National Statistics Administration Department
DANE [abbreviation of National Statistics Administration Department in Spanish]

Office of Census and Demography

Conceptual and Methodological Design
General Census 2005 – CGRAL [General Census]

April 2006

DANE
Conceptual and Methodological Design
General Census 2005 – CGRAL [General Census]
Code: CM-CGRAL-TEM-01
Version: 03
Page: 2
Date: April 15, 2006
Prepared by: CM [Concept and Methodology] Team
Revised by: CM Coordinator
Approved by: CGRAL Coordinator

Contents

Ruta:C:\Documents and Settings\LGalvisA\Configuración local\Archivos temporales de Internet\OLK46\CM-CGRAL-TEM-01.doc

Page: 3

Introduction
The National Statistics Administration Department – DANE as the rector of the strategic information politics for decision-making in all of the areas of society, is the official entity responsible for designing, implementing, and carrying out the transference of the technology for the methodology that guarantee its generation and use with the greatest standards of quality, coverage, and opportunity to satisfy the needs of the final users.

This obligation is seen even more, if you consider the strong dynamic of the social, political, economic changes that the country has experienced in that past decade, which has determined an important deficit in diagnostic information that permits the adequate direction of the public politics. In the local context, this

information deficit is even more extreme, determining large errors in the administrations for the adequate directing of the investments that permit a long-term vision to guarantee a substantial and sustainable economic development to improve the quality of life of the persons. For he above, the information not only is necessary to facilitate and improve procedures, but it also has great strategic value in the process of public politics decision-making and completion.

In this same train of thought, we have identified as one of the principal obstacles to the adequate development of a process of decentralization is the low level of capacity for local governments to implement action, motivated largely by the lack of local diagnostic information that permits the evaluation and prospection of the basic variables that determine the economic and social improvement of the communities.

For the above, and taking into account that the majority of the strategic information that is produced is done through research based on sampling designs, in which the problem is in its coverage, what is required is the development of a large macro-project that will permit us to obtain diagnostic information about all of the territorial entities about their basic unit, the persons, and the characteristics of where they live. This type of project can only be achieved through Population and Dwelling Census for each of the territorial units, obtaining strategic information that permits characterization of each of them and identify the principal deficits that cause problems of poverty, emigration, economic development, demographic growth, public services, dwellings, and in the normative framework of assigning basic fiscal resources for the development of projects with large local and regional impact.

The first line in the structuring of a Reliable Census Project, is the formation of a conceptual and methodological design that is robust from all points of view, which constitutes the binnacle for the development and implementation of successful basic processes such as carrying out balanced budgets, field operations with all of the

Page: 4

necessary supports and the design of all of the technological platform which permits the adequate and opportune access to the census information.

In this document we present each of the conceptual and methodological elements that are defined in the previous paragraphs, as well as the thematic content of the General Census. As for the Population and Dwelling section, the criteria are defined to define the content of the basic questionnaire, as well as the structural, general consistency and validation, and imputation norms necessary for the development of the intelligent questionnaire software for the DMC as well as for the process of the final depuration of the information.

Page: 5

I. Definition of Concepts

For the adequate realization of the General Census you should manage a series of concept with the goal of guaranteeing the quality of the information and its adequate use by the different users.

Since the Population and Dwelling census has three observation units that permit the adequate structuralization of the census information: Dwelling Unit, private home and the resident persons, all of them referring to a geographic unit, the adequate management of each of them, in each of the processes of the census Project, are the guarantee for obtaining the strategic information for each of the territorial

entities. For your identification you have the following fundamental concepts, upon whose adequate application the census information depends.

-Edifice: This is a building, independent and separate, comprised of one or more units.
Independent, because it has direct access from the public ways, roads, paths, or spaces for common circulation.
Separate, because it generally has walls that delineate it and differentiate it from other buildings.

-Unit: This is an independent and separate space that constitutes part or the totality of a building whose uses can be economic, dwelling, or mixed.
Economic: when it is destined for or is being used for industry, commerce, or services.
Industry: is the physical and/or chemical transformation of materials and components into new products, whether the work is completed with machinery or by hand, in a factory or domicile, and the products are sold via wholesale or retail commerce.
Commerce: Wholesale and retail buying and selling of merchandize (goods that go to the market), new and used and that are not subjected to transformations of their intrinsic nature during the activities inherent to commercialization, transportation, storage, packing, repacking, bundling, etc., or those of accommodating, exhibiting, presenting, or promoting the merchandise that is the object of the sale.
Services: These are heterogenic products generated each time that they are solicited, over which there are no property rights on the part of the user and that cannot be negotiated separately from their production nor can they be transported nor stored.
Economic activity associated with households: When in a dwelling the spaces that belong to the household are shared for the regular development of a commercial, service, or industrial activity that generates income.

Page: 6

Agricultural activity associated with the dwelling: This is only applicable in Class 3 (disperse rural). It is defined in the following manner: The agricultural activity is identified and located inside the rural lots or farms with dwellings, households and persons with habitual residence. For the purposes of the General Census 2005 this category has been adopted as an observation unit, defined in this manner: land where a rural dwelling with households and persons who are habitual residents is located and also occupied by agricultural crops; transitory and permanent, forestry crops, pastures or grasses and weeds and forage; breeding animals such as bovines, equines, porcine, goats, sheep, donkeys, poultry; and fish farming and shrimp farming, that belong to a physical and geographical unit best called FARM. For census purposes do not take into account the type of administration, titles, occupancy, or size of the unit.
Dwelling: when it is inhabited by or destined for habitation by persons;
Mixed: when both uses are combined: economic and dwelling;

Dwelling unit: This is an independent and separate space with areas that for exclusive use, inhabited or destined to be inhabited by one or more persons.
With the goal of facilitating the quantification of the stock of dwelling units and the analysis of the information for the purpose of the census we have the following typology:
House: This is the building constituted by one single unit whose use is that of dwelling, with direct access from the public way or from the interior of the building. The sanitary services and the kitchen can be inside or not.
For the census purposes, also consider those houses in which the garage, the living room or any other room is destined for economic uses.

Indigenous house: This is the building constituted by one single unit whose use is that of dwelling, built according to the custom of each ethnic group conserving the traditional structure.

Apartment: This is a dwelling unit, which is part of a building, in which there is/are (an)other unit(s) that is/are dwelling unit(s). It has direct access from the exterior or through passages, patios, hallways, stairs or elevators. It has sanitary services and kitchen available in the interior.

Room: This is the dwelling unit, which is part of a building and that has one or more spaces available. It has direct access from the exterior or through passages, patios, breezeways, hallways, or other spaces that are for shared circulation. In general it lacks sanitary services and kitchen in its interior, or only has one of these two services available.

Other type of dwelling: It is a space adapted for dwelling, where at the moment of the interview there are persons living. Generally it lacks sanitary services and kitchen,

Page: 7

such as: railroad cars, containers, boats, gypsy tents, caves, bridges, and shacks.

Habitual resident: This is the person who lives permanently or the majority of the time in a dwelling, although at the moment of the interview he/she is absent. Also considered habitual residents, because they do not have another habitual residence somewhere else:
The persons who are absent for special reasons such as: vacations, training courses, business trips; always and whenever the absence is for six months or less.
Traveling agents, merchant marines.
Persons who have been kidnapped, regardless of the length of absence.
Persons who are ill and interred in hospitals or clinics, regardless of the length of absence.
Displaced persons, regardless of the length of permanence in the household that is being interviewed.
The persons temporarily detained in police jails.

Household: This concept that is presented below is the result of a consensus among all of the Technical Departments of DANE[1], with which we look for homologation of the different research projects with the census, taking into account that this conforms the framework of the different research projects. Equally we carried out a homologation with that used by CAN[2], which has some differences with the Colombian context The Concept of Household that was formulated as follows:
It is the person or group of persons, related or not, that occupy a dwelling totally or partially; they attend basic needs, covered by a common budget and generally they share meals.

Head of household [male or female]: This is the habitual resident recognized as such by the rest of the Members of the household.

Household informant: This is each of the persons who are habitual residents who are over the age of 16 years old. In the case that the person is not present and cannot complete the interview, the information should be given by the Head of Household [male or female], or his/her spouse or a resident person who is over 18 years old and knows the information about the person. For those under 18 years old, the information should be given by the Head of Household [male or female] or his/her spouse.

---

[1] The DIRPEN as the Technical Department in charge of the standardization of the Statistical Systems officializes this concept.
[2] This is the person or set of related or unrelated persons who occupy a dwelling totally or partially; they share at least the principal meals and share other basic needs, covered by a shared budget.

Special Place of Housing [L.E.A. is the abbreviation in Spanish]: This is an institution in which a collective group of persons, generally not related, lives (sleeps).

Page: 8

In addition, there are other basic concepts in the development of the census. Some are related to the thematic content which can be seen in the Annex about the Glossary of terms. Others are related to the modification which DANE has carried out in regards to the execution of the general census 2005, concepts which, because of their methodological relevance, are listed below:

Conglomerate: Set of contiguous municipalities that conform a homogenous geographical area in the socio-economic, cultural and demographic aspects.

Sweeping unit: geographical area that can be enumerated with a team of six interviewers and one supervisor in one day.

Census period: duration time for the collection of information for one or various conglomerates.

Census moment: This is the 00:00:00 hours of the day of initiating the operation of collection for each conglomerate or group of conglomerates.

Enumerated population: This is the quantity of men and women, by age, that is obtained for each municipality during the census period defined for the data collection.

Single Statistical Moment: Single date of reference defined for restituting the information for all of the municipalities. Given that each group of conglomerates has different census moments, it is necessary to restitute the census information to one unique reference date, which permits restituting the information to aggregate geographical levels, department, region, and national. For this purpose, we will carry out a process of conciliation according to the characteristics of each conglomerate and starting with the compensation equation that is done in all of the municipalities on the closest date to the collection for the most dynamic conglomerate groups. Following this line of thought, we take the census moment of the conglomerates C4 and C5 as single reference date for the census information[3].

Conformation of the Conglomerates

Page: 9

This point is considered one of the methodological aspects of greatest importance for the adequate development of the Population and Dwelling census. Its importance lies in the fact that it permits us to minimalize the gaps in the facts caused by determinate factors of the dynamism of the resident population in each of the geo-political units into which the country is divided, such is the case of the spatial[4] movement and the vegetative or natural[5] growth.

Carrying out this process, we structured the following phases:

---

[3] Adopted as the median date for this collection period, corresponding to the 11th of November.
[4] As is the case of the population migration motivated by educational, labor, environmental factors, among others, as well as those determined by reasons related to preserving live due to threats of armed groups operated outside of the law.
[5] Determined by the deaths and births in each geopolitical unit.

-Phase 1: Structuralization of the municipal information with variables of migration, fertility, mortality, forced migration, and population density, each of them symptomatic of development.

-Phase 2: Development of the statistical processes[6] to identify the variables that adequately discriminate, to minimalize gaps coming from municipal level mobility.

-Phase 3: Development of multivariable processes[7] for the conformation of the homogenous conglomerates as related to mobility. In this process we include condition factors such as:

The municipalities of Orinoquía, Amazonia and those of the Pacific range- Chocó, Valle, Cauca and Nariño (census by routes)- are not included in the analysis of conglomerates.

New municipalities about which we do not have sufficient information, should be included in the conglomerates of the municipalities from which they were separated.

Bogotá as a city-region should be in the same conglomerate as the municipalities with the greatest influence by approximation.

The principal metropolitan areas or minor cities should be with their area of influence, which implies the identification of these central units.

-Phase 4: Final conformation of the Groups of Conglomerates, which by their characteristics determine the time order in which the census should be executed in each conglomerate and the final formation of a mathematical model for population restitution, in a single statistical moment.

Taking the characteristics of each of the different regions into account, the location of the central units and their categories, the population movement whether it is by choice or motivated by

Page: 10

external factors, generate interactions between municipalities of different regions or in the interior regions of the municipalities.

We consider this type of dynamic to be the principal factor for guaranteeing minimalization of the remnants by mobility both in the levels and structures of information. The large central units such as Bogota, Cali, Medellin and Barranquilla determine corridors where important sub-regional central units are found for the linked processes of migratory movements. This is the case of Pereira, Manizales, Armenia, Ibague in the western central part; Popayan, Pasto and Neiva in the southern part; Villavicencio en the western part; Barrancabermeja, Bucaramanga and Cucuta in the eastern central part; Valledupar Cartagena, Barranquilla, Santa Marta and Riohacha in the northern part.

Also identified as an equally important axis that for its agroindustrial and mining characteristics an internal dynamic is generated in said sub-regions. This is the case of the axis Uraba-Monteria-Sincelejo with an area of influence that extends to the north of the department of Antioquia: The axis of Aruaca-Yopal also has an important agro-ranching and petroleum impact.

This can be seen on the following map.

Page: 11

#tr note: This page presents an image, but no words.

Page: 12

---

[6] Method of Principal Components
[7] Hierarchical factorial analysis

2.1 Statistical Model for the definition of Groups of Homogenous Conglomerates of Municipalities

The objective of this exercise is to conform, starting with the statistical data of a varied nature, municipal groupings that present, in accordance with the available statistics for each one of them, similar conditions that join them as homogenous groups, with the goal of utilizing said groupings as part of the strategy for carrying out the General Census 2005 in the Colombian territory.

In particular, we present two primordial goals. In the first place, it pretends an approximation of the characterization of the current socio-demographic situation of the Colombian municipalities. Complementarily, it desires to obtain a municipal typology that illustrates the analysis and permits the advance in the establishment of a statistical classification based on socio-demographic aspects for the purpose of the census.

Given a set of individuals [in this case municipalities] and having for each of them the information in the form of variables [partial indicators], the analysis of conglomerates tries to classify them in groups [not determined "a priori"] in a manner that the individuals belonging to each group are, with regard to the available information, as similar as possible and that each group is distinguished from the rest as much as possible in terms of multi-variant distances.

The information that has been used for the conformation of the conglomerates comes from the following sources:
-Information about the evolution of the municipal population provided by the national census of population and dwelling from the years 1895 and 1993.
-The birth and death statistics, occurred in the country classified according to place of residence of the mother or habitual residence of the person who died.
-Municipal population projection numbers prepared by the DANE.
-Information about education, specifically elementary and secondary enrollment.
-Migration according to information provided by the Social Solidarity network.
-Information about the Number of enrolled electors.
-Value of municipal tax income.

2.2 Analysis of Conglomerates
The Analysis of Conglomerates [cluster analysis] is supported by a series of techniques [fundamentally algorithms], that have as their objective the search for similar groups of individuals who are grouped in conglomerates as homogenous as possible. These are not known in advance but are suggested by the information obtained with the data.

Page: 13

It was necessary to make tests of distinct existing methods and techniques to obtain a result that adequately adjusts to the objectives sought. Throughout the process it resulted evident that it was convenient to apply a hierarchical method, as much for the adequate use of the data, for the necessity of controlling the process and because we desired to obtain a typology that was coherent at various levels.

As a classification algorithm, after the pertinent tests, the Ward model was elected. This method defines the distance between two groups as the Euclid distance squared with regard to the medium of the observations. Each step of resulting conglomerates are those that present the least increase in the global sum of distances of the square within the possibilities. In addition, it is a technique sufficiently sensitive to the presence of outliers, for which it is convenient to eliminate the extreme cases of analysis and treat them as groups of a single individual.

In this type of analysis it is necessary to be especially careful in the election of the variables that are going to characterize each individual and that serve as a base for making groupings.

Once the variables are selected there are three fundamental problems for their joint treatment:
-The variables are measured in different units and present very different variability.
-Normally the are highly correlated, which supposes redundancy in the information.
-There is an excessively large number of variables [dimensions] which make the treatment and interpretation of the analysis difficult.

To solve the first problem, the entrance variables have been normalized. To resolve the other two problems, we have looked to the analysis of principal factors, which permit, in addition, analyzing the authentic meaning of each partial indicator with precision and orients on how to present some synthetic indicators.

Before utilizing the measurement of the distance, we should find the set of variables that best represnts the concept of similarity. [Concept with which we look for the maximum internal homogeneity of the conglomerates]. For this case, the interest lies in taking 1021 municipalities of Colombia and group them in four clusters. [Excluding the 98 municipalities for which it has been planned that they will be enumerated under a strategic operative based on routes].

The variables that best represent the concept of similarity, under the study, were:

Page: 14

-Enrollment in elementary schools
-Enrollment in secondary schools
-Births
-Deaths
-Number of lots
-Number of registered voters
-Value of municipal tax income

#tr note: what follows is a chart with the above terms placed in a diagram
The terms, from top to bottom, are:
Voters
Deaths
Income
Population 1993
Births
Elementary enrollment
Secondary enrollment
Migration
Remnant

2.3 Results of the model

For the conformation of the conglomerates that we expect to define for the census taking in the 1021 municipalities it has been decided to utilize the SPAD program.

The first result obtained is demonstrated in the following graph, where it can been seen clearly that starting with the variables of elementary enrollment, secondary enrollment, births, deaths, number of lots, value of municipal tax income, number of registered voters, the 1021 Columbian municipalities with information in all of the variables can be organized in four possible conglomerates:

Page: 15

#tr note: what follows is a chart with five classes circled, labeled class 5/5, class 1/5, class 3/5, class 2/5. For each group, there are cities listed in the accompanying circles.

The method demonstrates that the municipalities of Bogota and Cartagena are atypical entities, within said set and this in terms of use for the census operations will mean that for the goals of the census each one of them should conform a conglomerate. The third will be integrated by the municipalities of Cali, Barranquilla and Medellin; the fourth will be made up of the rest of the municipalities.

Setting aside the atypical municipalities and including the information on the variable of internal population migration collected in each municipality in the model, we obtained the result indicated below:

Page: 16

#tr note: the drawing shows class 2/3 and 3/3 with Colombian cities marked on the chart.

According to this last analysis we achieved a classification of the set of municipalities of the country identifying the following groupings:

-Group 1. Finally, a group in which 883 municipalities can be placed that expel and receive population on a small scale. They are without doubt the smallest municipalities of the country, and of course, the volume of the information of the symptomatic variables fluctuates a lot and their quality is not very reliable. For this group of municipalities we do not possess a statistical tool that can measure the population loss.

-Group 2. This set is conformed of 55 municipalities with ambivalent behaviors, this means, that present population migration expulsion and reception conditions that in a good measure are compensated with a negative final population movement result. They are municipalities that are smaller than the previous groups, whose data on births, deaths and school enrollments only reach a third of those of the previous group.

-Group 3. This is an additional set of 36 municipalities of intermediate size within which there are departmental capitals and nearby municipalities. This group is characterized as municipalities that are

Page: 17

clearly receptors and low-level expellers which can be thought to be process of return migration. They present high volumes of elementary and secondary enrollment. This group contains 18.9% of the population.

Group 4. Conformed by a set of municipalities that are receptors of population which are the most affected by internal migration and fundamentally by the effects of the impact of displaced population. The municipalities of Bogota, Medellin Cali and Barranquilla are in this category, which presents a clear

behavior in terms of condition of reception of displaced persons. Medellin and Cali are characterized as presenting elevated volumes of deaths.

These municipalities are central units[8] of other nearby municipalities, called bedroom-cities, operationalized based on criteria of distance and travel time between municipalities and the central city; the result is that this conglomerate is conformed by 47 municipalities in the following manner:

Bogota is the central unit of 17 municipalities
Medellin is the central unit of 13 municipalities
Barranquilla is the central unit of 4 municipalities
Cali is the central unit of 9 municipalities

This group contains 36% of the Colombian population.

The distribution of the groups of conglomerates that result from the model can be observed in the following map, which is consistent with the internal migration movements, linked to change of residence for selective factors due to work, study and environment. This is equally coherent with the movement of areas affected by the armed conflict that determines forced movement toward larger urban areas.

Page: 18

#tr note: there is an illustration of the departments of the Colombia.

DANE
General Census 2005
Conglomerates of Municipalities
Result of the Multi-varible Statistical Analysis

Conglomerates
C1: Low Dynamic
C2: Mid-level Dynamic
C3: Intermediate Cities
C4: Principal metropolitan areas
Census routes
Department limits
Source: DANE-National Geostatistical Framework MGN, 2003
General Census Group 2005

Scale: 1:750000

Page: 19

2.3 Conformation of the Groups of Conglomerates

Starting with the results obtained through processes of multivariable analysis with symptomatic variables[9] that give evidence of a series of central units of "attraction" located regionally, we obtain a first scenario of the dynamic population of the territorial entities which were demonstrated in the previous section.

---

[8] A municipality is defined as a central unit in the sense that this acts like a solidifying municipality of a set of municipalities that are located in its spatial surroundings.

This statistical result defines four groups of conglomerates conformed by all of the metropolitan areas[10]; another group conformed by all of the intermediate cities, the majority of which are departmental capitals and municipalities that for their regional position are important centers in the commercial and service sectors[11]; a mid-level dynamic groups conformed by municipalities with regional tourism characteristics, and lastly a group in which we find a large number of municipalities that from the demographic point of view can be considered a low dynamic level. In the following table the distribution in terms of population and number of municipalities can be observed.

Quantity and percentage of population according to groups of conglomerates

Group	Population 2005	%	Number of municipalities
Group 4	21,823,783	47.4%	72
Group 1	14,093,164	30.6%	883
Group 2	4,659,773	10.1%	55
Group 3	3,473,828	7.5%	11
Census routes	1,994,563	4.3%	98
Total	46,045,111	100%	1119

In the following phase we carried out a series of procedures in which the information about distances from these central units to municipalities in the area of influence, regional characteristics identified by the headquarters and sub-headquarters of DANE and aspects that are of operational nature that are related to results and number of interviewers per municipality.[12] With this information, we carried out a balance among groups of conglomerates, which defined the relocation of municipalities controlling that they have demographic characteristics homogenous to the group to which they were transferred.

Page: 20

The development of this process determines the necessity to open some of the conglomerates, with the goal of facilitating the operative process, specially the utilization of pocket computers o DMC maintaining the same principals of homogeneity defined in the statistical model giving, as a result, the following classification:
-A conglomerate in which Bogota D.C. and 24 municipalities in its area of influence that are highly correlated with said central unit.
-Group 1 opens up into two, with which we obtain a group of conglomerates with a low total of municipalities, which for its population dynamics characteristics will be the first group to carry out the census operative or given the operative conditions, this groups [#tr note: the Spanish version contains this grammatical singular/plural inconsistency] can carry out an overlapping form with the other groups of conglomerates.

In the following box we can observe the new classification:

Quantity and percentage of population by conglomerate
Group / Dynamic / Population 2005 / % / Number of municipalities / Number of interviewers
C5: Special Dynamic* /Special / 8,331,279 / 18.1% / 25 / 6346

---

[9] Births, deaths, school attendance, infantile population with prevalence of polio and triple viral, school attendance by levels, number of voters, consumers of electrical energy, beneficiaries of SISBEN, number of lots, population density, immigration since 1993, displaced persons.

[10] Include Bogota, D.C. They are cities with high development which determines a high dynamic.

[11] Mid-high dynamic.

[12] Urban and rural.

C4: High Dynamic** / High / 9,843,669 / 21.4% / 38 / 6883
C3: Upper-Middle Dynamic*** / Upper-Middle / 9,538,309 / 20.7% / 116 / 7032
C2: Middle Dynamic / Middle / 7,973,547 / 17.3 / 259 / 7568
C1A: Low Dynamic / Low / 5,868,561 / 12.7% / 402 / 6912
C1: Low-low Dynamic / Low-Low / 2,495,183 / 5.4% / 181 / 2957
Census Routes / / 1,994,563 / 4.3% / 98 / 2338
Total / / 46,045,111 / 100.0% / 1119 / 40036
* Bogota and Area of Influence
** Principal Metropolitan Areas
*** Intermediate Cities

On the following map we present the groups of conglomerates, which execute the census from the palest yellow until it finalizes in Bogota D.C. and its area of influence.

Page: 21

DANE
General Census 2005
Conglomerates of Municipalities

Result of the Statistical Analysis and the geographic and cultural evaluation of variables

[#tr: what follows is a color-coded map with place labels; the color-code is below]

Conglomerates
C1: Low Dynamic
C1A: Low-low dynamic
C2: Middle dynamic
C3: Intermediate cities
C4: Principal metropolitan areas
C5: Bogota and area of influence
Census routes
Departmental limit
Territorial limit DANE

Scale: 1:750000

Source: DANE National Geostatistical Framework MGN, 2003
General Census 2005 Group

Page: 22

According to the above, from the technical point of view and considering that the objective is to minimalize the gaps that can be introduced in the structure and levels of the census information by populational dynamic factors, the operative implementation of the Population and Dwelling census we should take into account the following:

-The simultaneous collection in all of the conglomerates, located in the different regions of the country, which conform each of the different groups.

-The interior of each conglomerate, the collection is carried out in simultaneous form in all of the municipalities that conform it. The census moment is the 00.00 hours of the initial day.
-The spatial distribution and execution of the census of the conglomerates is carried out from the group with the lowest dynamic to the highest. However, for operative necessities or the optimalization of the DMC, in these conglomerates the collection can be carried out by overlapping with the other conglomerates.

This process guarantees that that the census is carried out in the same order as the population movements, with which in some of the central units of attraction the persons or groups of persons will be enumerated. The second problem that the gaps introduce is the population information is the so-called "counter-flows" which can determine that a volume of population will not be captured in the census, given that after doing the census in a municipality these persons move toward the municipalities that have already been enumerated.

As you can observe in the following map, the migration movements have very particular regional characteristics, which can be classified as linked processes that go from areas of "fewer opportunities for development" toward central units of sub-regional, regional and national influence.

The above can be resolved in two ways: The first is through a strong awareness campaign, in which persons not enumerated are encouraged to report to the census offices in each municipality. The second way, which from a technical point of view is the most appropriate, is to carry out the collection in a period of less than twelve months in which it will possible to overlap the collection among conglomerates of low and medium dynamic, leaving those with greatest density for the final period.

The above means that to minimalize the gaps due to mobility and natural growth, the collection of the census data should be done in the following order:
First the C1; then the C1A; then the C2; then the C3; then the C4 and lastly the C5.

Page: 23

As far as the times for the execution, the existence of dead months or weeks in which the collection in the municipal capitals cannot be done should be taken into account, because this would imply high probability for the existence of gaps due to factors in the population dynamic, with significant effects on the structures by age and sex in the territorial entities. Below we present the table for the collection of census information.

The strategy of collection to minimalize these gaps should consider the following aspects:
-A number of days are required to facilitate the moving of the machines to the municipalities of the next group of conglomerates.
-The training should be done with sufficient advance time to guarantee the adequate formation of the enumerators.
-Processes should be done for each municipality with the goal of estimating a series of strategic indicators using the census information, which will permit the evaluation of the quality of the information (in addition to the coverage).
-A baseline should be generated for each municipality with the strategic indicators on fertility, mortality, structures, mobility, indexes such as the relationship children/women, masculinity, Myers, among other to define the adjustments necessary for each municipality and in this way, guarantee the validity of the information.

In addition, it should be considered that between May 22, 2005 and May 22, 2006, a series of "dead" months exist, in which it would be risky to carry out the collection of data because it would increase the number of revisits and, most importantly, increase the probability that a large number of persons would not be enumerated, periods that can be observed in the following graphic:

[#tr the illustration presents months of the year and abbreviations for months of the year]
May / June / July / August / September / October / November / December / January / February / March / April / May

The "colors" have the following meanings"
[ ] Period appropriate for collecting data
[ ] Period not appropriate for collecting data due to vacations
[ ] Period of risk for collecting data due to Election Period
[ ] Period for transition

Page: 24

Under this order of ideas, we can consider different strategies for collecting the information taking into account the characterization of the conglomerates that goes from using all of the period with consecutive collection to smaller periods of time and collection overlapping conglomerates.

Option 1: Would be to carry out consecutive collection beginning in the conglomerates with the lowest dynamic and ending with Bogota and its area of influence as shown in the following graph:

[#tr the illustration presents months of the year and abbreviations for months of the year]

Census by routes
May / June / July / August / September / October / November / December / January / February / March / April / May

C1 / C1A / C2 / C3 / C4 / C5

Year 2005 / Year 2006

With this format the principal metropolitan areas, as well as Bogota and its area of influence would dot the collection between February 15 and March 15, 2006, a process that would be in a period of parliamentary and presidential elections, whose effects on the credibility of the resulting information would be disastrous, in addition to going against the recommendations of Organizations like the United Nations.

Option 2: Carrying out the collection in consecutive form, but leaving only Bogota and its area of influence to do the collection between February 15 and March 15, 2006, and the rest of the conglomerates would be done in 2005. Using this strategy implies using part of the months of June and July for executing the census in the conglomerates with the lowest dynamic, taking into account that said period of vacation does not have the same dimensions of movement to vacation sites on the part of the members of a household as those of the end of the year. In the following graph you can observes this strategy.

[#tr the illustration presents months of the year and abbreviations for months of the year]

Census by routes

May / June / July / August / September / October / November / December / January / February / March / April / May

C1 / C1A / C2 / C3 / C4 / C5

Year 2005 / Year 2006

Page: 25

However, carrying out the collection for the conglomerate 5 in 2006, will continue with the problematic of the elections, increasing with the reality that it is the principal central unit and the one with the greatest political incidence, factors that can affect the quality of the information of said conglomerate.

Option 3: This strategy implies carrying out the totality of the collection in the year 2005 according to the calendar of optimal months and the rest of the time, not until May 2006, so that all of the process of consolidation, adjustments for coverage, restitution of population dwelling and household, conformation of the final census archives and production of tabulations. To achieve this goal it is necessary that the collection of some groups of conglomerates be done in simultaneous form or that they overlap in some moment, taking into account moving the PDA, the awareness and training of the operative personnel in each municipality. This alternative can be observed in the following graphic.

[#tr the illustration presents months of the year and abbreviations for months of the year]

Census by routes
May / June / July / August / September / October / November / December / January / February / March / April / May

C4
Process of restitution, consolidation, results

The technical viability of the third option is supported by the following:

-The adjustments to be made in the restitution to a single statistical moment are minimal, considering the lapse of 4 months in collection. The greatest adjustment would be on the 18 municipalities selected to carry out the collection in the month of May.
-By carrying out the some of the groups of conglomerates or with overlap, we are minimalizing the gaps caused by the effect of counter movement.

By doing the conglomerates with low-low and low dynamic in overlapping format with the other conglomerates, as is shown in the last graph, implies a reorganizing of said conglomerates to maintain the demographic principal that the

Page: 26

"central units" should be done simultaneously with their "area of influence", which is basic to guarantee the minimalization of the gaps caused by mobility. The above means that the group of municipalities of low dynamic, group C1 and group C1A, should be redistributed in two large groups under the following criteria:

-The size of the tow groups should be balanced in terms of the number of interviewers, to guarantee the utilization of the PDAs.
-They should be distributed in such a manner that they agree with the area of influence of the groups of conglomerates C2 and C3, the first large group to carry out the census collection, and with the C4 and C5, the last to be done. This strategy strengthens the design with which we can guarantee that the effects of the information for the "contra-flow" has a low probability.

The collection is done starting on May 22nd in a reduced number of municipalities (one for each DANE territory), of the group of conglomerates of low-low dynamic, low and medium, which will permit us to make the final adjustments to all of the processes. It is important to take into account that if the resulting corrections or adjustments affect the comparability of the information with the rest of the municipalities, these municipalities should have the collection in the census moment that corresponds to the respective conglomerates to which they belong. The selection of these 6 or the municipalities required as initial municipalities are done under the following conditioning criteria:
-They should have the adequate cartography completed for the collection.
-The municipality should not be a central unit.
-Municipalities should be selected that are in the rural area that have an indigenous reserve, with the goal of observing these operative processes in this type of territory.
-They should be municipalities with characteristics of the rural area, in terms of transportation times by surface size, which will permit trying out the operative design for daily return to the municipal seat and the part related to the synchronization of information.
-Municipalities in which the paper questionnaire is applied should be considered.
-Equal distribution of municipalities for each of the DANE Territories.
-Should not belong to conglomerates with middle, middle-high, high, and special level.
-Should not be close to capitals or intermediate cities or centroides.
-Should constitute a homogenous group in terms of characteristics and proximity.

In this order of ideas, the municipalities of low-low and low dynamic are distributed according to their location in regards to the centroids in each department and

Page: 27

also considering in various cases including various departments, maintaining the basic principal of the model that the collection is done in the same "direction of the special mobility". Said design can be seen in the following map:

[#tr note: the rest of the page is blank]

Page: 28

[map labeled:]
DANE
General Census 2005
Work phases according to
Conglomerates of municipalities

[on map:]
Venezuela
Pacific Ocean
Ecuador

Peru
Brazil

Phases / Municipalities
F0 / 6
F1 (C2-C3) / 374
F2 (C4-C5) / 63
F3A (C1A) / 312
F3B (C1) / 266
Dissolve_Territorials_01

[#tr note: page is blank]

Option 4: Given the inconveniences presented in the licitation process for the acquisition of the PDA, as well as the parameters defined by the General Management of DANE and the budget restrictions[13], this strategy implies carrying out the collection for a group of conglomerates in 2005 and another group in February, 2006.

The Wyauu ethnicity, whose territory corresponds to the middle and high Guajira, presents a special case as a group that is characterized by a high mobility that depends on the behavior of the rain and on their labors in the maintenance of their cattle, which generates a high mobility between Columbia and Venezuela, as well as towards the rest of the municipalities of the low Guajira.

Considering the above aspects and with the goal of minimalizing the gaps that can be introduced in the structure and levels of census information by populational dynamics factors, the operational implementation of the Population and Dwelling census under this strategy should take into account the basic aspects defined in the methodological model:

1. The collection is simultaneous in all of the conglomerates, located in the different regions of the country, that conform each of the different groups.
2. In the interior of each conglomerate, the collection is carried out in simultaneous form in all of the municipalities that conform it. The census moment is 00.00 hours of the day of the initiation of the collection of each group.
3. The spatial distribution and execution of the census of the conglomerates is carried out from the group with the lowest dynamic to those with the highest.

The necessity of carrying out part of the collection in February, 2006 and considering aspects such as elections and the period of regional festivities affects point three, which generates the following limitations:
-Due to the characteristics of the first trimester of 2006, the conglomerates with the greatest dynamic cannot be done given the implications of the electoral process.
-The high influence in great part of the Atlantic Region, especially of the Bolivar department towards the north, of the carnival of Barranquilla, means that all of this sub region cannot be done in 2006.

---

[13] Minimal utilization of paper; Maximum 14,300 PDA.

-The direction of the population movements towards the principal centroids must be taken into account, with the goal of minimalizing the effects of counter-flow.
-The group of municipalities that are done in 2006 should be primarily low-low and low conglomerates with complex geographic characteristics or far from the principal centroids.
-For this selection the unique dynamics that exist in the interior of certain regions of the country should be taken into account, as is the case of the south of the department of

Page: 31

Bolivar, south of Cauca, the axis of Arauca-Casanare, west of Nariño, as well as the axis Aruaba Antioqueño-Cordoba-Sucre.
-The restitution of the single statistical moment for the group of municipalities done in 200 implies the development of process of conciliation with the goal of obtaining the adequate parameters for carrying out the [retrepoyección #tr note: this word is unique to this document. It is not in any dictionary that I used and a web-search leads to this document only], an aspect that requires time. It is important to take this point into account, considering that the date May 22, 2006 defined as the official date for turning in the census data, given that in addition to said processes of conciliation, all of those for depuration of the census information must be done.

Considering all of the above, and with the goal of guaranteeing the basic principals of the model supported by the conglomerates, as well as the adjustments for countering the possible effect of not doing the collection from the groups with lowest to highest dynamic, the following additional parameters are defined for the collection model:

a. Carry out the collection in all of the department of the Guajira and the Resguardos de la Sierra Nevada of Santa Marta between the 1st and 14th of September, 2005. With this, we guarantee the following aspects: avoid the effect of the rains in the middle and high Guajira and simultaneously do all of the area of influence of the regional centroids of Riohacha, Uribia, and Maicao with which the gaps due to mobility are minimalized for this sub region.
b. Carry out the collection for the conglomerates of middle, middle-high and special[14] between October 8th and December 15th of 2005. Considering these large centroids, all of the municipalities of their areas of influence will be done simultaneously as well as those that define the corridors of populational mobility.
c. Carry out the collection between the 1st and 28th of February of 2006 for the municipalities that are more distant from the principal centroids, as well as the axis Arauca-Casanare and Uraba Antioqueño-Cordoba-Sucre. These last are included considering that the centroids of Uraba, Monteria, Sincelejo and Arauca, Yopal generate in each sub-region a very individual dynamic for which the collection should be done simultaneously with regards to their areas of influence.

Page: 32

d. III. Restitution to a single statistical moment.

As is described in the previous point, each group of conglomerates will have a different census period and, therefore, a census moment at the initiation of the collection[15]. This means that the census information for each of the municipalities of the country have different reference times, which does not

---

[14] Bogota DC and 24 nearby municipalities.

[15] See definitions of the point 5.1 of this chapter.

permit obtaining the accumulations that are departmental, regional, and national as well as dwelling, households, and population.

The above requires taking the basic population, dwelling and household information to a single statistical moment, for which the following considerations should be taken into account:

-It should be located close to census moment for the group of conglomerates with the most populational dynamic, which means that said municipalities that conform it will not have any type of adjustment.

-With the goal of avoiding problems with the population data by the territorial entities, taking into account that these know the data observed, it is better to take these values forward and not retroproject [#tr note: retropoyectar [not standard spelling] is translated here as retroproject.

-The totals of population for each municipality are projected and the structures by age and sex are maintained constant.

-The projection of these totals, carried out considering the compensating equation [#tr note: compensadota is not standard spelling; here it is translated as compensating], from which a growth factor is obtained for the period given by factors of spatial mobility and natural growth.

-With the questions of control an additional factor can be estimated per conglomerate, which would be applied to the population structure, determined by the information obtained from the questions of the control section regarding the volume of persons who transfer to municipalities that have already been enumerated and who were not enumerated in the municipality of origin.

-An adjustment to the population structure is made when duplicated persons are detected. For this sampling systems should be made in which the information is taken for the control variables and crossed with the name, age, dates of birth, family relationships, marital status and educational level. In the case that this comes up for each municipality and, among municipalities, the following situations should be taken into account:

a. If the person moves to the interior of the seat or rural part of the same municipality and is detected on more than one occasion, he/she will be counted one single time, and for the purposes of the census will be left in the last place in which he/she was enumerated.

Page: 33

b. If the person moves from the center to the rest or from the rest to the center of the same municipality the person is counted in the last place where he/she was enumerated and he/she is erased from the initial place.

c. If the person moves among municipalities, he/she is left in the last and erased from the rest.

-The percentages of leaders are calculated by municipality and according to its dynamic which can be used for projecting the households and dwellings. According to the dynamic, it can be considered that those groups of conglomerates classified as low and low-low, said volumes in short term as is the time of development of the census, are maintained as constants.

-The demographic indicators that are used will permit characterizing each conglomerate, as well as other additional [indicators], with which monitoring should be done on the quality of the information of each municipality. The additional indicators are: masculinity index, relationship children/woman, brute level of birth, general fertility rate, brute mortality rate, structures by age and sex. At the same time, estimated volumes will be obtained for population as the census advances in each municipality, a value that is obtained in this manner:

a. The value of the population in the first two days is taken and under the supposition that the products are the same project the population lineally for the skipped days to complete the census period of the conglomerate.

b. On the fourth day the accumulated is taken and the rest of the census period for the conglomerate is projected. This is done successively as the census advances.

c. Each estimation is compared with the indicators contained in the Sampling Base and the consistency is analyzed. The first level of analysis is done by the person responsible for the municipality and should be reported to DANE Central.

Taking into account the previous points as well as option 3 defined with the conglomerates, the statistical moment will be located on the date that corresponds to the census moment of the last conglomerates that begin on October 20, 2005 and the correspond to those with the highest populational dynamic. This means that the group of conglomerates C4 and C5 do not require any type of adjustment.

Below we present the mathematical formula for the restitution of the population to a single statistical moment.

The process of populational change and its components in a municipality z (z= 1, 2, 3, …..1119), belongs to a conglomerate j (j= 1, 2, 3, 4, 5, and 6), is resumed in the so called Compensational Equation that is presented below.

[#tr note: the following should have subscripts and superscripts. The first ( ) after the capital letter should be subscript; the second ( ) should be superscript.]

$$N_{(t+n)}^{(z,j)} = N_{(t)}^{(z,j)} + B_{(t, t+n)}^{(z,j)} - D_{(t, t+n)}^{(z,j)} + I_{(t, t+n)}^{(z,j)} - E_{(t, t+n)}^{(z,j)}$$

Where:

Page: 34

$N_{(t+n)}^{(z,j)}$ is the population of the municipality in the moment t+n.

$N_{(t)}^{(z,j)}$ is the population in the census moment of conglomerate j.

$B_{(t, t+n)}^{(z,j)}$ are the births in the period t to t+n.

$D_{(t, t+n)}^{(z,j)}$ are the deaths in the period t to t+n

$I_{(t, t+n)}^{(z,j)}$ are the immigrants to z in the period t to t+n

If $N_{(t+n)}^{(z,j)} - N_{(t)} / N_{(t+n)}^{(z,j)} = B_{(t, t+n)}^{(z,j)} / N_{(t+n)}^{(z,j)} - D_{(t, t+n)}^{(z,j)} / N_{(t+n)}^{(z,j)} + I_{(t, t+n)}^{(z,j)} - E_{(t, t+n)}^{(z,j)} {}^{(z,j)} / N_{(t+n)}^{(z,j)}$

The equation is equivalent to:

[#tr note: the following should have subscripts and superscripts. The first ( ) after the capital letter should be subscript; the second ( ) should be superscript.]

$$r N_{(t+n)}^{(z,j)} = TBN N_{(t+n)}^{(z,j)} - TBM N_{(t+n)}^{(z,j)} + TNM N_{(t+n)}^{(z,j)}$$

Where:

$r N_{(t+n)}^{(z,j)}$ rate of growth of the municipality z in the period t to t+n.

As contributions relative to this growth are:

TBN N(t+n) (z,j) gross birth rate of the municipality z in the period t to t+n.

TBN M(t+n) (z,j) gross mortality rate of the municipality z in the period t to t+n.

TNM(t+n) (z,j) net immigration rate of the municipality z in the period t to t+n.

With this estimated relative growth for each municipality of the conglomerate j, it is then possible to project the population totals for the single statistical moment in the following manner:

If the projection is done forward, the total in the moment t+n (t in the census moment of conglomerate j) is:

[#tr note: the following should have subscripts and superscripts. The first ( ) after the capital letter should be subscript; the second ( ) should be superscript.]
$PT_{(t+n)}^{(z,j)} = PT_{(t)}^{(z,j)} * EXP(r(t,t+n)(z,j)*\Phi)$

Where:
PT (t+n)(z,j) is the projected population of the municipality z of the conglomerate j in the moment t+n

PT(t)(z,j) is the total population obtained through the census in the census moment t

Φ is the time, in days, between the census moment and the statistical moment t+n for the municipality z of the conglomerate j.

If the projection is backwards, the total in the moment t+n is:

$PT_{(t+n)}^{(z,j)} = PT_{(t)}^{(z,j)} * \{EXP(r(t,t+n)(z,j)*\Phi)\}^{-1}$ [#tr note: the last "-1" is superscript]

To obtain the population by structure according to simple age and sex, as it is supposed that the period defined by Φ is relatively short, we have:

If Φm and Φh is the structure by sex of women and men respectively and βx is the percentile distribution by age x and sex, then the population by age and sex of the municipality z of the conglomerate j is:

For the structure of women by simple age:
$PT(t+n,x,m)(z,j) = PT(t+n)(z,j)*(\Phi m*\beta(x,m)$

For the structure of men by simple age:
$PT(t+n,x,h)(z,j) = PT(t+n)(z,j)*(\Phi h*\beta(x,h))$

With the goal of carrying out the monitoring to guarantee the consistency and quality of the information, we should build the Table of Indicators which forms part of the System of Monitoring and Control in which all of the existing information for each municipality on population, births, deaths, lots, users of public services, as well as the strategic indicators mentioned in the above formula should be placed.

We should take into account the population data observed in the census of 1993, as well as the data from the projections for each of the municipalities[16], with which we build reliable intervals with the goal of determining the lower and upper limits in which we can locate the population estimates. This estimate is done in the capitals taking into account that only for said area can we count on the census information in a timely manner was it is collected through the DMC[17]. The process of estimating is done taking the information by days or weekly, depending on the size of the capital and with an expansion factor we estimate the population that will be obtained at the end of the collection period.

This estimate, depending on the location within the interval, determines a series of alarms that can be related with coverage problems, management of the concept of habitual resident among others, with which we can make the respective corrections in the field. This information is contrasted with some indicators such as Masculinity Index[18] or Relationship Children/Women[19] which can explain effects of migration, male excess-mortality, infant mortality, omission of boys(girls) among others. See the annex where you will find some of these indicators.

[#tr note: terms found on illustration:]
45 Days Municipal Collection
Week(s) later
ENC Coverage
Indicator 1
Indicator coverage dwelling, household, persons

Page: 37

As can be seen in the above graph, this process is done by accumulating the information, with which we can obtain other strategic indicators, according to the conglomerates that they belong to, information that should be managed not only at a central level, but also at local level and at the level of the Territorial Offices of DANE.

These indicators are of great use in the case of using the capitals as a coverage control with verification and correction of errors, guaranteeing not only a high index of coverage in observation units, but also an adequate consistency and statistical completeness of the information both in the basic questionnaire and the extended questionnaire.

Other important aspects to take into account is that with the information of "Control of mobility", we can construct indicators that will permit quantification of the effect of the mobility and identify the number of cases of "duplicated persons", which, if the values turn out to be significant (more than 5%), can be considered in the restitution part.

Page: 38

IV. Thematic Content

---

[16] As much in the capital as the rest
[17] Mobile Collection Framework (Diapositivo Movil de Captura)
[18] This index is an indicator of the structure of the population and sample of the differences in the composition by sex and its formula is IM=(total men)/(total women)*100. This expresses the fertility conditions in a simple manner as well as the effect of infant and juvenile mortality. Its formula is calculated RNM = (Children from 0-4)/(Women in reproductive age).
[19] This indicator of fertility is of much usefulness when there is no existing information or for small populations.

The Census, due to its population and geographic universality characteristics, as well as the diversity and integral characteristics of the themes that it seeks to cover, constitutes the fundamental and un-substitutable base of knowledge about the demographic, economic, social and cultural conditions of the population of the country.

The thematic defined to obtain strategic information through the general census, is built starting with the following premises:

-For the definition of the census themes we have taken into account the informational needs of the country, the actualization of the existing data, its nation-wide comparability (with previous census) and international comparability (with other countries of the world).

-Idealness of the questions to guarantee the quality of the information.

-For the collection of the census data, we will use intelligent questionnaires developed under specialized software for pocket computers or PDA. In special cases, questionnaires designed on paper will be used.

-Each variable should have associated geographic variables that permit, on one side the geo-referencing of the information and production of tabulations or crosses of variables according to the needs of the users.

-Guarantee the basic information for the calculation NBI, migration, mortality and fertility, as well as those that permit obtaining the volumes of dwelling, households, and persons. This last according to structures of age, sex and ethnic groups[20].

-Characterization of the dwelling units for each geographic unit.

-Characterization of the population for each geographic unit.

-Construction or actualization of framework for carrying out polls on social, demographic and economic themes.

-One single questionnaire will be used considering that the census of population and dwelling will be done simultaneously and in this manner guarantee the relationship between each dwelling unit and the data related to the habitual residents.

-A basic content is defined which should be applied to all of the universe of each municipality and an amplified content to be applied to a representative sampling in each municipality (capital and the rest), for each of the strategic thematic contents as is the case of the variable for calculating the NBI.

-Basic questionnaires will be used for variables on economic and agricultural units for building frameworks.

-The thematic content is constituted in the basic input for the structuralization of the Basic Statistical System of Columbia.

Page: 39

5.4.1 Population and Dwelling Census

---

[20] Indigenous according to each ethnicity or town or black population.

The thematic content of the census questionnaires is defined [#tr note: the Spanish version is literally "The thematic content are defined"] according to the institution and technical requirements in terms of the need for representation of basic census information for small areas. In effect, it is important to consider a series of conditioners that define the contents of the basic formulary, which will be applied to all of the universe [group]. These parameters are the following:

-The prevalence of the many phenomena requires very large samples with high benefit costs, which determine their obligatory nature for inclusion in the Basic. This is the case of the variable that permit measuring the indexes of NBI and ICV, as well as population by age and sex, migration, movement, fertility, and infant mortality.

-Aspects of legal nature such as is the case of the information that must be certified by the DANE.

-Information that because of its characteristics is constituted in strategical information for the establishment of base lines in the processes of planning, follow-through, and evaluation of the local public process.

-Given the frequencies or totals that are needed in subpopulations, in the case of information by sex and age, to obtain robust information requires its application to the total of the universe in each municipality. This is the case of the information on migration, movement, level and grade reached in education as well as fertility and infant mortality.

Under these two parameters, the Thematics for each of the questionnaires are the following:

Basic questionnaire

It is applied to the entire universe, capital and the rest, of each of the Columbian municipalities with a total of 28 questions. Its content is the following:

-Identification: This is a chapter made up of 8 items that permits relating the census information in terms of dwelling, household, and persons at the level of each territorial entity (capital, population center, disperse rural), as well as the territorial part for indigenous reserves and collective territories of black communities.

-Dwelling and household: This contains two questions with which we can quantify the inventory of the dwellings according to topologies defined for the census (House, apartment, room, boarding house, and other type of dwelling) and, at the household level, the part related to international emigration.

-List of persons who make up the household: Under the concept of Habitual Residence[21] and ordered according to relationships, this permits carrying out a previous control related to size of household. Additionally, there are two controls included with the goal of detecting persons who have been excluded and who are habitual residents of the household[22] and in the case of the contrary, minimalize the gaps by

---

[21] This is the person who lives permanently or the majority of the time in a dwelling although at the moment of the census he/she is absent.
[22] The persons tend to omit as residents minor children, the elderly, persons interred in clinics, the persons kidnapped or those who at this moment are on vacation.

including persons who are not residents of the household as in the case of the students who spend the majority of the time in the city where they carry out said activity. A question is included about "deaths in the past year" which is of great importance for the evaluation of the death registries.

-Data about Resident Population: This includes questions with the goal of carrying out controls to avoid omissions or duplicity. The questions are additionally controlled be done according to age and sex of the person. The thematic is the following: sex, age, relationship to the head of household (male or female), ethnic background[23], disability, migration[24], force movement in the past five years[25], education[26], fertility and morality[27], variables that because of their frequency at the level of age and sex are required to be applied to all of the population.

-Control Questions: This includes a series of questions that permit identifying: if [the person] is a habitual resident of the census household or of another; if [he/she] was enumerated in another place in the same municipality or in another municipality. Equally, the is a question included that permits controlling if there is an economic activity associated with each household, which determine if it is necessary to apply the additional questionnaire about industry, commerce, or services, as well as in the case of the rural area if there is an agricultural activity present.

Amplified [Extended] Questionnaire

This instrument is applied to a representative sampling for each of the municipalities[28] of the country with a total of 80 questions. In the case of the very small municipalities, as well as the indigenous reserves and the collective territories of black communities, this questionnaire will be applied to the entire universe. The thematic content includes that of the basic questionnaire plus the following themes:

Page: 41

-Dwelling: This is a chapter with a total of 10 questions that permit the characterization of the dwelling; estimate qualitative and quantitative deficit, as well as the required variables, in terms of dwelling, for the calculation of NBI and ICV.

-Household: This is a chapter with a total of 15 questions in which variables are include for estimating the subjective poverty and international migration.

-List of persons who make up a household: This contains 6 questions. Additionally this asks about the contribution to the household expenses, which permits carrying out an analysis of the information with perspective of genres for the first time.

-Information of Resident Population: This chapter contains 43 questions, some correspond to new thematics such as poverty measured from the perspective of food consumption, health, and social security among others. This complements the part about education by asking about the cause for not attending school, foreign language, computer use, and reading habits. Equally, in the economic part it asks about the place where the person does work, the address of work, occupational category and income. According

---

[23] The black population is measured by two aspects: auto recognition of phenotype and culture.
[24] This permits measuring internal and international migration. The first is from the capital or rural part of a territorial entity.
[25] Identify the year and the origin.
[26] Literacy, attendance and last year passed.
[27] This applied only to women 12 years old or more.
[28] Capital, the rest, and some cities at the level of locality.

to a technical meeting with officials of the Ministry of Agriculture, this includes a question for women 12 years or older who live in the rural part about principal economic activity that is done by women.

Control questions: They are identical to the basic questionnaire.

Basic questionnaire for Special Place of Housing – LEA [initials in Spanish for 'Special Place of Housing']

This questionnaire is necessary to be able to capture the basic information of habitual residents of each territorial entity, which because of special characteristics of work, study, members of religious groups or reclusion, fulfill the criteria of Jure for the census of population and Dwelling.

The above means that the population by age and sex of each of the different municipalities is conformed by all of the persons enumerated both by the questionnaires applied to the private homes as well as in the LEA that are found in the urban and rural areas.

The thematic content for the population that resides in the LEA should not only capture the structures by age and sex in each of the types of LEA[29], but also the basic characteristics that are obtained for the population that resides in private homes. Equally, this includes some varibles that permit the characterization of the different insitutions.

Page: 42

Note: For the characteristics that exist in each municipality, given that many of these LEA are located in specific places, the personnel that applies this questionnaire should be different from that of the private homes. An activity that should be done by the line personnel is identify in advance the LEA that exist in each municipality, and including apply them before the population and dwelling census.

In the annex document "Thematic Content of the General Census 2005" we can observe the structures of the different questions that contain each theme in each of the questionnaires.

5.4.2 Economic and Agricultural Units.
The content of these two questionnaires was coordinated with the groups in charge of the theme, arriving at some basic contents that can be seen in detail in the Annex.

5.4.3 Urban Environment
This is a new theme within the census process with which we amplify the use of the information in a significant manner by being able to relate the population data and its habitational characteristics with the urban environment that surrounds each dwelling unit.

However, it is necessary to be careful with the operative process of collection to guarantee that the information that is obtained is pertinent.

From the methodological point of view, the unit of observation is the "side of the block", taking into account that operatively, all of the information is "tied" to the AG [#tr note: there is no definition of AG here.]. According to this, the questionnaire should be "unique [or single, solitary] for the side of the block" and apply as much the habitat criteria as the physical characteristics of the roads, paths, etc.

---

[29] Jails or centers of reclusion, orphanages or homes, convents or seminaries, boarding schools, barracks, work camps, persons who live on the street, refugee shelters, houses of prostitution, among others.

To achieve putting this process in operation it is necessary to identify if each AG is for "residential"[30] or "non-residential". Only in the AG for residential use do you apply the criteria of habitat.

Page: 43

V. Norms for managing the census information of population and dwelling.

For the conformation of the archives of the general census 2005, as much in the collection of the information as in the final consolidation a series of aspects of a technical nature should be considered with the goal of guaranteeing the consistency of the information and therefore its posterior use on the part of the different users.

In the development of a census project we should consider an important phase for obtaining quality information for an adequate use on the part of the users, in which we carry out the depuration of the census data.

This phase of depuration, in the model of the General Census, has the following characteristics.

1. The logical and physical design of the Data Bases that contain the census data.

2. The norms of consistency that should be implemented in the DMC for the intelligent collection of the census data.

3. The validation norms of consistency and data imputation. This process is delicate and therefore it should be done as soon as all of the data is consolidated by municipality, given that it is required to evaluate the effect of the imputation on the structure of the data.

In the generation of the computational tools, as well as the physical and logical design of the census archives the following norms should be taken into account:

-Those of structural nature: This consists in the verification and correction of the structural variables that come in the census form, which guarantee that the information for each of the units of observation can be geo-referenced and feed the conformation of the different tabulations or crosses ov variables for the analysis of the information. These norms have to do with:

Identification variables that define the pertinence of the questionnaire to a geographic unit and are: Department, Municipality (capital, populational center, the rest of the rural area, block) and that are contained in the AG; number of the dwelling; number of the household; number of the questionnaire; number of the additional form[31].

Content variables constituted by variables of each chapter that are basic for each unit of observation. In the case of the census of Dwelling and

Page: 44

---

[30] If in the AG there exists as a minimum on dwelling.
[31] For the case of the paper questionnaires. For the intelligent forms, the system should assign a consecutive [#tr note: the sentence ends in this manner, which seems to be incomplete]

Population they are the following: Condition of occupation of the dwelling; number of the persons; sex; age; relationship; educational level and years completed; civil status.

Those of Consistency: They are norms that give the guide [option] for verifying the census information. The consistency data are those that are within the permitted range for the variable and that also maintain a logical relationship with other variables within the same chapter or with variables from other chapters.

Below the general norms are defined in each of the types previously announced:

1. Structural for the Population and Dwelling Census.

In the annexed document "Norms for the Control of Structure" all of the processes that are required to be taken into account are detailed.

2. Internal consistency of the Population and Dwelling Census

The norms of consistency that should be taken into account to automatically complete the form at the moment of collection when using the DMC should be the most general ones that do not slow down the collection.

Below we present some of the norms of general consistency for the chapters of dwelling, household and persons, considering that the contents of the questionnaires are not officially closed yet.

-In the variables of municipality, department and country there should be valid codes.

-The total of households existing in the dwelling should be equal to the same as the count of the registries of households for said dwelling.

-If the question Condition of occupation of the dwelling is "Occupied with all persons absent", in the question Total Households of the dwelling should be 01.

-If the result of the interview is dwelling "Unoccupied", in the question of Total households of the dwelling should be 00.

-If in the type of dwelling you have Apartment or Type of "Room" and the material of the walls is zinc, fabric, cardboard, cans, discarded materials, etc. or without walls, the type of dwelling in House.

-If in the type of dwelling you have Apartment and all of the households of the dwelling say that "They do not have sanitary service", the dwelling is Type "Room".

Page: 45

-If the dwelling does not have energy service but some of the households of the dwelling cook with electrical energy, or have: TV or air conditioning, then the dwelling does have electrical energy.

-If there exists only one household in the dwelling, the use of the sanitary service is exclusive for the household.

-If in the household they do not cook, the question of where the water comes from for preparing food should be blank.

-If in the household they cook and they count with service of aqueduct, then for where the water comes from to prepare the food should be aqueduct.

-The questions of age, sex and relationship should not be blank.

-Each woman who is 12 years old or more should have information for fertility.

-If in the age of the person you have 00 in the question of migration in the past 5 years should be that the person had not been born and the question of where the person lived should be blank.

-The codes for department and municipality associated with the AG are the same as the name of the municipality and department of the option 3 in "in another municipality", the information is incorrect and it should be changed to option 2 "in this municipality".

-The level passed in education should be consistent with the age (the control table should be constructed).

-If in relationship you have "Domestic employee" (male or female), in "the greatest part of the time last week" should be code 1 or 2; in the question of occupation there should be the code related to household jobs; in the occupational category code 4 should be marked and in the question about economic activity you should write private home with domestic employee.

-The total of surviving children, sum of the total of men and women in question 75, should not be greater than the total of children born alive.

-Tables for homologization should be built for processes of imputation of sex, age, educational level and civil status.

The following norms are defined in a specific manner for each of the themes:

Chapter II. Dwelling information

Page: 46

-This chapter should only be done for household 01 of each dwelling.

-This chapter should not be completed in the additional forms, when the household is number 01 of the dwelling and it has more than 16 persons.

-If in the question "predominant material of the exterior walls" there is a mark for code 6 (cane, branches, other vegetable) or code 7 (zinc, fabric, cardboard, cans, discarded materials, plastic), then, the type of dwelling should be code 1 (house).

-If in the question "type of dwelling" there is a mark for code 5 (other type of dwelling), then in the question "predominant material of the exterior walls" the code of the answer should be 8 (zinc, fabric, cardboard, cans, discarded materials, plastic), or code 8 (without walls) and in question 8 (floor materials) the code should be 6 (dirt, sand).

-If in the question "wall materials" the code of the answer is 6 (cane, branches or other vegetables), 7 (zinc, fabric, cardboard, cans) or 8 (without walls), then in the question 8 (floor materials) the code of the

answer should be 3, 4, 5 or 5 (rough wood, boards, plywood, other vegetable, cement, gravel, dirt or sand).

-If in the question "type of dwelling" the code of the answer is 5 (other type of dwelling), then in question 10 (public services), the only service the can have a mark in code 1 is the electrical energy service.

-If in the question "type of dwelling" the code for the answer is 2 (apartment), then, in the question "sanitary service", the code for the answer should be 1, 2, or 3 (toilet connected to a sewage system, toilet connected to a septic tank or toilet without a connection, latrine, pit) and in the question "site for food preparation", the mark should be in code 1 (in a room used only for cooking), or in 3 (in a living room/dining room with a sink for washing dishes) or in code 6 (in no place, they do not prepare food).

Chapter III. Information about the household

-This chapter should be unanswered in the additional forms.

-the number of rooms in the questions "rooms that are available to the household" should be less than or equal to the rooms in the question "rooms available in the dwelling" in Chapter I.

-The number of rooms in the question "rooms used exclusively for sleeping", should be less than or equal to the number of rooms in the question "rooms available to the household".

-In the question "rooms available to the household" the number should be between 1 and 20.

-In the question "rooms used exclusively for sleeping" the number should be between 1 and 20.

-If in the question "type of dwelling", the code marked is 3 (room) or 4 (other type of dwelling), then, in question "rooms available to the household" the number of rooms cannot be greater than 3.

-If in the question "type of sanitation" the code for the answer is 1 (toilet connected to sewage system), then, in the question "services of the dwelling" should have the mark in code 1 for sewers.

-If in the question "type of dwelling", the code of the answer is 5 (other type of dwelling), then, in the question "site for food preparation", the mark should correspond to codes 2, 4, 5, or 6.

-If in the question "connection to public services", the dwelling does not have electrical energy (code 2), then, in question "cooking fuel" the answer should be different from code 1 (electricity).

-If in the question "water for food preparation", the mark corresponds to code 1 (aqueduct), then in the question "public services", aqueduct should have a mark in code 1.

Chapter IV. List of persons who conform the household

F1. Resident Population

-The total of persons in the household should be equal to the number of persons interviewed.

-The number of order registered in question "greatest contributor", should correspond to one of the numbers of order registered in the question "list of persons who make up the household". In some cases the number can be "00" or "98".

-The number of order of the person registered in question "greatest contributor", should correspond to a person who in the question "age" is 12 years old or more.

-The valid age to be the head of household (male or female) is: for men it is 14 years old or more; women 12 years old or more.

-The number of order that corresponds to each person in the list of persons who make up the household, should be the same that is registered in the question "order" for this person.

-If there is a date of birth, it should be consistent with the declared age of the person.

-If in the household there is spouse, husband/wife, companion (male or female) or couple (question "relationship to the head of household=2"), then, this person should be of the opposite sex from the head of household. During the taking of information a window for verification should appear, since there can be an error in assigning the relationship or in the sex. If you corroborate that the sex is the same as the head of household, for relationship 2, leave it as it is.

-In every household the person with order number 01 in the list of persons who make up the household, there should be a question "number of order of the person" code 01 and in the question "relationship" code 1 (head of household, male or female).

-In the household there is always a person in the question "relationship" with code 01 (head of household, male or female) and this should be the only one.

-In the household there should only be one person in the question "relationship" with the code 2 spouse (husband/wife), companion (male or female), couple and this should correspond to the order number of person 02.

-If in the question "sex" the answer corresponds to code 1 (man) and the age is less than or equal to 13 years old, then, in the question "marital state" the code of the answer should be 5 (single).

-If in the question "sex" the answer corresponds to code 2 (woman) and the age is less than or equal to 11 years old, then, in the question "marital state" the code of the answer should be 5 (single).

-According to the age of the person, the following assumptions should be followed:

[#tr: the following is the text only from a table]
Age (years) / Sex / Relationship / Filters (blank) / Migration / Highest year completed

Man or woman / Son, grandson, brother, other relative, other non-relative/ Son, grandson, brother, other relative, domestic employee, other non-relative

Man / any relationship / according to the table in the question "last level passed"

Additionally, the following relationships:

-(Age of the head of household minus the age of the son) >=14
-(Age of the head of household minus the age of the grandson) >=30
-(Age of the father minus the age of the son) >=14

Migration:

-The codes should be valid according to the divipola [División político-administrativa= Political-administrative division] or of the countries.

Page: 50

-If the age is under 5 years old then in the question "where did he/she live five years ago" should be the option "had not been born".

-If the department and municipality are equal to the AG:
     -If it is the question of birth, then the code is 1.
     -If it is the question of residency five years ago then the code is 2.

Movement:

-So that the information is valid it should fulfill: 2005 minus the year of movement should be less than the age.

-Codes of municipality and department according to the political-administrative division.

-The year is valid if 2005 minus the year of movement is less than or equal to five.

Ethnic group:

-If the person auto-identifies himself/herself as indigenous he/she should be asked "group" and if "he/she speaks the language".

-If the person auto-identifies himself/herself as "ROM", there should be information about if "he/she speaks the language".

-Any of the other options should not have information about "group" and "language".

Information about poverty:

-If in the question "due to lack of money he/she did not eat", the answer is code 1, then in the field for "how many days" should be asked and only accept numbers 1 to 7.

-If in the question "food consumed" "yes" is marked, necessarily there should be an answer in the field for number of days with values from 1 to 7.

Education:

-If the age of the person, minus the total years of study (sum of years completed counted starting with basic elementary through the last year of studies passed), should be greater or equal to 5.

-If in the question "place where the computer is used", there is a mark in code 1 "in this dwelling", necessarily in the question "ownership of goods" computer should have the minimum mark "01".

Page: 51

-If the person is between 3 and 5, then the only answer possible for the question "in the last week he/she used the computer" is the alternative "entertainment."

-If the person in the question "place where he/she uses the computer" is the option "in the place where he/she studies" then the person should be attending an educational establishment.

-To carry out the process of validation of consistency of the variable "Last year passed", the age of the persons must be taken into account, for which the following hears of studies are considered as the maximum that a person can have according to his/her age[32].

Economics:

-If the relationship comes as "Domestic employee (male or female)", in "During the past week" should be code 1 or 2; in the question of occupation there should be a code related to "household occupations"; in the occupational category option 4 should be marked and in the question about economic activity should be written "private household with domestic employee".

Fertility and mortality:

-The total of surviving children, the sum of the total men and women, should not be greater than the total of children born alive.

-For women from 12 to 15 years old, the difference between the census year and the year declared[33] as "birth of last child", plus 11 should be less than or equal to the age of the woman.

-For women from 51 to 54 years of age, the year of birth of the "last child" should be less than or equal to the difference between the census year and the age of the woman minus 50.

-The year declared for the birth of the "first child born alive" of a woman 12 years old or more, should be consistent with her age, in this manner: the difference between the "census year" and the "year declared", plus 11 ((2005 minus one year) plus 11), should be less than or equal to the "age of the woman."

Independent from collecting the census data with DMC and, considering that in some case the cases of collection is with a "paper form", the final census archive should be purified, which means that carrying out the process of "Validation of Consistency and Imputation". In the annexed document "Norms for Control of Validation of

Page: 52

---

[32] See annexed document "Norms for control of Validation of Consistency and Imputation".
[33] If the option "6" is marked (2000) or earlier, assimilate "2000" as the year declared.

Consistency and Imputation: indicates for each of the questions of the Sections of Dwelling, Household, and Persons the structure for the validation and/or the imputation.

Page: 53

Annex: Indicators of Control of Quality and Evaluation

Page: 54

Code / Indicator / Type / Unit

Dwelling
1 Total dwellings
1 Rate of increase of dwellings
2 Total dwellings according to typology
1-2 Total dwellings according to condition of occupation
1-2 Percent of occupied dwellings
1-2 Percent of unoccupied dwellings

Households
1 Total households
2 Total households by type of services
2 Users of public service according to typology
1 Coverage of public services
2 Number of rooms per household
2 Kitchen or room for cooking per household
2 Households according to source of water for food preparation
2 Sanitary services by household
2 Households with exclusive sanitary services
2 Households according to form of trash elimination

Persons
1 Total persons
2 Persons per household
2 Persons per dwelling
1-2 Population density
1-2 Inter-census population growth
1-2 Population in the capital and in the rest
1 Persons by age -0-4 years
1 Persons by age -5-25 years
1 Persons by age -26 years or more
1 Persons by sex
1-2 Birth rate
1-2 General fertility rate
2 Global fertility rate
1-2 Relationship children/woman
2 Gross mortality rate
2 Infant mortality rate
1-2 Masculinity index

1-2 School attendance
1-2 Literacy
1-2 Literacy rates
2 Total persons without education according to age
2 Total persons according to educational level
2 Percentage of persons with higher education
2 Total persons according to type educational establishment that they attend
2 Total persons according to type of disability
2 Percentage of persons with disability

1-2 Total persons by Ethnicity or Indigenous culture

1-2 Total native persons
1-2 Total immigrant persons
1-2 Immigration rate
1-2 Total foreign persons
1-2 Total of displaced persons
2 Proportion of displaced persons
2 Total of persons according to Condition of Activity
2 Proportion of persons who study
2 Proportion of persons who work

Economics
1-2 Percentage of population economically active PEA
1-2 Percentage of population of age to work PET
2 Total of persons by branch of activity
2 Total of persons according to occupation
2 Total of persons according to occupational category