# Wholistic Approach to Multifaceted Integrated Census

**Olivia Blum**

Central Bureau of Statistics, Jerusalem, ISRAEL
E-mail: oliviab@census.cbs.gov.il

**Abstract**

During the last 20 years, new census philosophy is implemented via the integration of data sources, the integration of different modes, methods and tools to collect and compile census information, and the deployment of census activities over time. The planning and implementation of the deriving new censuses necessitate a wholistic approach, which aims to answer some of the challenges presented: It encompasses and organizes the diversified census modules along central elements, and it entails integration-infrastructures that exist independently or can be built as an autonomous playground for the census game. The perception of a census as a whole refers to three basic elements: time, area and data sources and scope, whose bonding glue is the integration infrastructures. The suggested infrastructure systems are: information, spatial information and management and monitoring systems. Those elements and systems are expected to remain relevant when development in statistical methodology and census technology will further change and shape future censuses.

*Key Words: Census Philosophy, Integrated Census, Census elements, Integration-Infrastructures.*

## 1 Introduction

Traditional processes of census taking, no longer serve as an exclusive source of census information. Generating detailed population data develops into a costly project with an intrinsic complexity, causing the accuracy of the results to deteriorate. Consequently, statistic bureaus have accelerated the search for accepted functional solutions, during the last two decades. The use of registers and administrative records for different uses and in various phases of a census is already an integral part of conducted or planned censuses in many countries. For example, Finland and Denmark make use of and maintain registers and administrative records for censuses and ongoing statistics (Denmarks Statistik, 1995; Harala, 1996; Spieker, 1999). Norway and Sweden have performed their last integrated census that uses administrative files, for individual statistics, and fieldwork operation for household statistics (Thomsen, Vassenden and Laberg, 1996; Utne, 2003). Their next census is planned to be a pure administrative one. Israel and Slovenia have executed their last conventional census that uses registers for ensuring population coverage and for editing and imputation of the census file (Blum, 1997; Dolenc, 2003). Both plan their next census to be an integrated one, where administrative records are fully exhausted and field sample surveys supplement and correct it. USA and Switzerland lean heavily on administrative data because of the need to mail out census questionnaires (Buscher and Haug, 2003). Further developments that are contemplated lately relate to *the philosophy of censuses*; the idea of changing boundaries of content and time, that emerges in countries like France and the US (Griffin, 2003; Desplanques, 2003), puts in question some of the classic census principles. The tolerance to different time and population frames, as presented by the rolling samples and the separation between the geo-demographic census and the socio-economic survey, means that core principles of a census as well as principles that have been evolved as secondary goals and steamed mainly from operational considerations, are reconsidered.

The overall tendency is to recruit statistical logic and tools in order to achieve censal goals: *The characterization of small population groups of individuals and households, in small geographic areas, in optimal time intervals.* The ultimate goal of a population census has not been changed. The census is the main source of detailed information to be used as an input for decision-making and policy outline. However, census mechanism is modified constantly. Central characteristics of a census, like a snapshot in Census Day and an Actual Count of each individual in the population, are followed by a question rather than an exclamation mark. The use of 'optimal time intervals', rather than 'dictated dates', in the above definition of census goals is a result of users needs and suppliers ability to supply data, under changing and moving

constrains. Time reference is altered to a moving statistical average around predefined and rather frequent dates. Moreover, better accessibility to data in a world of developing statistical methodology enables the flexibility of associations between data collected and census results, therefore, the representation of all individuals in census file is changed to their inclusion in it (Blum, 2000).

The new census philosophy is implemented via the integration of data sources, the integration of different modes, methods and tools to collect and compile census information, and the deployment of census activities over time. This integration apparatus presents challenges of defining and identifying boundaries and frames of the census project and consequently, challenges to the management and control of census scope and essence. Furthermore, because of the fusion of different data-worlds, which are neither mutually exclusive nor consistent, the frame of the target population remains obscure throughout the process and can actually be defined only a-posteriori. It denotes uncertainty as an integral part of these multifaceted contemporary censuses.

The suggested wholistic approach in this paper aims to answer some of the challenges presented by the new censuses. This approach is based on two main features: It encompasses and organizes the diversified census modules along central elements, and it entails integration-infrastructures that exist independently or can be built as an autonomous playground for the census game. Integration systems are planned to enable the continuous follow-up, management and control of the operation, and to be stable enough to tolerate expansion in the number of census elements and processes without loosing the internal integrity or endangering the end products.

## 2 Basic Elements in a Wholistic Approach

Census process is changing and challenges its planners in its operational and substantive aspects. Carrying out censuses is an increasingly convoluted process that modifies census essence through a change of the operational apparatus. For example, collecting data from administrative records, in addition to field generated data from the target population, is a modification of the data source that also compromises the content, timing and timeframe of a census. The new ideas behind the contemporary censuses go even further and rethink census philosophy by contemplating census essence as a target for a change in itself.

Census goals and content are a reflection of legal and normative institution, crystallized over time. The social contract between census bureaus and users has been established in a continuous interaction, led by the needs of the users. Honoring the social contract means that the statistics organization should supply the user with detailed census statistics suitable to the agreed upon uses. Flexibility is a critical feature in this agreement. It allows the census scope to be expanded from enumeration of people to their geo-demographic and socio-economic characterization. It allows the separation between the enumeration and the characterization of the population and it does not obstruct the use of evolving statistical methodology and technological innovations. Nevertheless, this flexibility may lead to a fragmented process of generating census information, which loses its feature as a complete project. The perception of a census as a whole refers to three basic elements whose bonding glue is the integration-infrastructures: *Time, Area and Data sources and scope*.

### *2.1 Time*

The time that the data refers to in a single census has several correlated dimensions: The time and timing of data collection, the time to which the data relate to and the frequency of census taking. Most census models, suggested lately, are on the continuum between ‘once a decade census project’ and a ‘breathing census information system’ based on continually updated registers (Blum, 2000). For example, the new censal statistics planned to be produced in the US and France suggest a continuous process, which produces information for the whole country in small area level, several times throughout a decade. In other countries, census information is produced at least once a decade for each small population group. It seems that most countries try to keep the decade timeframe, but some wish to produce the full census information as a compilation of partial information generated over time rather than in a single time point. This approach has its toll: As in an ongoing process, the reference time is different and changing for different population groups and geographic areas. As the time range of data collection increases, the accuracy of the estimates of national census is reduced. Changes of geo-demographic or socio-economic attributes, have a similar effect; the more dynamic the

population, the less accurate is a cross sectional analysis throughout a country. Yet, users of data which are collected simultaneously, like those of a small or medium geographic units, are indifferent to the global census operation and therefore, from their viewpoint, census goals are achieved. A wholistic approach vis-à-vis time means that census planners should keep the ability to supply comparable census information within the country and between countries. There are three cycles to be considered about census time:

a) The intra-national cycle of activities defined as a single census, but extends over few months or years;
b) The international one, a cycle that extends over an agreed upon time interval;
c) A possible cycle of integrated or administrative censuses that ends up with a conventional census once in several decades.

Time intervals between data collection operations within a single census, should be minimized but adjusted to local circumstances and effective limitations.
Census time-span should coincide with the international time reference at least once a decade.
A cycle of integrated or administrative censuses that ends up with a conventional one is less needed as statistical methodology is developed and improved.

### *2.2 Area*

Census area is usually the area within the borders of the nation state. As such, it has a territorial continuity, divided by natural and administrative partitions. Since census population can be defined only a-posteriori, the area can serve as an a-priori census frame (Blum&Calvo, 2001). However, census area and census population do not coincide perfectly. There are short time visitors residing in the census area but do not belong to the census population, while servicemen and diplomats abroad, who are included in the population, do not reside in the census area. The former does not present a problem when the area is a mediator boundary, their omission should be part of the main process of the census. As for the later, the area is an irrelevant boundary. Nevertheless, area can be considered as a basic element of a wholistic approach as long as the global village is not becoming a melting pot with regard to access to services and goods, rendered in the international arena. Regular residence in more than one country is not a frequent phenomenon. Most people have a geographic base in which they get most services and goods. A wholistic approach vis-à-vis area means that the relation between area and population is two-sided: On the one hand, covering an area means covering its population, since area is the carrier of the target population. On the other hand, since census goals refer to information in small geographic units, all census entities and their attributes should have a geo-reference in order to be anchored and spatially analyzed (Blum&Calvo, 2001). It means that the integration of different data sources is facilitated when area is the common denominator.

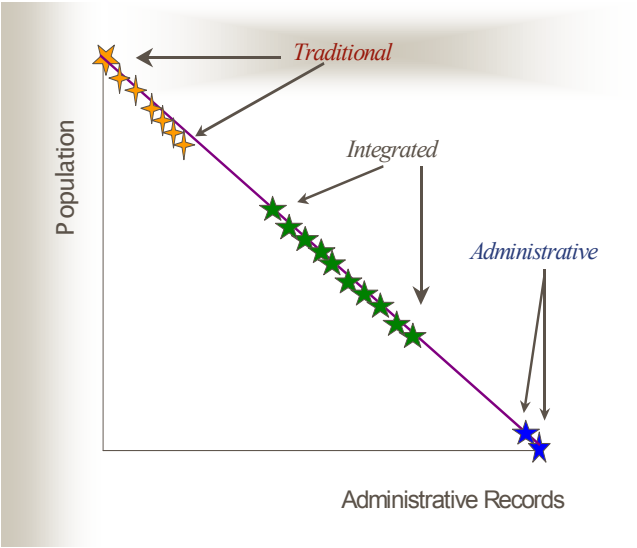
### *2.3 Data Sources and the Scope of Data Collection*

Census processes address the questions of size and characteristics of the target population. Census methodology, which put forward these processes, is developed and improved over the years. The contemporary censuses are led by two correlated elements that dictate and are dictated by the statistical method: D*ata sources and the scope of data collection*.

The first censuses in the 18th century that have used the parishes' notebooks are actually administrative censuses. Most censuses of the 20th century collect the information directly from the population and are referred to as traditional or conventional censuses. At the beginning of the 21st century, censuses lean heavily on the population as a source of information, but use registers and administrative records as a substitute, enhancement or as an improvement medium, for the information generated from the population. Registers that are maintained and sometimes built, by the statistic bureaus, should be defined as a direct and ongoing, rather than an indirect, administrative data collection. If the data source for this type of registers is the population and the data collector is the statistic bureau that uses other administrative records for census purposes as well, the out going census is an integrated one rather than a pure administrative census. An illustration of census type according to data source is presented in the diagram below.
Data sources are on both axes and they are either the population or the administrative records. Under the assumption of linear relation between data sources, the stars along the line represent different combinations of them, i.e. different types of censuses:
Both extremes denote the dichotomy of traditional census verses a pure administrative one.

A shift from traditional census to an integrated one goes through several types of *traditional* censuses, which use administrative records to support the traditional apparatus. However, an administrative census is changing its type to an integrated one, once it uses the population as a source of information. The difference between the two is a result of the fact that an interaction with the population is never minute. When data is collected from the population, the census is either integrated or traditional. When data is collected from administrative files, the census is either administrative, integrated or a traditional one.

**Diagram 1: Census Type by Data Source**




Regardless of the information source, the end file should represent the target population and supply a picture of the whole population. As mentioned above, this picture is not necessarily a snapshot, but rather a representation of the population in an accepted time range. It is, inter alia, a result of the selection of unsynchronized data sources, but it also influences the selection itself. The new flexibility allowed, with regard to a representation of reality in a time range instead of in a point of time, allows for data sources that are not updated regularly, to be used for census purposes.
As for the second element, the scope of data collection, it is either a full enumeration or a sample survey. Historically, the geo-demographic part of the census is obtained by means of a full enumeration, while the socio-economic part is based on a sample survey. Nowadays, both are compromised. Population size may be estimated through rolling samples while registers of socio-economic characteristics, like education or employment, may cover the whole relevant population. Although the law enforces full enumeration in many countries, statistical methodology allows for surveys to serve census goals. A wholistic approach vis-à-vis the selection of information sources and the scope of data collection place them as the cornerstones of most methodological aspects of a census, from the substantial statistical-methodology needed to compile the estimates to the mere tools selected for data collection.

The different combinations of various attributes of time, area and data source and scope are the trunks from which the branches of most census types grow. As such, they can serve as axes of the integration-infrastructure systems, needed in complex census apparatus.

## 3 The Integration-Infrastructures

Modern censuses are multifaceted in their basic attributes: Census day is a blurred concept when data sources are updated according to different procedures and customs of measurement, report and recording procedures. Start and end points of a single census operation are a function of definition and are extended over a few months or a few years span. The information collected is fragmented along content or geographic area lines; the demographic attributes of the population

are collected apart from the socio-economic ones and not all localities are included simultaneously. Consequently, management and control require integration infrastructures. They are needed to support the flow of updated process information and for rational decision-making throughout the census process, from draft planning to the dissemination of the last product. Monitoring and adjustment of census plans, within a single census cycle, is stipulated not only by the information flow to the managers and decision makers but also by their ability to actively manage the process in return. For that matter, three integration infrastructures are suggested: *Information, spatial information and management and monitoring systems*.

### *3.1 Information System*

Information items in a census apparatus are of four types, nested in each other: *Data* - presented mainly by alphanumeric values of variables within a file or by data-files. They are raw, as collected, measured, documented or processed. *Metadata* - refers to the information needed in order to use the data. It includes definitions, codebooks, path pointers etc. *Method* – comprises of census static components and dynamic processes, such as questionnaires, data collection methods and tools, editing and imputation, estimation and so forth. *Meta-information* – would be the ideology of the census, its structure and logic.

What to document and what to include in the information system do not usually receive identical answers in a short-term operation. However, a complex, multi-item and long-term census process does not allow for information to be excluded from a computerized system. It is needed for the mere operation of census processes and as a means of conveying information and knowledge to consecutive censuses. Yet, a distinction can be made between census core information that requires a ramified set of links to enable smart navigation and to stipulate or launch processes, selective data, metadata and method items, and information that can be stored as autistic items to be retrieved directly, including, among others, all meta-information items. The literature with regard to metadata and related documentary information systems is quite extensive and statistic bureaus report gradual implementation despite facing obstacles (Highsmith, 2002; Oakley, 2002). The issue remains on their agenda because of the growing and urging need, owning to the ongoing technological improvements, which facilitate the storage, retrieval and processing functions of information management**.**

The new integrated censuses emphasize or add unique features to the census core information. These censuses main function is the integration of all parts of a single census and they refer to the basic elements listed in section 2. Time, area and data source and scope gain dynamic traits, translated to a set of values and a set of new interactions between them. The matrix of 'time-area-data source-data collection scope' has many potential values and therefore, produces a much required information and meta-information. Moreover, a specific combination represents a stage in a census and therefore connected to a specific set of processes, which are the input or output of the interfacing stages. The following table is an example for combinations of values of basic elements in different stages of the same census.

| Element / Stage | *Time* | *Area* | *Data Source* | *Data Scope* |
|---|---|---|---|---|
| *Administrative Demographic Census* | 1.2004 | Census area: All localities | CPR<br>Local authorities | Census population |
| *Coverage Evaluation* | 2-3.2004 | Selected Localities | Population,<br>Demographic- Estimates | Sample of census area cells |
| *Socio-Economic Field Survey* | 2-3.2004 | All Localities | The population, supplemented by administrative sources | Sample of census population |
| *Quality Evaluation* | 9-12.2004 | Selected Localities | LFS | Sample of census Households |

Each stage could have been defined as a single project. However, since all stages are part of one census project, the information system supplies the project frame and serves as an integration-infrastructure.

### *3.2 Spatial-Information System*

A spatial information system has static information layers of physical elements, such as buildings, streets and geographic areas, as well as additional capabilities of dynamic analysis and the production of spatial information. It is perceived as one of the most important technologies for integration of information through synthesizing and organizing functions (Dangermond, 1999; Schaefer, 1999). More so for censuses whose goals are defined in terms of area and whose implementation relies on spatial analysis; Area is the carrier of all census information and therefore, all census entities should have geo-references, which are anchored to physical elements or to the virtual grid of xy-coordinates. Furthermore, in the new censuses geo-reference becomes a necessary condition for implementation, especially those integrating data from different sources and those conducting series of surveys covering the census area over an extended period of time.

The spatial system as integration infrastructure supports and enables main census processes through the following (Blum&Calvo, 2001):
*Linking information* related to the same entity (individual, household) and ascribing it its attributes not only by aggregating data from one level to the other but also by collecting or generating data on a macro-level spatial unit. At times it entails the creation of ad-hoc complex entities when the information to be linked does not share a known spatial entity. For example, individual data, with an apartment as a geo-reference, can be fuzzy linked to individual data anchored to the building or to an electric pole outside the building. When the individual level information has different geo-reference in different sources, spatial analysis support the decision what information relates to the same person.

*Defining census frame and sampling units* are another aptitude of the integration infrastructure. As mentioned above, the census area is predefined and known beforehand. It is divided for sampling purposes according to expected population size and expected number of households. The working and analysis units of a census, within the geographic boundaries of the area, are sampling area-cells and enumeration area-cells (sampling area-cells that were sampled). Both types are spatially created and are based on the updating of the buildings and addresses layers in the geographic information system.

*Ensuring coverage* by spatial analysis and follow up of the progress of census processes are capabilities of a dynamic system, which integrates spatial information. In integrated censuses, individual records from different administrative files can be anchored to a spatial entity and a visual picture can reveal spatial patterns of coverage of these files. It is to be considered when defining the samples for the field surveys and may lead to over-representations of areas with visible under-coverage. Ensuring coverage by spatial analysis is also needed for survey follow-up when a census is based on rolling samples, especially when estimates are required on an aggregate level, across the geographic divisions of the survey. Many other census processes that can be characterized in spatial terms are supported by this integration infrastructure system: Imputation, enumeration follow-up, geographic distribution of enumerators' recruitment etc.

As long as the census population is defined in terms of area, spatial information system is a most needed census infrastructure, which becomes critical in the contemporary censuses since the basic elements of area and data collection scope are spatial derivatives.

### *3.3 Management and Monitoring System*

Management and monitoring system of a census incorporates static operation rules and dynamic ignition of processes like checks, preparation of job-packets and generation of reports. It has always been a complex mission of commanding a huge operation, performed in a relatively short period of time, and described in terms of mass production in a blitz operation. With the new combinations of time, area and data in censuses, additional components are introduced and the ability to see the whole census picture, at any point of time, is developed into a methodological and technological challenge.

The input for the management and monitoring system comes from the other integration-infrastructure systems and its output is feeding them back. Meaning that the census tripod, which is comprised of three integration infrastructures, is tightly connected at its base. This is the very same feature needed for and enables the implementation of wholistic approach in census planning and conducting.

Managing each sub-system of a census, on different managerial levels in the census hierarchy, implies the following main processes: The creation of a selective managing-information item as a result of the completion of a necessary condition;

the automatic activation of a subsequent process; or displaying the item and allowing for manual activation in order to continue or to change a process. Defining the managing-information items defines all inputs-processes-outputs of a census system. Those items that reach the human-machine interfaces should be located in the critical and risky points along the census process.
As for monitoring, it requires the passive acknowledgment only. A monitoring-information item is created and displayed but no manual interference is allowed. Monitoring is required for an ongoing follow-up, yet, corrective action, if needed, is implemented out of the computerized system, in a face-to-face interaction.

## 4 Concluding Remarks

A census is a mean to produce detailed and comparable statistical information, needed for policy and decision-making. Contemporary censuses enable the production to be more efficient and frequent, yet comparability is hampered. Conditions and stipulations are introduced because of the changing perception of time, area to be covered and source and scope of data to be used. Further developments in statistical methodology, in areas such as sampling, small area estimates and time series analysis, will allow for additional changes of these basic census elements. Related technological developments in areas of data collection tools, accessibility to databases, storing, retrieving and analyzing large volumes of data etc. will also influence the selection of data source and the sophistication of data use.

The perception of a census as a whole stipulates its success to achieve census goals, however, it cannot be done with yesterday's tools. A wholistic approach, that put integration infrastructure systems at the base of each census, is a possible way to face new and growing challenges.

## Bibliography


Blum, Olivia. Israel Central Bureau of Statistics, 1997.
"Evaluation of Data Editing Using Administrative Records". WP 16 presented in the UNECE Work Session on Statistical Data Editing. Prague.

Blum, Olivia. Israel Central Bureau of Statistics, 2000
"Combining Register-Based and Traditional Census Processes as a Pre-defined Strategy in Census Planning". Statistical Policy: Working paper 30. 1999 FCSM Research Conference.

Blum, Olivia and Rinat Calvo. Israel Central Bureau of Statistics, 2001
"Geospatial Data Collection and Analysis as Crucial Processes in an Integrated Census". Paper presented in the 1999 FCSM Research Conference.

Buscher, Marco and Werner Haug. Swiss Federal Statistical Office, 2003.
"The Population Census 2000 in Switzerland". WP 2 presented in a joint ECE-EUROSTAT Work Session on Population and Housing Censuses. Ohrid.

Dangermond, Jack 1999 "GIS in the Next Millennium" in GIS 2000: the next millennium. ESRI.

Denmarks, Statistik 1995
Statistics on Persons in Denmark: a Register-based Statistical System. Eurostat and Denmark Statistik. Luxembourg.

Desplanques, Guy, INSEE, 2003
"The New French Population Census". WP 12 presented in a joint ECE-EUROSTAT Work Session on Population and Housing Censuses. Ohrid.

Dolenc, Danilo, The Statistical Office of the Republic of Slovenia, 2003.

"Register-Based 2002 Census of Population, Households and Housing in Slovenia and New Solutions in Data Processing". WP 4 presented in a joint ECE-EUROSTAT Work Session on Population and Housing Censuses. Ohrid.

Griffin, Deborah H., US Census Bureau, 2003.
"The Role of the American Community Survey in Re-Engineering the US Census of Population and Housing". WP 8 presented in a joint ECE-EUROSTAT Work Session on Population and Housing Censuses. Ohrid.

Harala, Riitta Statistics Finland. 1996
"Continuous quality assessment of the register based census and regional employment statistics". Working paper No. 6. SCECE Work Session on registers and Administrative Records in Social and Demographic Statistics. Geneva.

Highsmith, Samuel, US Census Bureau, 2002
"Building a Metadata Repository to Support the 2002 Economic Census". WP 4 presented in a joint ECE-EUROSTAT Work Session on Statistical Metadata. Luxembourg.

Oakley, Graeme. Australian Bureau of Statistics, 2002.
"New Strategies for Metadata in the Australian Bureau of Statistics". WP 5 presented in a joint ECE-EUROSTAT Work Session on Statistical Metadata. Luxembourg.

Schaefer, Mark 1999 "GIS and Emerging Jeffersonian Technologies" in GIS 20: the next millennium. ESRI

Spieker, Finn Statistics Denmark 1999
"Formation of central variables in a decentralized statistical system". Working Paper No. 25. Joint ECE/Eurostat Work Session on registers and Administrative Records in Social and Demographic Statistics. Geneva.

Thomsen, Ib, Elisabetta Vassenden and Britt Laberg Statistics Norway. 1996
"Availability and use of administrative record systems in the ECE region". Working paper No. 7. SCECE Work Session on registers and Administrative Records in Social and Demographic Statistics. Geneva.

Utne, Harald. Statistics Norway, 2003.
"Future Censuses in Norway – A Register-Based Approach". WP 7 presented in a joint ECE-EUROSTAT Work Session on Population and Housing Censuses. Ohrid.