer data management will provide. It is evident that when the technical aspects are handled by experts, the gain in terms of adherence to the standards required is considerable; for this reason in the future Census the goal is the maximum involvement of these ones.
- 35 From the experiences conducted so far, we have seen that more detailed and precise the standards of transmission/acquisition data are, the higher the level of formal quality of received data. The goal to be attained before the Census of Population and Housing in 2011 is to provide a system of transmission/acquisition data that minimizes the workload for the Municipalities and the time they need to extract and send the MAL and the complexity for the ISTAT Central Division of General Census of the computerized management of MALs.

The target will be achieved by programming:

- clear and compelling standard and in particular the "optimal" record format that balances the needs related to the Census and "sustainability" for Municipalities in terms of skills and quality assurance
- ii) the presetting of software for the correction, contextual to the mailing, of data remediable with deterministic procedures
- iii) the implementation of a set of controls and indicators to validate the basic data so as not to acquire MALs which do not meet certain quality constraints
- iv) a system to ensure that the data transmission and any possible decoding is very easy for the Municipalities.
- 36 The whole experience that can be gained to achieve these goals is considered crucial to the success of a Census which is in many ways unique and innovative than its predecessors. At this regard, ISTAT has planned to implement during 2010 an experimental survey, foreseen in the National Statistical Program,

# Working paper 3 Page 10

whose overall objective is to obtain in advance information necessary to optimize the acquisition process of MALs also for the development of the software that will enable Municipalities to submit data. The experimental survey will involve a sample of 1,850 Municipalities. The sampling criterion has the operational objective to observe the largest number of formal errors, for which an *ad hoc* intervention will be required on each MAL assessing the impact on all Municipalities and the resources - human, financial and time – to be devoted to their solution. The main objectives of the survey are to allow to the sampled Municipalities to familiarise with the procedures for preparing and sending lists, allowing them to identify also possible internal problems both in terms of technology and content, to involve the largest possible number of Software Houses and to explore the realities in which they are not present.

- 37 ISTAT will allow the transmission of MAL exclusively through a web application on secure protocol that:
  - i) allows the registration of municipal technical contacts
  - ii) enables them to transmit MALs on secure protocols
  - iii) monitors the respect of certain required standards
  - iv) restores, where possible, the formal correctness of certain information with the reserve to accept or reject, if inaccurate, the sent data
  - v) makes available to Municipalities some reports in which they can view summary information and quality indicators. The reports will be useful to Municipalities because more the high quality of MAL data will be, the lower the workload of the Municipality at the next Census.
- 38 The obtained results in terms of validity of process and product, on the basis of the tools used will be available before the requirement of the MAL for the next Census (1/01/2011). The experimentation will enable, in fact, to finalise different aspects of the initial phase of the next Census.

\*\*\*\*