

### **Name of the public institution**

Federal Public Service Economy, SMEs, Self-employed and Energy (General Direction of Statistics and Economic Information - Statistics Belgium: see the website <http://www.statbel.fgov.be/> )

### **Name of the project**

Microcensus 2006

### **Acronym**

Microcensus 2006

### **Project description and definition within the framework of the institution**

The Federal Public Service Economy, SMEs, Self-employed and Energy - General Direction of Statistics and Economic Information, hereafter Statistics Belgium - is organising "General Censuses" since 1846. They were not only used to determine population size and structure but were also of a scientific nature. Statistics Belgium processes and disseminates results as tables and publications, but also as source material for research institutes of the government, universities and other interested parties researching in any field of social and economic interest. To date, the census is the most important source of social and economical information in Belgium. For many census data no other source is available.

In the future however, a major change will revolutionize census taking: in its session of 4 May 2000, the Council of Ministers has approved the proposition to organise a General Socio-Economic Survey on 1 October 2001 and at the same time has charged the Minister of Economy to draw a proposition amending the Statistics Act of 4 July 1962 in order to repeal the obligation of organising a decennial exhaustive census and to develop alternative possibilities to collect these data without having to appeal to an exhaustive survey, especially by making use of administrative databases, extant or still to establish.

The research on alternatives for a classic census was entrusted to the Higher Council of Statistics. In the meantime, some five academic research teams are investigating partial fields of the matter since two years.

Besides this research, it is of the utmost importance to start the practical development and feeding of the statistical databases, in order to guarantee the valorisation of former and actual research as well as the continuity of the census towards 2011. An "Administrative Pilot Census" will be taken to this effect in 2006. It should give the opportunity to build technical expertise and logistics for the transition to administrative censuses, but should also bring a first round of useful socio-economic data from administrative sources, right in the middle of the ten-year interval between two censuses. The pilot census will be taken from a 20 % sample of reference persons and the members of their

households. This will cover the entire Belgian territory with the required resolution. Consultation of major administrative databases (National Register of Natural Persons, Crossroads Bank for Social Security, Crossroad Bank for Enterprises, Land Register, databases on education and diplomas...) will cover most of the classical census fields.

Taking the census in 2006 is expedient for different reasons:

- it is exactly between the Survey 2001 and the planned follow-up in 2011: as one of the aims of the transition to administrative censuses was the shortening of the interval between censuses, this could initiate a five-year census system.
- There is enough time left before the administrative census to incorporate current research into the project.
- The preparation time of about one year enables us to undertake the necessary organisational steps: starting up statistical databases within Statistics Belgium; making arrangements with administrators of administrative databases etc.
- A five -year interval contributes to the practicability and the reference value of the results.

This project is part of a broader vision on the future development of the Belgian statistical system. The important role played by administrative databases will make new demands on data collection by means of large-scale additional sample surveys. Important demographic and social evolutions can be traced by making good use of administrative data. The main advantages of using administrative data are their high precision and their exhaustiveness. On the other hand, such statistics present several disadvantages. As the data reflect a purely administrative reality and not necessarily a social reality, they do not always allow for in-depth research into causal relationships. Hence the increasing demand for large-scale national sample surveys.

Besides existing surveys held by Statistics Belgium, a completely new panel survey was initiated: SILC (Social Income and Living Conditions). Statistics Belgium also started a joint venture with the regional authorities to investigate the possibility of a new demographic survey: GGS (Gender and Generation Survey). For the SILC survey, Statistics Belgium did considerable investments in manpower and hardware.

It is customary with statistical surveys to test the results against large-scale representative sources. To the present day, the census or "Social-Economic Survey" played this role. The reference value of the census could however be impaired by the new method of data collecting. This is why there is an urgent need for additional research based on data already gathered through the new system. The new situation should be explored soon in order to assure a lasting leading role for the administrative census in the future of the Belgian statistical system.

We finally would like to point out the ever-growing European dimension of the collection of social en economic statistics. Many large-scale sample surveys fit in with larger European projects or are used in reporting to European Statistical Institutes. The most common example is the Labour Force Survey. The European dimension is equally important for the transition towards administrative censuses. Some countries, especially in Scandinavia, have successfully made use of administrative databases for statistical purposes. Many others are preparing for the transition or are inquiring into the possibilities of the method. Within these countries a large range of possibilities has been tested and/or implemented, while statistical infrastructure was adapted to new methods of data collection in order to make it profitable to other functions of public interest. Obviously this proposition will also appeal to this expertise. The integration of results from databases and from surveys must also be seen in a European perspective.

## **Task to be accomplished**

### a. by the research team

The tasks of the research team can be broken down into five domains:

- 1) Drawing a list of variables and their specifications (moment, format...) to be requested from the administrators of databases, including the metadata required to exploit this information and to verify its validity and reliability. This includes that all necessary measures are taken concerning the data request and the sampling in order to harmonise sampling from different databases.
- 2) Assisting and advising Statistics Belgium with data exploitation. This includes all activities that are part of the process from the transaction of data as they are supplied up to the production of finished statistics: selecting sources and variables, coupling information of different origins, securing confidentiality, testing and quality analysis (e.g. comprehensiveness, double counting, internal consistency, time index of measurement, etc.) The research team will also have to set up data cleaning procedures. It will have to take into account specific features of the information from administrative statistical sources: i.e. the fact that this is fragmented information, having different time indexes and a method of collecting aimed at administrative rather than statistical purposes.
- 3) Systematic reporting and mapping the entire process, data flow, quality analysis and all features concerning coverage, reliability, topicality, comparability and other aspects of quality of the administrative method of gathering socio-economic information, as well as formulating recommendations to minimise or counteract all negative effects on the quality of the final results.
- 4) Preparing a scientific publication to valorise and disseminate the results of the administrative pilot census 2006, taking into account the uncertain quality of the final product. If this quality is sufficient, one could think about a very limited "Census Monography", focusing on the socio-economic shifts since 2001. If the quality and comparability to 2001 are doubtful, the final product will rather be a critical source study, comparing the newly acquired data with other sources from the same period of time.
- 5) Examine the data in an international/European perspective in order to allow for harmonisation/comparisons. The team will issue a report in this respect.

### b. by the public institution

#### *1. prior to the actual project*

Statistics Belgium and ICT will prepare the transaction of data files together with the principal administrators of relevant administrative sources: the Land Register (FPS Finances), the National Register, the Crossroads Bank for Social Security and the regional authorities in charge of education statistics. In general, cooperation with these authorities is good and administrative data are already exchanged. What matters is to structurally introduce the element of the administrative census into the co-operation and to organise consultation about this matter. The authorities in charge will be invited to deliberate

about the problem of data exchange for statistical purposes and to set up a structure for a long-term collaboration on census statistics.

Statistics Belgium will report about the state of affairs concerning databases that are being prepared to ensure the longitudinal follow-up of population, level of education, habitat etc. (DEWOLO project). This includes activities already taken up within Statistics Belgium and ICT in order to prepare or evaluate coupling and exploitation of alternative sources (socio-economic information from the Crossroads Bank for Social Security, Labour Force Survey, address information from the National Register, Land Register, etc.).

Statistics Belgium will take all necessary administrative and logistic steps to ensure the presence and availability of manpower in order to start up the further development of these databases as soon as the project is launched. This will require from Statistics Belgium and ICT about 5 FTE of qualified manpower. An internal project request P890 concerning these activities has already been submitted.

## *2. during the project*

The submitter will be in charge of:

- organising consultations with all administrators of administrative databases and delegating persons within the institution and the research team;
- treating with requests for information exchange and with the technical aspects of exchange protocols and formats, as well as with the follow-up of these requests;
- managing the administrative information for statistical purposes, making all couplings and supplying anonymous samples to be evaluated by the research team;
- publishing the final results of the proceedings, the format of which can to a certain extent vary (as stated above) in function of the quality of the information, and disseminating the final results along the usual channels (infoshops, internet,...).

### **Expected Final product**

#### a. by the research team

The final product expected from the research team consists in three parts:

- a *report* mapping and discussing the process and proposing alternatives and solutions along the line of action research. For the public institution, this will serve as a logbook to all activities concerning longitudinal statistical databases and their optimal development.
- a *statistical database* following as faithfully as possible the classical census format as a picture of social and economic reality. The degree to which this first "administrative picture" will be useful for further dissemination and scientific exploitation is to be specified but the possibility is left open from the outset. Obviously, this database is located within and administrated by the submitting public institution. The usual restrictions and clauses concerning privacy are as a matter of course applicable to research teams inside and outside the institution.
- A *published* final scientific product, ranging - depending from the quality reached - from a *limited census monography* to a *critical overview* of the present state of socio-economic indicators from different sources. It will include other sample surveys by Statistics Belgium.

b. for the public institution

The public institution agrees to deliver as a final product the set-up and updating of an exploitable database of all data gathered in the framework of this project and to observe to this purpose the same rules and practices that are usually applied to census information. However, the possibility is left open to observe some limitations regarding the dissemination of results of doubtful quality.

The final result made up by a database and the published results constitute an important surplus value for the FPS Economy, SMEs, Self-employed and Energy (General Direction of Statistics and Economic Information). It should however be emphasised that it is at least as important to anticipate in time the technical and intrinsic challenges relating to the method of gathering social and economic information. The information as a whole must indeed be assembled from elements collected from different time dimensions, from different viewpoints and with various purposes. For a transition of this magnitude and complexity it is essential to have a qualified research team, methodologically competent to tackle new challenges.

**Valorisation of the final product**

a. for the public institution: integrating the final product into its functioning

Paragraph 21 of the law of 2 January 2001 puts an end to the compulsory decennial general censuses and lays down instead the set-up and updating by Statistics Belgium of statistical databases fed from administrative sources. The Minister of Economy has charged Statistics Belgium to further develop all activities in this field.

This research project is oriented towards methodological and intrinsic progress in several related fields. Moreover, preliminary research in this field is done since two years. The long tradition of census-based socio-economic statistics in Belgium, the ongoing efforts in this field and the importance attached to it by the competent minister guarantee the integration of the final product into the activity of the FPS Economy.

Moreover, Statistics Belgium counts on a positive interaction of this project with major sample surveys and places it in a broader vision of the development of the Belgian statistical system. This system ensures that Belgian socio-economic statistics are made up longitudinally and reported to the European statistical institutions and monitors their quality.

b. for the general public: access conditions and procedures for scientific users and the general public

The valorisation of the final results for scientific users and the general public is guaranteed on the long term, as the research project is the continuation of the census (or socio-economic survey). As the results of the Survey 2001 become available for dissemination and further research, the immense interest of the mass media and the research world for this unique source of information is once again confirmed.

As stated above, it is less sure if the results of the proposed pilot census using the administrative method will be equally valuable. This is however anticipated by letting the published final results depend from the quality of the data, in order to ensure that valuable and scientifically reliable final results will be available.

As for access conditions and procedures for different groups of users of census data, Statistics Belgium uses a well-tried method, with a central place for privacy protection. Data that could directly or by crossing lead to identification of individual situations cannot, under any circumstances, be passed on; on the other hand, information is disseminated at the highest level of aggregation still useful to the end user. Moreover, sample surveys are systematically being preferred to exhaustive surveys. Only the Privacy Commission can permit departures from this rule.

**Timing, planning of the activities and allocated budget**

a. timing of activities

1 September 2005 – 31 August 2008

Cf. attached document "Timing of partial targets". It should be stressed that this timing is dependent on data supply from third parties. Therefore, there should be some tolerance in the timing of partial targets.

b. planning of activities: phasing of the tasks

1. *for the research team*

Four partially overlapping phases; tentative timing due to dependence from third parties for data supply.

Phase 1: (September 2005 - June 2006)

- consultation with administrators of administrative databases: a series of meetings resulting in the establishment of fixed structures for data transaction in the framework of administrative censuses
- Formulating the precise data request from the administrators: determining variables and their specifications (moment, format, sampling method...), including the metadata required to exploit this information and to verify its validity and reliability.
- Mapping and integrating current and extant research in the field

Phase 2: (March 2006 - October 2006)

- processing of test files on aspects of quality: 2 aims:
  - (1) further specification and, if necessary, adjustment of the data request (selection of the best quality source, specification of technical aspects of the data format and/or harmonisation in function of the coupling with other sources)
  - (2) preparation for the final processing of administrative census files (Phase 3): preliminary research on making a cross-section of the administrative database and evaluation of problems; effects of the new observation method
- Subjects treated: all partial activities from the transaction of data as they are supplied up to the production of finished statistics, coupling information of different origins, securing confidentiality, testing and quality analysis (e.g. comprehensiveness, double counting, internal consistency, time index of measurement, etc.), fixing procedures for data cleaning.

Phase 3: (November 2006 - October 2007)

- administrative census files come in (ref. date 01.01.2006)
- exploitation of files in multiple phases (with logistic and technical support from Statistics Belgium and ICT)
- data cleaning, harmonisation of the anonymous identification code for coupling without identification, consistency and internal verification of the source
- coupling, consistency checks between sources, treatment of missing values, documenting variables and codes (meta-information)

- creation of finished, documented and exploitable census files
- drafting of report "Data quality of administrative census data"

Phase 4: (July 2007 - August 2008)

- production and description of census results 2006
- production of tables and crossed tables, maps, socio-economic indicators
- final scientific result: The Socio-Economic Situation in Belgium 2006 (to be reformulated in function of data quality)
- final procedural result: towards an administrative system of data collection and production of socio-economic statistics in Belgium.

## 2. for the public institution

Phase 1: (September 2005 - June 2006)

- (renewed) contact with administrators of administrative databases to prepare the consultation about the launch of data transaction for administrative censuses
- designation of persons in charge of different partial targets and of the follow-up of data exchange
- consultation phase: delegating persons in charge from Statistics Belgium and ICT to meetings, staffing fixed structures of data transaction in the framework of administrative censuses
- specification of data request from administrators: technical specifications (format,...) and practical aspects of the transaction

Phase 2: (March 2006 - October 2006)

- implementation and development of the statistical database, further development of the DEWOLO project: hardware, intrinsic database structure, data models, queries, objects and their interrelations (geographical units, housing, households, family nuclei,...)
- introduction of files into the database (administrative test files, census information, database of addresses and their records) and exploitation (sampling and coupling for research purposes by the research team, procedures for updating information within the database)

Phase 3: (November 2006 - October 2007)

- administrative census files come in (ref. Date 01.01.2006)
- creation of a file "census 2006" within the database
- allocation of targets for data cleaning: design and testing on samples of a procedure of data cleaning by the research team, implementation of the procedure for the whole population and feedback by Statistics Belgium/ICT: data cleaning, harmonisation of the anonymous identification code for coupling without identification, consistency and internal verification of the source
- Exploitation of census files 2006

Phase 4: (July 2007 - August 2008)

- exploitation of database: further dissemination of anonymous results (if quality is guaranteed)
- longitudinal follow-up: updating of administrative information, introduction of new sources
- publication and dissemination of the final scientific result: "The Socio-Economic Situation in Belgium 2006" (to be reformulated in function of data quality)
- Publication of the final procedural result: "Towards an administrative system of data collection and production of socio-economic statistics in Belgium".

**Particular conditions**

a. suggestion for the composition of the steering committee

- representatives of Federal Science Policy, Statistics Belgium and ICT;
- representatives of the Privacy Commission and of all public institutions in charge of the required administrative registers, such as:
  - National Register of Natural Persons;
  - Crossroads Bank for Social Security
  - Land Register
  - regional and community institutions in charge of education.

b. Data Confidentiality Clause

barring previous amendment of the Statistics Act, the research team will be liable to sign a Confidentiality Contract before anonymous individual data are supplied. On the other hand, the research team will have to comply with conditions stated by the institutions supplying data or by the supervising committee.

c. presence of the research team within the institution (if applicable)

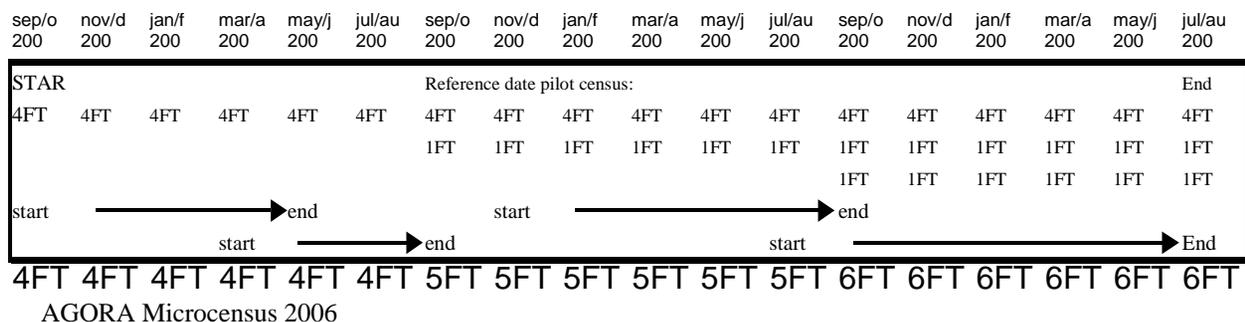
the presence of the research team within the institution is not required. However, team members and especially the person in charge of the co-ordination of all activities and of the liaison with Statistics Belgium and ICT, should permanently be in contact with the institution, take part in consultation meetings (if invited) and comply with the conditions stated by the supervising committee.

d. Other

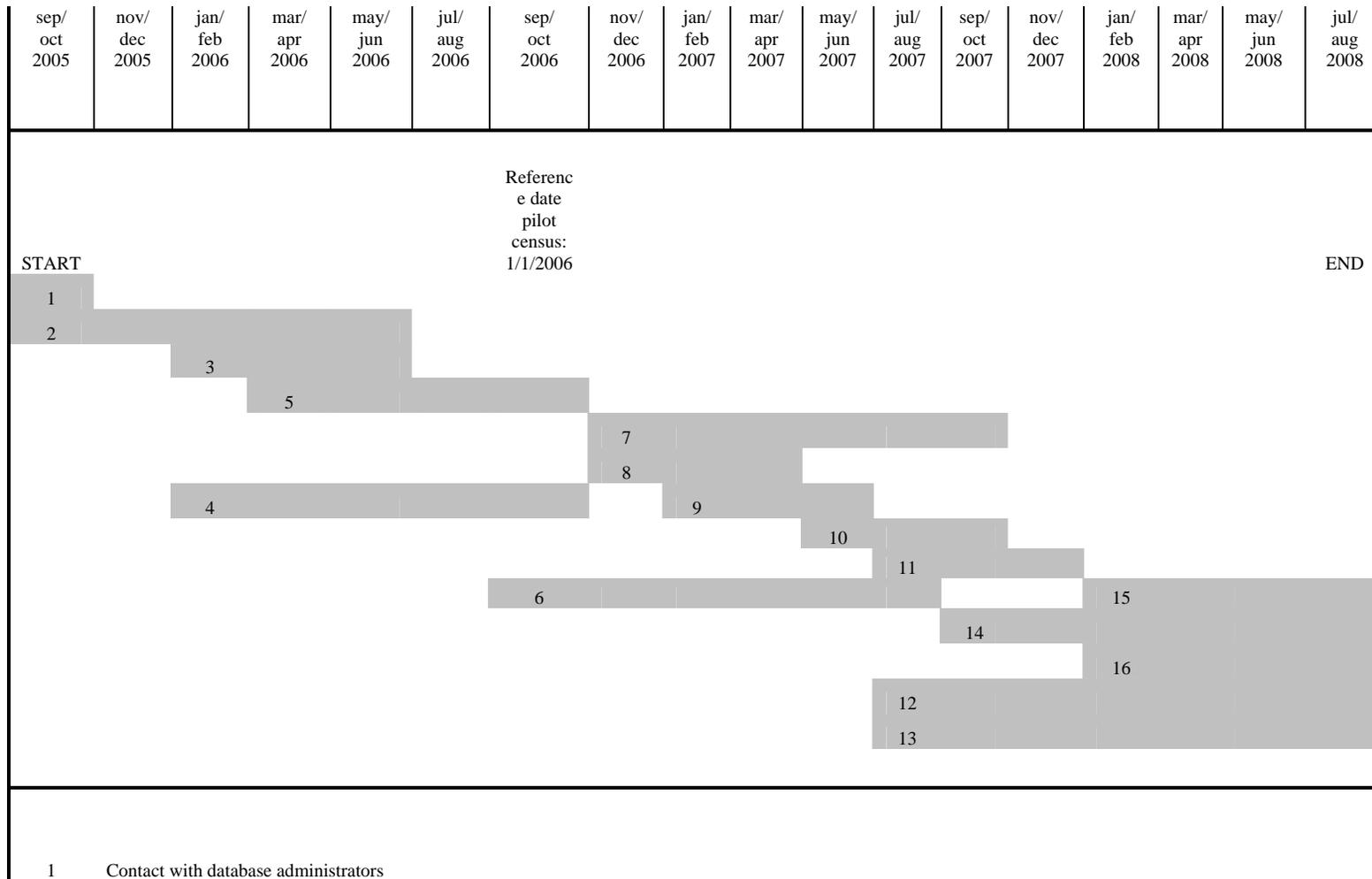
Statistics Belgium and ICT will supply 5 FTE of qualified manpower for the development of databases in order to exploit and update census information. A new internal project request was submitted to this end. In addition, the needs for scientific manpower to execute research activities within the research team is estimated at 5 FTE over 3 years, gradually increasing as the project unfolds. The allocation of target packages to the staff of the research team(s) could be done as follows:

- First package: scientific and organisational co-ordination, liaison with Statistics Belgium and ICT, scientific and administrative follow-up and reporting of the entire project. Staff: 1 FTE for the complete duration of the project.
- Second package: 1 FTE co-ordinates the aspects database, administration and exploitation of data, updating, coupling, integration of partial information and other co-ordinating technical, statistical and methodological aspects. Staff: 1 FTE for the complete duration of the project.
- 2 FTE (first year) to 3 FTE (later) for thematic activities concentrating on specific sources and/or fields of socio-economic information (demography and households, education, employment, health, housing, mobility,...). It is necessary that researchers in every partial field are involved from the beginning. However, research teams will be increasingly staffed as soon as test samples or final samples are available (from the second year on).
- 1 extra FTE for the third year. Targets: finishing and reporting of research, consolidation of the database, dissemination and documentation of the results and transfer of knowledge from the research team towards Statistics Belgium and ICT.

The following time graph "Manpower" gives a schematic overview of allocation and timing.



The time graph of the project is summarised below:



- Consultation period administrative databases:
- 2 Phase 1  
Specification of data
  - 3 request  
Statistics Belgium / ICT: database, interrelations and structures, injection of test
  - 4 files  
Processing test files: Phase
  - 5 2  
Statistics Belgium / ICT: introduction of administrative files, intensive collaboration with research
  - 6 team  
Exploitation of administrative census files: Phase
  - 7 3
  - 8 Transfer of census files ref 1/1/2006  
Preliminary treatments: cleaning,
  - 9 coupling  
Evaluation of data quality and
  - 10 reporting
  - 11 Final census files, production of tables
  - 12 Final report: Administrative Census: logbook, data quality and recommendations  
Final report: The Socio-Economic Situation 2006, format according to data
  - 13 quality  
Statistics Belgium / ICT: exploitation of stat. database admin. census
  - 14 2006
  - 15 Finishing and final reporting: Phase 3
  - 16 Statistics Belgium: publication + dissemination of results