The Geospatial Perception and Its Impact on The Content and Processes of a Multi-Source Data Collection

Olivia Blum

Israel Central Bureau of Statistics
blum@cbs.gov.il

1 Introduction

The introduction of the geospatial perception to the statistical information world has stimulated the imagination of people as of what information can be produced and has enhanced the abilities to do so. It has enabled the integration of different information sources, the specification of complex spatial models describing population characteristics and behavior, the analysis of volumes, areas and distances in space, over-time etc. The derived statistical system is comprised of a geospatial infrastructure carrying heterogeneous information by means of spatial entities. Their quality and integrity are pre-requisites for the geospatial processes to be part of the statistical production.

The geospatial perception has been recruited and developed as part of the census methodology in Israel. The qualitative geospatial infrastructure has been the enabler of the transition from a traditional census to an integrated one. Different tasks of data processing, like data collection, integration and correction are embedded in the infrastructure, and therefore performed as geospatial processes. Moreover, a strategic plan has been developed to exploit the advantages presented by the system, so that field surveys will be effective and parsimonious.

This paper is comprised of two main parts, the geospatial perception and its implementation in data collection processes in the integrated census in Israel. In addition, an example of a special challenge that such systems generate is presented, and it is the identification and localization of geospatial entities in localities without addresses.

2 The Geospatial Perception in the Statistical Production

Events of statistical interest happen in space, over time. Space matters since the interactions between data items in space matter (Calder, 2000). However, until recently, measuring those events in the statistical production process, has neglected its multi-dimensionality, mainly because of limited supporting technology. The accelerated technological development of geographic information systems has first introduced the geospatial dimension to the military and space industries. Its diffusion to the civilian industry has enabled statisticians to have a better insight and understanding of the studied reality and to explore and exploit the new opportunities presented. The statistical production needs have been answered by solutions supplied by the improved technology, yet the beneficial process has further evolved. The new means have instigated new needs in the spatial realm, and have altered the perception of the measured reality, its significant factors and the links between them. In a way, thinking geo-spatially is a revolution in which basic questions, accompanying the statistical operation, like what to collect, how and to what end, are expected to be followed by rethought-of answers. The geospatial perception that includes time as a fourth dimension, and the corresponding capabilities, imply a dynamic system in which processes of progression are its core nature, and the statistical action is conducted to fight ‘emergencies’ and to exploit ‘opportunities’ (Backer et al, 2001). As such, the geospatial foundation is a potential key to unsolved problems, to better solutions to already solved ones and to emerging challenges.

2.1 Engaging the geospatial perception in the habitual production process

Many processes in the statistical production are changed and most of them are improved, once the geospatial perception and the pertinent technological tools are introduced. They provide the vastly sought for, common denominator, needed for the *integration* of heterogeneous information-sources and data types (Blum&Calvo, 2001; Calder, 2000). The harmonization needed for the integration is embedded in the geospatial infrastructure, using geo-references as a pivot for each data item. This harmonization process is applicable to information sources that could not be used before; each data source carrying geospatial information can be linked via the infrastructure to the existing data. Consequently, unconventional data sources and data types are introduced to the statistical arena.

Immediate benefit of the integrated geospatial infrastructure is the improvement of *data processing* tasks like record linkage, address coding, editing and imputation. The first two are mostly a side product of the integration; they are an integral part of the infrastructure. Editing and imputation are enhanced by the graphic presentation and by the spatial context of data-analysis. The visual presentation simplifies the identification of missing or erroneous data, and spatial analysis improves imputation and hence, data quality. For example, nearest neighbor imputation in space rather than in a two-dimension environment, can be more accurate if the proximity is across information layers and additional data can be included in the model. An empirical example for a better-specified imputation model is the use of snow, air and weather data to improve the nearest neighbor imputation model to forecast snow avalanche across Europe (Brabec, 2000).

Another closely related infrastructure process of a geospatial nature is a *territorial division*. Area cells serve as sampling and estimation units. The traditional territorial division takes into account demographic and socio-economic attributes of the population in neighborhoods. However, since the basic building block is a unit of administrative nature, homogeneity within area cells is not optimal from the very beginning. Moreover, in the long run, the original homogeneity is further deteriorated. The introduction of spatial relations and the assumption of implicit social relations in them are used to solve the problem of heterogeneity (Grasland,2000). The boundaries of area-cells do not have to coincide administrative or natural outlines, and can be a model-based product. The spatially homogeneous area-cells serve the design of an efficient and sophisticated *sampling of* individual records or area-cells, by stratifying statistical units used for census and surveys purposes (Blum&Calvo,2001; Crescenzi,1998; Gould&Hecht,2001; Insard,2001).

Territorial divisions also serve the organization of the fieldwork operations. Ordering the sampled units in an efficient coverage route reduces futile travels between them. This advantage is less protruding when the sampled unit is an area cell. The model-based boundaries of the area cell can be difficult to identify. However, this challenge is met by the improved *spatial orientation*, facilitated by the use of navigation tools, like the GPS.

The geospatial infrastructure and the linked information improve the quality and diversity of the *statistical products*. The dataset is richer and therefore opens new options for qualitative map production, modeling and analysis. For example, the static characterization of the workplace is enriched by the possibility to link the attributes of its individual workers, the air pollution data and taxation policy.

In all above functions, the statistical production process takes advantage of the technology to simplify and improve solutions for well-known questions, and to some extent, to alter the perception of a problem, so that the search for the solution will take a different avenue. These capabilities and perception have been a necessary condition for the planning and implementation of new types of a census in several countries, during the last decade, among which are France, the United States and Israel.

2.2 Instigating new needs and facing new challenges

The emerging challenges are of two types; system related requirements and newly created statistical needs:

The two pre-requisites of the geospatial statistics *system* are the completeness of its geo-infrastructure, and the data integrity as presented by the entire set of linked geo-referenced data-items. The geospatial infrastructure has to be built and updated continuously, in a register-based approach; all the population has to be covered and all events have to be recorded as close as possible to their occurrence. The available data sources usually include geospatial entities, however, they are incomplete and their updating is not frequent enough for register needs. Therefore, designated data collection is conducted, using advanced technology like digitized satellite images and remote sensing. As for *data integrity*, the main challenge is to find possible and functional geo-references rather than the most accurate ones. Attaching x-y coordinates to data used in aggregates, which are defined as geospatial polygons, can be unproductive, if it is not an automatic, straightforward action. Yet, the flexibility of drawing spatial patterns, using point-based references, and the need to locate the position of statistical micro-data or small territorial areas, impel statistics bureau to adopt the point-based approach (Nael&Nordstrom,1998). The geospatial system, which is the result of data collected by remote sensing integrated with static maps and with alphanumeric geo-referenced data, comprise the new geospatial database (Crescenzi, 1998).

The newly created statistical needs stem, as mentioned above, from the new perception of the measured phenomenon and the exposure to the technical and technological abilities of the geospatial system. The answers to the questions asked engage the unique characteristics of this system; geospatial analysis and multi-dimensional graphic presentation. The rising needs are expressed in *spatial patterns of volume, area and distance, measured over-time*. The exposure of a population in a defined area, to air pollution generated by a refinery is measured as a phenomenon in space, where variables like the topography of the area, the atmospheric conditions, the relative altitude of the houses, the distance from the refinery are relevant to the analysis. Travel data of Moslem women in the U.S. after 9-11, have been graphically presented as a result of spatial analysis, taking into account the interaction between time and space (Ding&Kwan, 2004). The well being of the population is widely measured, as a phenomenon in space, across time.

The facilitation and simplification of defining boundaries of areas as boundaries of spatial patterns stimulate needs of *ad-hoc estimation units*. End users are sensitive to the expansion of possibilities, and they change their demands to better suit their needs. If a studied phenomenon, like the socio-economic status of the population in rehabilitated housing units, crosses administrative borders, it is more efficient to create an exhausting and exclusive area, where all the relevant houses are included and only them, rather than to stick to independent administrative boundaries. In this type of analysis, respecting administrative borders should not be expected (Backer et al, 2001). Creating ad-hoc spatial units better specifies the research population. An empirical example of that is the estimation of population density across Europe, using CORINE land cover inventory. It has been found that land-cover type improves considerably the estimates. Moreover, since the land type crosses administrative borders, their inclusion in the model hampers the estimates (Bjornsson,2001; Gallego,2001).

The new challenges and needs, presented in the spatial world, will expand the existing accumulated knowledge and will serve as a break through vehicle for new ideas and research. In order to create the conditions for the maximization of the benefits the geospatial discipline offers, several stipulations are to be followed. One of the main pre-requisites is a designated data collection.

3 Geospatial Data Collection in an Integrated Census

Data collection is use and tools driven. Data is collected to serve the direct needs of the users and the indirect needs of the enabling system, whether it is an organizational system or a substance directed one. The geospatial system, being a technological tool and a substance domain, inspires the data collection methodology, its processes and content. Moreover, sources of information are relevant only if they can be geo-referenced to spatial entities and if they carry the wanted information. Since they are heterogeneous in nature, data collection strategy is required (Nebert,2000) and has to be adapted ahead of time.

The actual and potential uses and the availability and accessibility of the relevant data sources induce the decision regarding the spatial data-collection unit:

If the estimated phenomenon is based on exact position in space, point-based approach is adopted and x-y coordinates are attached to each data-item. The major advantage is the expansion of the production potential. Pre-planning is not a pre-condition to produce a tailor-made spatial product. However, the implementation of this approach can be expensive and may introduce additional sources of error. Moreover, the basic process of information integration is led by the logic of information segregation defined by flexible geospatial patterns or by pre-defined units like grid-net. Hence, there is a need for useful geo-referenced information for disaggregation (Gallego,2001; Bjornsson;2001) and sources of information that answer this stipulation are limited.

A polygon-based approach in data collection is addressing needs with formal aggregated geographic units. It is easier to implement and more sources of information answer the basic requirements. However, users have to adapt the offered solutions to their needs, instead of vise versa.

Census information is used in both forms, micro-data and small area aggregates. Therefore, the selection of data sources should be guided by the ability to attach the smallest geo-reference to all meaningful census entities. It can be an indicator of a point in space or a centroid of a house polygon, or larger units like a city-block or an enumeration area-cell. In traditional censuses, data source is the population and the ability to attach x-y coordinates to the dwelling unit, depends on the existence of agreed-upon addresses. In administrative and integrated censuses, the meaning of addresses is extended to include spatial entities, not necessarily residential addresses, which carry relevant census information and can be linked to an estimation unit.

Strategic plan of data collection has to distinguish between processes by their objectives; system and tools related, or end-users related: Identifying and locating *spatially fixed elements* serve operations like sampling polygonal units (buildings, area-cells), field orientation and navigation, and infrastructure updating; Defining infrastructure indicators that trigger an updating operation has to be included in the geospatial system, since its relevance depends on its register-based approach. Identifying the spatial location of *spatially mobile elements,* like people and households are needed in direct and especially in remote data collection (mail, telephone etc.).

In censuses, finding and scanning spatially fixed area, and identifying all mobile elements geo-related to fixed elements in space, is measured in a given point of time, rather than over-time. This requirement emphasizes one of the weaknesses of the existing geospatial infrastructure; real-time updates are not an option and the system is updated in a lag of months and at times of years. Consequently, part of the geographic identifiers, collected in the field, are not recognized as fixes spatial elements and the mobile data-elements 'float' until a correction operation is undertaken. Another problem accentuated in censuses is the absence of recognized geographic reference related to the mobile data-elements, the census population.

The association between the data collection unit and the geospatial infrastructure is never perfect. The lack of formal addresses is a common problem, shared by many countries, in a non-urban area or in a non city-style address. Their spatial anchoring requires the intertwining of special processes, supported by designated tools, as exemplified in the integrated census in Israel.

3.1 The geospatial context of the 2008 census in Israel

The next Israeli census is basically an administrative one. It is based on the central population register (CPR) that has to be corrected for census purposes. Two sources of information are used in the correction process, administrative files and field surveys:

The administrative records support mainly the formation of the census-population frame, the correction of individual addresses in order to get the right geographic population distribution, and the marginal imputation of demographic variables. These corrections are made on the individual record level and therefore all sources of information have to be linked on this level. In the geospatial system, records are anchored to their dwelling unit, via the centroid of their building polygon. Since, not all individual records can be spatially anchored to this smallest fixed spatial entity, the attached geo-reference is the closest geographic aggregate, which is a street, an area-cell or even a locality. These linkages, although not refined, have an added value in the imputation process needed for small area estimation.

The attempt to assign a geo-reference to each census record contributes not only to the quality of the census file, but also to the completeness and quality of the geospatial system. Encountering an unknown geo-reference, whether it is a formal address or an informal neighborhood name, leads to the improvement and updating of the geographic database and its digital presentation. Saving the informal names and addresses, as synonyms, in the geospatial database, is very useful for data collection processes, when the population is the source of information. People tend to report the frequently used address even if it is not the formal one. Turning it to a known informal address supports the geospatial localization needed in the following surveys. The administrative records hold informal geo-references when the data collector does not insist on recording the formal address. The central population register in Israel, although tend to collect formal addresses, still have informal temporary addresses and old addresses that have not been changed since the initial reporting, even though new formal addresses have already been assigned.

Assigning geo-reference to the addresses in the different administrative data sources, using the geographic system and its geospatial analysis abilities, is a pre-requisite for the geospatial integration. It creates a unified information infrastructure for rational decision-making, in the process of building the geo-demographic census file (Blum, 1999; Blum&Calvo, 2001; Blum, 2003). The harmonization process allows for a comparison between common characteristics of the same population, coming from different information sources. In cases of disagreements between values a data selection process has to be engaged. In order to select the most accurate value, several quality criteria are built, based on the quality of the administrative data, evaluated by the following geospatial characteristics: The extent and accuracy of the possible geo-references, the resolution of the geo-reference, and the coverage completeness of the file with regard to the census area. The end product of the selection process is an integrated administrative file (IAF).

Once the administrative sources are exhausted, field surveys are engaged to allow for further correction of the IAF. However, while the administrative files correct the CPR data directly, through editing and imputation processes, field surveys supply the information needed for small area estimation, and their results are translated to weights assigned to each individual record in the IAF rather than to a changed population or value. The new product is the geo-demographic census file.

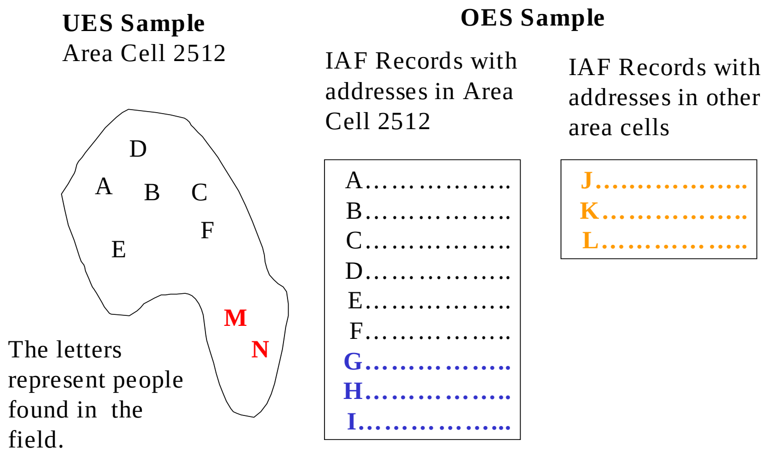
Two sample surveys are carried out, one to evaluate the undercoverage of the file, and the other to evaluate its overcoverage, on the local and global levels, using statistical methodology developed by Glickman and Nirel (Glickman et al, 2003; Nirel et al, 2004). These evaluation processes are geography dependent and therefore are stipulated by the ability to identify and locate fixed and mobile data elements in the *census space*.

The Undercoverage Evaluation Survey (**UES**) is based on area sampling. The census area is divided to area cells of about 50 households, with boundaries created by buffers around building polygons. These boundaries are not well-identified border-entities, like streets or other natural borders. Their definition is stipulated by the size of the CPR population, anchored to each polygon of a residential building, and by the boundaries of the statistical area. The sampling process itself is also a product of geospatial analysis of the tendency to be under or over-covered in the administrative files, the proximity between sampled area cells, the borders of statistical areas and the fieldwork organization requirements, related to the data collection routes and to the interviewers workloads. More area cells are sampled in localities that their CPR records are known to be wrong. Areas with taxation benefits are expected to be over-covered in the CPR records, while new residential areas are expected to be under-covered in the CPR records, since not all people have had the time to report their new address to the CPR. In both cases, more area cells are sampled, comparing with an area with a stable population, in order to provide a better population estimates. Operational considerations may also influence the sample, since a scattered and not well-organized sample has financial and quality costs.

The undercoverage estimation, enabled by the UES, is based on the comparison between the population found in the sampled area cells in the survey and the population in the IAF records that carry addresses anchored to the same building polygons as in the sampled area cell. The people, who have been found in the field but have no record in the administrative file, or have a record with an address in a different area, represent the local undercoverage. The estimation process and the quality of the results depend on the ability to assign the right geo-reference to all entities involved, on the quality of the geospatial infrastructure and on the ability of the interviewer to identify her spatial location. Failures are expressed by erroneous geographic location of people interviewed in the survey, or inability to assign them a geo-reference, which lead to estimation error of the undercoverage. One of the potential sources of error is the unusual borderline of the sampled area-cell. Since these borders can be a virtual line between houses, interviewers may cross them and interview people from other statistical areas. Since their addresses in the IAF are in a different statistical are, the result is an over estimation of the undercoverage. Failures are also expressed by erroneous geographic location of administrative records in the IAF that leads to undetected undercoverage of people missing from both, the administrative records of people in the investigated area, and from the corresponding survey records. Yet, the use of sophisticated spatial sampling for evaluation purposes, and engaging geospatial record-linkage in order to detect undercoverage in the administrative file, are a product of the adopted geospatial perception, and the quality of the outcome overrides the obstacles created along the implementation process.

The Overcoverage Evaluation Survey (**OES**) is also relying on geospatial capabilities. The sample is not of area but of individual records in the integrated administrative file. However, in order to save resources, this sample overlaps, fully or partially, with the records that carry addresses in the area-cells sampled for the UES. A full overlap between the two samples means that most of the sampled units in the OES, have already been found in the UES, and only those people whose administrative records carry a residential address that have been anchored in the sampled area cells, but have not been interviewed in the UES, are to be found in the OES. Those records are a potential overcoverage of the administrative file and the goal of the OES fieldwork operation is limited to the finding out if they are in the same statistical area in the field as in the administrative file, or not.

In the following illustration M and N are the undercoverage of the IAF, as found in the UES. G, H and I are a potential overcoverage of the IAF, since they have not been found in the UES, although their addresses in the IAF are within the sampled area cell. J,K and L are the part of the OES sample that does not overlap with the UES sample and together with G, H and I are to be looked for in the Overcoverage Evaluation Survey.



The integration of the UES data and the administrative data in the geospatial system is the enabler of this process, and its benefits are in quality and costs. A smaller number of sampled OES individuals to be searched and interviewed, implies a reduction in the potential non-sampling errors and in the costs of the fieldwork operation in the census.

The administrative record of the sampled individual in the OES has a geo-reference to a building in a sampled UES area-cell. The interviewer of the OES may find this individual in the same building and clear him from the suspicion of being overcoverage of the statistical area (He is the undercoverage of the UES fieldwork). Otherwise, the search for the sampled individual is done in an address, or by telephone, provided by proxies like family members and former neighbors, or by supporting administrative files. The decision if the individual resides in the investigated statistical area is based on the ability to assign a geo-reference to a reported address. It is a simple task if a face-to-face interaction occurs between the interviewer and the interviewee, and when the address supplied is in a different locality. It gets more and more complicated as the reporter is less reliable (not directly related) and as the quality of the address declines. Quality is evaluated by the resolution of the spatial reference that the address permits. Erroneous decisions vis-à-vis the location of the sampled individuals may lead to estimation error of the overcoverage.

As in the UES, the estimation process and the quality of the results of the OES depend on the ability to assign the right geo-reference to all entities involved and on the completeness and timeliness of the geospatial infrastructure. However, these are the enablers of the integrated census methodology described above. The statistical production process depends on them completely, and therefore, they may become its Achilles heel. Their quality has to be monitored regularly and resources are to be allocated to their constant improvement and update.

Another survey of the integrated census, not discussed in this paper, is the socioeconomic survey. It is integrated in the UES and serves also as the sole source of household statistics.

3.2 A special case: Localities without addresses

An example of an incomplete geographic infrastructure that calls for special data collection processes is the case of localities without formal addresses. In the administrative records, the geo-reference of the population is on the locality level only, and in the field, streets and building have no formal names or unique numbers. Small villages and Arab towns in Israel have missing, completely or partially, formal detailed addresses. In most localities, it does not have any effect on the census results since the whole locality is small and comprises only one estimation area. However, a detailed geo-reference is required in relatively large localities, which are divided to several statistical areas. This is the case in most Arab towns in Israel. The challenge presented is finding the way to attribute geo-references to the sampled individuals, in an estimation unit within the locality.

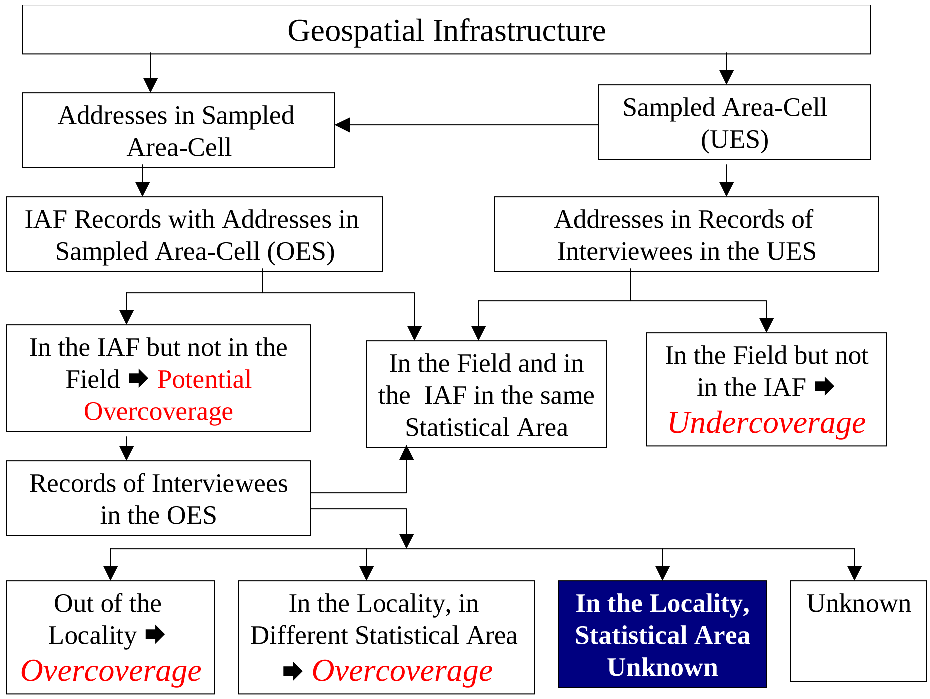
In the previous traditional censuses, the whole area has been scanned and streets and buildings have been labeled with synthetic numbers. Therefore, the whole population has been anchored geo-spatially and the boundaries of any estimation unit or any geographic aggregate have been drawn at will. This is not the case in the integrated census; the frame is supplied by administrative sources, and when there are no addresses, the population is separated from the detailed geographic infrastructure. The whole population in the integrated administrative file is anchored to one ‘cloud’, floating over the locality.

The data collection process in the UES is basically similar to the traditional census. Although only a sample of area cells is scanned, once a questionnaire is answered, the whole household is anchored to the building polygon that carries a unique synthetic address. The area and the buildings are fixed in space and can be identified in the geographic infrastructure. However, since the results of the UES are locally anchored (to the statistical area) but the matched records in the integrated administrative file are anchored to the locality level, an evaluation of the under-coverage is possible only on the locality level. This problem has been tackled by assigning *ad-hoc* geo-references to the administrative individual records, based on the previous traditional census. The logic behind this decision derives from the low migration rates of the Arab population within Israel, about 1% a year. The population is geographically stable. As for those who have not been found in the 1995 census, the geo-reference of their nuclear family is imputed for them (young children, female spouses). Once, most of the population has addresses, the process of evaluating the undercoverage is similar to the one in localities with addresses. Yet, the orientation in the field is tricky since streets and building in the field do not carry the synthetic numbers, assigned in the geo-spatial database.

As for the OES in localities without addresses, the challenge is presented once the information is generated from proxies. The identification of the spatial location of a sampled individual, in a place with no agreed-upon name, calls for a short run improvisation and a long run solution.

Those who are not found in the UES in their 1995 addresses, are looked for in the OES in three proximity circles: The nuclear family, the extended family, and the local authority. These proximity circles are relevant to the Arab towns because of several culture-dependent attributes: Living in an extended family pattern, low migration rates, and familiarity within the locality. A source of information for the family location, either nuclear or extended, is the UES, limited to the sampled area cells. The family that has been interviewed in the UES is questioned again in the OES, about the whereabouts of the sampled relative. The third circle consists of informants in the local authority, like the water-meter reader, who is potentially a reliable source of information, since he supplies the population with services in its physical geographic location.

The information gathered is sufficient if it provides the required geo-reference to the statistical area or locates the sampled individual totally out of the locality. Sufficient information is gathered in a direct interaction with the sampled individual or his residence. However, if the information gathered implies that the sampled individual lives within the locality but not where the interaction with the proxy has taken place, data collection processes have to be extended further. The following diagram presents the processes and products along the evaluation surveys in localities without addresses, up to this junction point:



When the interviewer is physically directed to the individual's house, the geo-reference is enabled, and the statistical area is identified. However, this solution is shortcoming when there is a large number of potential overcoverage.

A simulation of a physical guidance of the interviewer to the residence of the sampled individual is navigation on the locality's map. If the interviewee points at a building or a block polygon on the map as the residence of the sampled individual, the geo-reference is enabled and the statistical area is identified. Yet, map reading is not widely spread and moreover, since the pointing action is error prone, the quality of the answer has to be monitored.

A short-run spatial solution for the identification of the statistical area within the locality is the adoption of the common geographic terminology related to main spatial entities in each statistical area. The local population does refer to physical elements by agreed-upon names: The mosque, the church, Aladdin high school, the convenient store of Abu Sheriff, etc. Moreover, since many extended families live together in the same location for many years, the family name becomes a geographic indicator as well. Learning the names of principal spatial pointers, in a short preparation stage, and assigning a geo-reference to each one of them facilitate the identification and location tasks. Interviewees and proxies will be able to describe a specified neighborhood using recognized geographic indicators. Moreover, it will open the option to conduct OES interviews by phone, rather than in a face-to-face interaction.

In the long run, the 1995 census will supply less and less reliable ad-hoc geo-references for the administrative records and not only the OES will be hampered but also the UES. The 2008 census will not be able to supply the same solution since it relies on administrative sources and only part of the population is contacted in the evaluation surveys. This census is perceived as the first in a series of censuses, in which the use of administrative data will be increased overtime. Therefore a long run solution is required. As long as the lack of addresses is not an ideologically selected option, it is possible to start an operation of naming streets and numbering buildings in these towns and encouraging the people to report their detailed address to the relevant address collecting agencies. Appointing formal addresses by the local authority and their ongoing use by the local population are pre-requisites for future integrated or fully administrative censuses.

Further Contemplation

The adoption of the geospatial perception and the enabling technology in the statistical production process contributes directly to the expansion of knowledge and to better understanding of socio-economic and demographic processes in a studied population. Moreover, they alter not only the end-result but also the way to achieve it.

Population censuses around the world differ mainly in their data sources and their data collection processes. Geospatial data collection has been the enabler of multi-source data collection, in administrative and in integrated censuses. This trend of the integration of information in a geospatial infrastructure is expected to encompass an increasing number of statistical operations. Censuses, being costly and involving the whole population, are particularly expected to benefit from it, not only in the data integration capabilities but also in efficient sampling and in the collection process itself. Moreover, The pivotal role of the geospatial infrastructure implies the ability to perform *simultaneous processes*, like data integration and small area estimation, which will further enhance the quality of the information generation products.

However, those benefits are conditional. They depend on the identification and recognition of each data item as a spatial entity, with unique spatial ID, not only in databases, but also in the physical reality. A *register-based approach* should be engaged in order to cover the whole population of spatial entities and to record all the events as close as possible to their occurrence.

The implication is that statistical offices have to initiate spatial tagging operations, like giving names, codes and numbers to streets and houses in localities without addresses. The challenge resides in the cooperation of the population in the adaptation to the spatial perception and in the adoption of spatial tags given to the surrounding entities.

References

Bjorensson, A. (2001). “The Landscape Group – Integration of Statistical and Administrative Data with Land Cover Information”. Joint UNECE/EUROSTAT Work Session on Methodological Issues Involving the Integration of Statistics and Geography. WP10. Tallinn, Estonia.

Backer, L.H., M. Tammilehto and P. Guiblin (2001). “In Search of a System of Small Statistical Areas”. Joint UNECE/EUROSTAT Work Session on Methodological Issues Involving the Integration of Statistics and Geography. CRP1. Tallinn, Estonia.

Blum, O. (1999). "Combining Register-Based and Traditional Census Processes as a Pre-defined Strategy in Census Planning". Statistical Policy: WP30. 1999 FCSM Research Conference, Washington DC.

Blum, O. and R. Calvo (2001). “Geospatial Data Collection and Analysis as Crucial Processes in an Integrated Census”. 2001 FCSM Research Conference, Washington DC.

Blum, O. (2003). “Wholistic Approach to Multifaceted Integrated Census”. 2003 FCSM Research Conference, Washington DC.

Brabec, B. (2000). “A Nearest Neigbor Model for Regional Avalanche Forecasting”. UNECE Work Session on Methodological Issues Involving the Integration of Statistics and Geography. WP26. Neuchatel, Switzerland.

Calder, A. (2000). “Spatial Analysis – Why Bother?” UNECE Work Session on Methodological Issues Involving the Integration of Statistics and Geography. WP12. Neuchatel, Switzerland.

Crescenzi, F. (1998). “Applications of Geo-Statistics at ISTAT”. UNECE Work Session on Geographic Information Systems. WP7. Ottawa, Canada.

Ding, G. and M.P. Kwan (2004). “3D-VQGIS: 3D Visualization and Qualitative Analysis of Geospatial Data”. UCGIS Assembly 2004. Maryland, USA.

Gallego, J. (2001). “Using Land Cover Information to Map population Density”. Joint UNECE/EUROSTAT Work Session on Methodological Issues Involving the Integration of Statistics and Geography. WP21. Tallinn, Estonia.

Glickman, H. et al (2003). “New Methodology for an Integrated Census”. 2003 GSS Methodology Conference. London, UK.

Gould, M. and L. Hecht, Jr. (2001). “OGC: A Framework for Geospatial and Statistical Information Integration”. Joint UNECE/EUROSTAT Work Session on Methodological Issues Involving the Integration of Statistics and Geography. WP27. Tallinn, Estonia.

Grasland, C. (2000). “Spatial Homogeneity and Territorial Discontinuities”. UNECE Work Session on Methodological Issues Involving the Integration of Statistics and Geography. WP11. Neuchatel, Switzerland.

Insard, M. (2001). “Geography and New Census in France”. Joint UNECE/EUROSTAT Work Session on Methodological Issues Involving the Integration of Statistics and Geography. WP19. Tallinn, Estonia.

Nael, I. And L. Nordsrom (1998). “Register Data, Point-Based Statistics and GIS, in the Swedish and the Estonian Statistical Systems”. UNECE Work Session on Geographic Information Systems. WP26. Ottawa, Canada.

Nebert, D. (2000). “Building a Geospatial Data Clearinghouse for Data Discovery and Access”. UNECE Work Session on Methodological Issues Involving the Integration of Statistics and Geography. WP8. Neuchatel, Switzerland.

Nirel, R., H. Glickman and D. Ben-Hur (2004). "A Strategy for a System of Coverage Samples for an Integrated Census." Proceedings of Statistics Canada Symposium 2003: Challenges in Survey Taking for the Next Decade.