Development of a National Policy Framework for Census Data dissemination utilising a Geographic Data Management Systems (GDMS)

Harold Wall

IDB/UNFPA Project Coordinator

Common Census Framework

Caribbean Community (CARICOM) Secretariat

ABSTRACT

Improvements in information technology have provided unimaginable opportunities to support Census data analyses in the last two decades. Geographic Information Systems (GIS) has provided new and exciting ways of displaying this data. GIS is designed to display all forms of geographically referenced information. Moreover, one of the most significant valued advantages of a GIS is the ability to provide comparable Census data on a variety of characteristics at small area level, creating the perfect hierarchical model for data sharing, spatial query and data analysis of household information.

However, Geographic Information Systems has proliferated the Region at an incredible rate, creating the need for a National Policy Framework to resolve many data sharing issues. Administrative boundaries constitute essential layers that must be coordinated if effective data sharing, integration and analysis between organisations is to eventuate. Historically many countries have divided social, economic and political responsibilities amongst a variety of Government Ministries. In turn, these organisations have established independent administrative, planning and political boundaries which causes the inability to accurately share socio-economic and demographic information from a Population and Housing Census due to their non-alignment to Census enumeration areas.

The problem of incompatible boundary alignment units restricts the implementation of a National GIS in a country. It is therefore a fundamental prerequisite to ensure that your base unit is accurate; this base unit is the Statistical department's Enumeration District. Consequently, the accurate referencing, demarcation and delineation of Enumeration Districts maps are fundamental prerequisites to ensure that Census dissemination is spatially accurate and that a National GIS is on a strong foundation for geographic hierarchical aggregation, query, analysis and data sharing.

The views expressed in this paper follow the usual caveat relative to being the responsibility of the writer and not of any organisation to which the writer is affiliated. Any errors and omissions are also the responsibility of the writer.

BACKGROUND

The main objective of this paper was the development of common standards and interoperability through the establishment of statistical enumeration districts as the smallest unit of geographic query across Ministries and Agencies to share and use National statistics compiled from Population and Housing Censuses to drive government's social and economic policies. The establishment of a coordinated approach to support Government and private sector applications of geospatial data (data linked to digital maps) in such areas as land information, demography, transportation, social development, agriculture, emergency management, environmental planning and information technology can be achieved through a National Policy Framework.

Administrative and political boundaries constitute essential layers that must be coordinated if effective data sharing, integration and analysis between organisations and data layers is to eventuate. However, historically many countries have divided social, economic and political responsibilities amongst a variety of Government Ministries. In turn, these organisations have established independent administrative, planning and political boundaries which causes the inability to accurately share socio-economic and demographic information from a Population and Housing Census.

Consequently, the accurate referencing and delineation of Enumeration Districts maps are fundamental prerequisites to ensure that Census dissemination is spatially accurate and that a National GIS is on a strong foundation for geographic hierarchical query and analysis. The Enumeration District is the smallest unit of geographic query that data can be collected systematically at a national level. In other words, the problem of incompatible boundary alignment units used by Government and States agencies restricts the implementation of an efficient GIS and subsequently becomes a waste of valuable time and resources. It is therefore a fundamental prerequisite to ensure that your base unit is accurate; this base unit is the Statistical department's Enumeration District.

INTRODUCTION

An organization's strategy for managing Information plays a crucial role in the success of the technology used within the organization. As with any technological innovation, the key factor in its success is how it is applied to inform the organisation's strategic plans, which drives the organisation's objectives.

Geographic Information Systems (GIS) is fundamentally an information system capable of the collection, input, processing, manipulation, analysis and dissemination of spatially referenced data. A GIS play many different roles in various organizations. To some, it is an integrating technology that provides a new perspective and basis for operations and involves major organisational impacts. The role a Decision Support System (DSS) plays in an organization affects its strategic operating environment, the role it plays in a holistic national approach can be delivered through a Geographic Data Management System (GDMS).

A GDMS is fundamentally a concept about facilitation and coordination of the exchange and sharing of data between stakeholders from different jurisdictional levels in the spatial data community. In principle, a GDMS allows the coordinated sharing of data, which is extremely useful, as it enables users to save resources, time and effort when trying to acquire new datasets by avoiding duplication of expenses associated with generation and maintenance of data and their integration with other datasets. This development entails moving from a product-based approach to a process based-based approach.

Many decision support systems that rely on geospatial information have proliferated over the past few years and are increasingly embedded in the operations of Government Ministries at all levels. Societal demand for wide-ranging and accurate geospatial data that can be quickly integrated with other socio-economic and demographic data sets is rising. This places a premium on effective geospatial data coordination and sharing. Achieving more cost-effective use of geospatial data requires concerted and creative efforts to adopt common standards and interoperability across agencies and sectors.

Increasing demands by researchers for immediate access to accurate and documented geospatial data across agencies and sectors are pushing open-source and standardization initiatives to the extremes for many Statistical Organisations in the Region.

Subsequently, fundamental to the effective execution of government policies aimed at optimizing the use of land and land related information, is the knowledge of, and access to accurate, up-to-date, timely, complete and comprehensive socio-economic, demographic, environmental and spatial data. Spatial or Geographic data encompass digital maps with their respective attribute information on natural resources, environment, land ownership, land use, transport, communication, demography and socio-economic factors. Such data can be related to geographic positions including, terrestrial, aquatic, atmospheric, and subsurface regions.

All Governments in the region are attempting to move towards more efficient and effective delivery of goods and services through the use of an effective DSS. Due to the cost and complexity of data collection and analysis of a Population and Housing Census however, the tasks involved in achieving these objectives are both very expensive and time consuming. In addition, in the CARICOM Region, the rapid growth of communities and the diverse conflicting demands for access to the limited available resources, make it especially difficult and challenging to adequately address the needs of all stakeholders in these communities. Manual systems are no longer adequate for the provision of current and reliable information to decision-makers and policy makers. This leads to delays in identifying problems, making decisions, and ensuring that available resources are optimally utilized to provide goods and services to all stakeholders.

In a statistical environment, a GDMS provides an information and operations framework for a major portion of the activities and applications within the organization. The GIS database and software are designed to be used by myriad applications. Benefits result from the reduction of duplicated effort and the widespread availability of resources in many operational areas. The integrated information and operation provided by a GDMS also serve corporate decision making and operations.

This level of integration is typical of many local government and utility applications. Likewise, in many businesses the same data and functionality are integral to many operations.

Therefore it is essential that the demarcation and delineation of Enumeration Districts (EDs) are established accurately prior to a Population and Housing Census so that, clear identifiable features are used as boundaries; EDs are geographically named and coded; EDs can be aggregated into communities and subsequent Administrative areas of various Ministries and demographic data accurately match local geographic areas. It is to be noted that the major component of a GIS is 80% data and 20% maps. Without the data accurately matching the geographic areas on the ground then it is impossible to accurately develop a National GIS or a GDMS.

The problem of incompatible boundary alignment units restricts the implementation of GIS for urban and regional planning. Many utility agencies often attach personal data to postal codes, while demographic data is attached to enumeration district boundaries making accurate analysis between these non-coterminous boundaries virtually impossible. In the example of child immunisation rates, records detailing the number of children immunised are attached to health districts, so to calculate if this number is below or above the average population data is required.

However, due to the incompatible alignment of health districts and enumeration district boundaries demographic data cannot be cross analysed with medical information. Thus, without additional information it is impossible to establish if child immunisation rates, within a particular health district, are below or above an acceptable limit, restricting the accurate planning of health services. As the powers of GIS for data analysis are increasing, countries world-wide are realising the problems associated with incompatible boundary alignment.

DEFINING THE PROBLEM

Consequently, Government agencies using different criteria and methodologies for defining management boundaries that rarely coincide with Census Enumeration Districts make it quite difficult to share accurate data spatially for effective sharing of data for the same geographic area.

A number of countries experience this problem where management units have evolved in an uncoordinated manner. For example many countries use Enumeration Districts (EDs) to collect and disseminate demographic information, post codes to define regions for mail delivery, electoral boundaries form voting districts, school catchment areas for education planning and natural boundaries the basis for environmental monitoring. Due to the structure of boundaries as polygons, problems occur when technology such as GIS is used to integrate and cross analyse data based on these non-coterminous boundary units.

Conceptually many organisations have divided the spatial environment into different management units. The size and shape of these management units varies between organisations. Subsequently, each organisation aggregates these boundaries at an entire state or country scale, creating independent layers of information. (Serryn,1999) The acceptance of Enumeration districts as the smallest unit of geographic query with the accompanying socio-economic and demographic data provides the basic foundation for an efficient, robust information decision support system. However it must be noted that respective Statistical organisations must take the responsibility to ensure that their Enumeration districts are accurately delineated, if it wants other agencies to use their associated data.

In the absence of a National Policy Framework on land and land related Geospatial information many debilitating factors for Census data management have developed. These issues have precipitated government's need for a national policy framework for land information systems through geographic data management systems and are as follows:

- Census Data comparison and cross-correlation are severely hindered by the several non-standardized referencing and operating systems and map scales across Government agencies.
- Lack of knowledge and inaccessibility to existing data stores has resulted in difficulties in compiling basic inventories of land and household assets
- Interagency coordination does not exist and in the few areas is, is restricted to facilitate the ease of sharing government collected information, therefore leading to costly duplication of data collection, redundancy and analysis.
- Delays in the development of National Physical Development Plans due to the limited access to up-to-date relevant Census data, geographic data and satellite imagery.
- Critical shortage of Geographic Information Systems (GIS) management specialists and skilled operational staff.
- No metadata compilation of collected information.

COMMUNITY DELINEATION

The method used for the delineation of Communities was to assign each ED a community name based on field information collected from the Continuous Sample Survey of Population, Population Census Visitation records, Map-comment forms and visits to problematic areas.

Consequently, all contiguous EDs with the same community name were then aggregated and delineated to form the boundaries of the respective named community. EDs with more than one community names i.e. EDs that are rural in nature and had ribbon-type development stretching along a road and/or EDs with Housing developments in them, were further analysed using individual household addresses from Visitation Records to identify and use the dominant name or both.

These communities were then geographically coded into a four digit code to reference community administrative areas. Consequently, data integration from various social surveys were now able to impact on the planners for an improved understanding of social ills by visualizing spatially, in relation to other social infrastructures/facilities, such as proximity analysis to schools, hospitals, environmental areas and Police Stations.

LITERATURE REVIEW

South Africa

South Africa is currently expressing concern about the lack of boundary coordination and demonstrating interest in new methods for organising spatial units within the country. Gavin (1999) explains the country's division into nine separate provinces. These provinces have been further segmented into 364 magisterial districts with boundaries overlapping the 834 local government boundaries. It has also been noted that the magisterial district and local authority boundaries also cross provincial boundaries. In addition to these boundaries there are approximately 85,000 enumeration districts and postcodes defined largely by the routes travelled by postal service workers, similar to those established in Victoria (Gavin, 1999). In summary, there are plans for the future as a strong push in South Africa for all 'higher-order' spatial units to be derived by aggregating smaller units, starting with the enumeration areas (Gavin, 1999).

United Kingdom

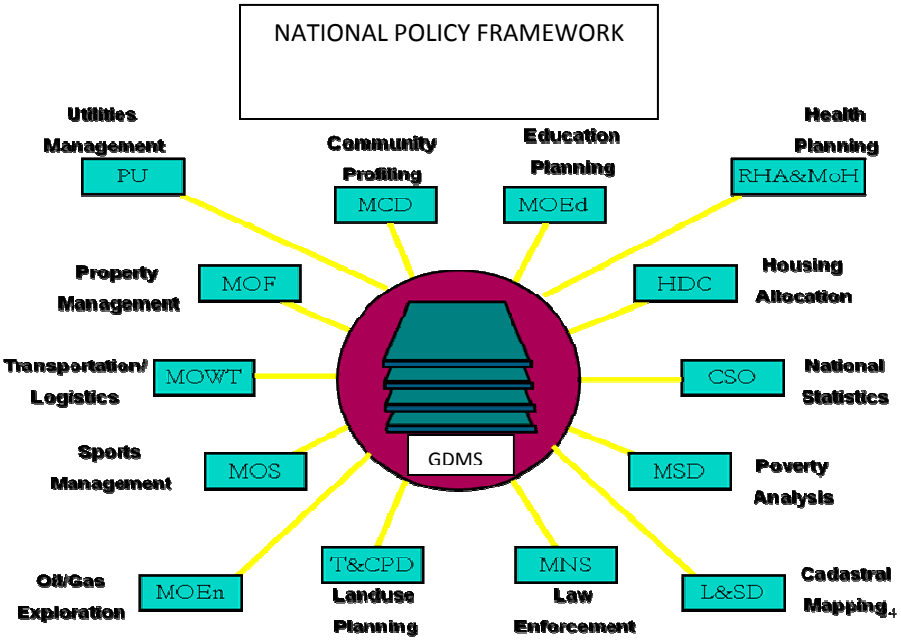
In the UK, problems associated with incompatible boundaries have been well documented by authors Martin (1991), Openshaw (1992), and Duke-Williams and Rees (1998). Structurally the smallest units for census data are the Enumeration Districts (EDs) in England, Wales and Northern Ireland, and Output Areas (OAs) in Scotland. However, due to the differing spatial requirements of organisations, census data boundary units do not provide a spatial boundary system satisfying the majority of users. As a result, there is a high demand for data to be produced on a range of alternate boundary units (Duke-Williams & Rees 1998). One prime example is Local Government Users in England and Wales requiring census data for the local government area. The reason is a direct relationship between the "central government grants to local authorities based on size and characteristics of local authority populations," (Duke-Williams & Rees 1998). The primary argument being parallel to the health example of Victoria where postal codes are used for the collection of patient related information and require cross analysis with demographic information.

CONCLUSION

Previously GIS problems associated with incompatible boundary alignment were negligible, as the data produced on analogue maps were rarely cross analysed. Consequently, organisations collected copious quantities of data with various boundaries to meet their individual needs. Now the technology to integrate, and cross analyse data sets is available, however severely restricted by non-coterminous boundaries. In an effort to take full advantage of current technology, changes in design of the spatial framework are imperative.

The proposed solution is to ensure that enumeration areas or enumeration districts are accurately defined and geographically delineated on a national base map; this ensures accurate enumeration coverage and eliminates omission of census areas. By using this approach, it is anticipated that each political and administration boundary will be formed through the aggregation of enumeration districts, where the smallest spatial unit is the household. In turn, this system would enable rapid and efficient cross analysis between data sets. It is envisaged that the reorganisation of boundaries into one well designed structured system will have the ability to revolutionise data integration and analysis methods at a relatively high accuracy and low cost compared with previous solutions.

A National Policy Framework identifies the need for a GDMS to coordinate the efforts of the various agencies that have to rely on data for informed decision making to meet their administrative needs. The following diagram explains.



A Framework Structure for Co-ordination and Integration of Census Data

Although GIS analysis is restricted by incompatible boundary units, GIS technology is proving valuable in the formulation of solutions to the problem. There are at least 40 International Organisations for Standards (ISOs) and ISO Agreements (ISO/TC211) that have completed hundreds of classification of content related standards for a National GIS development, namely ISO19101 - ISO19140. These standards can be researched for the preparation of a more comprehensive technical document to be developed and further ratified by a steering committee. One of main goal of the five goals of ISO/TC 211 is to increase the availability, access, integration and sharing of geographic information and enable inter-operability of geo-spatially enabled computer systems.

The development of strong relational database software and modern technology has opened the door for improved GIS software to be used for spatial and aspatial analysis. Effective reporting can lead to informed decisions that directly affect people lives positively on the ground, rather than just the formulation of utopian plans. The linking of census demographic data and administrative maps has been hampered by the lack of up-to-date National topographically referenced base maps.

This has been a debilitating factor in the delivery of GIS technology over the years. One major problem that GIS can solve is to unravel the complex databases that have confused users of census and social surveys data, by providing a visual element that offers popular user appeal and can provide an ideal starting point for information discovery. This can replace traditional tabular based information retrieval strategies with one based on visualisation and new knowledge discovery.

REFERENCES

Australian Bureau of Statistics (1996) “Statistical Geography Volume 2 Census Geographic Areas Australia”, Commonwealth of Australia.

Car, A. (1997) “Hierarchical Spatial Reasoning: Theoretical Consideration and its Application to Modeling Wayfinding”, PhD Thesis, Department of Geoinformation, Technical University, Vienna.

Gavin, E. (1999, February 15). Masters Research

Martin, D. (1991) “Geographic Information systems and their Socioeconomic Applications”, London: Routledge.

Trinidad, G. and J. Crawford. (1996) A tool for transferring attributes across thematic maps. “AM/FM” 127-137.

Serryn Eagleson1 et all 1999 Spatial Hierarchical Reasoning Applied To Administrative Boundary Design Using GIS

Duke-Williams, O. and Rees, P. (1998) Can Census Offices publish statistics for more than one small area geography? An analysis of the differencing problem in statistical disclosure. “International Journal Geographical Information Science” 12(6): 579-605.

Wall, H (2008) The application of a Geographic Data Management System for dissemination of Census and Survey Data for Government Ministries in Trinidad and Tobago

Wall, H (2007) A Proposal for a National Spatial Data Infrastructure – Trinidad and Tobago