Intra-Censal Geographical Information Systems: Application To Binational Border Cities

James Pick, W. James Hettrick, Nanda Viswanathan, and Elliott Ellsworth
University of Redlands
Redlands, CA 92373-0999 USA

***Abstract*-This paper presents a new design for a census-based binational geographic information system. GIS has had limited application intra-censally i.e. between nations. The reasons are the formidable obstacles of standardization of attributes, matching of meaning of attributes, consistency of data-bases, and consistency of boundary files. Many problems in public administration can be solved by using GIS systems that encompass data from several nations. This challenge is even greater for GIS applications involving small area data. The present paper presents a general framework for dual-nation GIS's and then offers an example of its application to the United States and Mexico. The example is a successful GIS constructed for the twin cities of the U.S.-Mexico border. The results of the paper can apply to combining European censuses, many of which lack standardized formats or to combining censuses of European and neighbor nations, where the differences are greater. The paper discusses the results and makes suggestions for improvements in dual-censal GIS applications in public administration.**

I. Introduction

A geographical information system (GIS) is "an information system that is designed to work with data referenced by spatial or geographic coordinates" [1]. A GIS contains a spatially referenced data-base, to which analysis procedures are applied. GIS's are utilized extensively in the public and private sectors. In public administration, GISs are utilized for tax records, land parcel maps, utilities management, city and regional planning, disease analysis, public works projects, police and public safety work, and many other applications [2]. A GIS is underutilized if it is applied only to mapping applications. A true GIS is utilized for analysis and modeling that have spatial referencing. Examples of analytical techniques are simulation, spatial analysis, statistical analysis, and optimization.

A GIS may also be classified as a Decision Support System [3]. It meets the DSS definition by having the requisite data-base management system, model management system, and user interface [4]. A GIS is a special DSS in having a spatially referenced data-base. The spatial referencing of attributes is to the points, lines, and polygons that are the building blocks of GIS boundary files. As a DSS, a GIS is special in having models and analytical tools that are spatially referenced. For instance, a public utilities GIS may seek to minimize the cost of putting in a new sewer line, and the sewer line is spatially referenced through linkage of its geographical features to attributes.

For applications in public administration, a GIS is helpful to decision makers at the township, city, regional, state, and national levels. It has been adopted extensively for decision-making in U.S. cities [2] and in most European cities. On the other hand, it has been applied to a limited extent at the local level in developing nations. The present paper offers a framework for GIS's to be built based on the census information from two nations. This is in contrast to the vast majority of GIS applications, which utilize the information from a single census. This framework becomes increasingly important as a GIS's areal units i.e. polygons become smaller. Multi-census data are often combined at the national levels and sometimes at the state levels, but less so for counties and rarely for city small area units, referred to as census tracts or block groups in the U.S. This paper will point out the challenges in building a dual census GIS and how the challenges may be overcome through an integrated data-base design.

The research questions are the following:

1. What are the key problems in designing a dual-census GIS, and how can the problems be surmounted?
2. How can GIS analysis and modelling be achieved with a dual-census GIS and what are some of the constraints?

II. Creating a Binational, Dual-Census GIS -- The Challenges

Two or more national censuses rarely coordinate with each other on their design, operations, and standards. For instance, census unemployment definitions are different in many nations (for one example, see [5]). The United Nations [6] and World Bank [7] have difficulty in achieving consistency of definitions in their compilations of worldwide data-sets. They often have to eliminate nations that do not comply to standard definitions and/or nations that do not collect particular attributes. Another challenge to establishing consistent data-sets is the reduction in the number of attributes, as unit areas get smaller. This stems from the sheer size of collecting small area data and from the needs of confidentiality. Regarding confidentiality for instance, at the tract and block group levels, most information of the U.S. Economic Census is suppressed due to confidentiality [8]. Another large intra-censal difference is in the approaches to coding and table construction. Some censuses such as the U.S. Census are systematically hierarchical in their table construction, while others including Mexico are a-hierarchical and lack a systematic overall approach to table construction.

The four major challenges in dual-census GIS are as follows:

Attribute definitional matching. Although some socio-economic attributes such as population, gender, births, and population density have standard definitions, other important attributes such as labor force participation, unemployment, and poverty lack standard definitions that are well accepted worldwide. The challenge for dual-census GIS is to find a limited set of attributes that match exactly and/or to identify a tolerance for other attributes that match closely enough, or can be adjusted to match closely. An example for the U.S. and Mexico is unemployment. In the Mexican Census underemployed can be defined as the ratio of persons working 1 to 32 hour per week to the economically active population. In the U.S. Census, the closest available match to this definition is persons working 1 to 35 hours per week to total labor force. These definitions are slightly different, not only in the range of hours of worked, but in shadings of meaning between "economically active" (Mexico) and "labor force" (U.S.). However, the present researchers considered the definitions close enough to "match." It is important to point out that perhaps twenty percent of the attributes available in both the U.S. and Mexican censuses can be matched exactly or closely. Thus, a limitation is that a dual-census GIS will have a reduced attribute set for binational maps and analysis than a single-census GIS.

Data-base design. The challenge here is to design data-bases that can systematically support a GIS for a large number of attributes and complex coverages. Two censuses produce individual data products that provide the data either in hardcopy, electronic, or web-accessible form. The problem for a dual-census GIS is that, between censuses, the data are arranged quite differently, in respect to coding, table dimensionality, order, calculation procedures, aggregation/ desegregation, and spatial referencing. It is almost impossible to construct a large-sized GIS accessing two nations' sets of data products directly, i.e. without translation or conversion.

Our approach is to convert unique data products of censuses into standardized sets of relational tables supported by a common data-base (see Figure 1). The definitions and pointers can be set differently in the relational tables for each census, so that the censuses' data dictionaries and relationships are equalized in format. One small example is the Access coding defining economic activity ratio, for the relational data-base tables of INEGI attributes. It consists of three macro commands: P34 AS ECONOACT, P01 AS POP90, [ECONACT]/[POP90] AS RECONACT. Pointers are applied differently for the two censuses, since each has differently arranged tables. There are three advantages to our approach: (1) data formats and definitions are standardized and efficiently maintained, (2) formula computations can be performed systematically on the data of both censuses, (3) the data can be easily aggregated as output with a single unique key field to reference the geography, and (4) the output data also can be transferred in a standard format for further statistical analysis or other modeling.

Although the process of aligning the dual-census data could be accomplished through tedious and complicated flat file conversion steps, the dual relational data-base design is enormously more efficient and more easily documented. Among other things, the macro scripts used to create formulas can be included directly as part of the documentation, especially since formulas for derived attributes are often quite complicated.

Boundary file Matching. Another problem for dual-census GIS is the matching of boundary files i.e. sometimes called coverages [9]. In most nations, the census or an allied federal agency, is responsible for maintaining national boundary files, sometimes called "base coverages" [2]. In the U.S., for instance, all the national geography is maintained by the U.S. Census Bureau in the nationwide "Tiger File" [8]. In other countries, the national geography is not unified into one national base file, but is maintained in a number of different base files that follow common definitions. Third party commercial vendors often build their national boundary files utilizing the national base file.

A particular challenge for coverages of dual census GIS's is to match the geographic boundaries at national borders. A starting point is that the two censuses' geographic projections need to be adjusted to be the same. A projection is a particular rendition of the 3-D surface of the earth to a flat surface and includes such well known ones as Albers equal area, asimuthal, cylindrical, Lambert conformal conic, and transverse [9]. However, the boundary file matching problem for borders extends beyond mere projection alignment. The problem is that the procedures to produce two nation's base files are different. The point, line, and polygon information is drawn from different sources such as satellites, GPS and other field measurements. At national borders, the differences show up acutely. When one nation's boundary file (or one derived from third party providers) is aligned with a second nation's, such border features as shared bridges, highways, railroads, rivers, etc. do not align. This requires development of special GIS macros to make the compromises necessary to align the borders. This is best done in an advanced GIS package, such as ArcInfo, that has a powerful macro language, "snapping" and other alignment features.

Modeling/Analysis with dual-census data. The essence of any GIS is its analysis and modeling features. This is tied to the idea that a GIS is a form of DSS, and that DSS is based on modeling features. In a dual-census GIS, there is the potential to do binational spatial analysis and modeling that is frequently restricted by lack of standardized data sets and geography. The spatial analysis can yield powerful results. For instance, joint environmental analysis can be done for air-polluted twin border cities; bi-national transport corridors can be planned spatially taking into account constraints, land features, and important determinant variables on both sides of the border. The obstacle to binational modeling is that the modeler needs to consider a valid model in both countries, not one with assumptions applying only to one nation or the other.

In summary, this section has examined the four major problems to achieving a dual-census GIS, and suggested for each ways to overcome the problems.

III. Example: The U.S.-Mexico Border Twin Cities Project

The Twin Cities GIS Project utilized the general framework of this paper to construct a dual-census GIS for the twin cities of the U.S. and Mexico. There are eight major twin cities located along the 2,000 mile U.S.-Mexico border. The twin cities include the large ones of San Diego-Tijuana and El Paso (Texas)-Ciudad Juarez. They also include medium sized cities such as Brownsville (Texas)-Matamoros (Tamaulipas) and smaller cities such as Nogales (Arizona)-Nogales (Sonora) and Eagle Pass (Texas)-Piedras Negras (Coahuila). The total population of the twin cities in 1990 was 5.1 million, but this is expected to rise to 6.8 million by 2000 and an estimated 10.6 million by 2020 [10]. The twin cities are particularly important as a result of the NAFTA agreement, since they serve important conduits for trade, commerce, and transportation between the two large nations.

The cities have suffered from lack of dual public administration planning. For instance, the metropolitan planning agency, SANDAG, has an advanced GIS for the metropolis of San Diego, but does not include any data on Tijuana, even though that city has surpassed one million population. Likewise, Tijuana's limited GIS from INEGI, the Mexican Census, includes no data on the U.S. side. Such planning deficiencies are repeated in all the border cities. Planning is better on the U.S. side and often utilizes GIS, but dual-census GIS's are lacking.

The Twin Cities GIS Project overcame these problems based on the system model discussed earlier and illustrated in Figure 1. Small area data were utilized for block groups in the U.S. and AGEBs (Area Geografía Basica) from the Mexican Census. Overall, these small area units had average population of 1,739. The U.S. Census data were converted and re-arranged in relational tables of a Microsoft Access data-base. Likewise, data from the Mexico Census of 1990 (CIEN files) were converted and re-arranged in tables in Access. In many cases, it is necessary to perform formula calculations to convert raw variables into finished ones, for instance, the project calculated "one-room housing" as the ratio of single bedroom housing units to all housing units. Likewise, it computed "dependency ratio" as the ratio of population less than 18 plus population 65 or over to the population between 18 and 64. The formula computations for 16 matched variables were done in the Access scripting language. The variables included population, social, economic, and housing ones. The variables were output with systematic geographic key fields. The key field is a unique identifier for each block group and AGEB. The files output were merged binational attribute files that merged the Mexican relational table and U.S. relational table. There was a unified geographic key field that coded each polygon in the binational coverage.

In another procedure, the U.S. small area boundary files were acquired from the GIS provider ESRI and the Mexican files from the Instituto Nacional de Estadística, Geografía, y Informática (INEGI), i.e. the Mexican national statistical bureau. The U.S. and Mexico border and its features were aligned by utilizing macro language and spatial commands in the advanced GIS software, ArcInfo. The alignment consisted of applying algorithms to modify the boundary files so the boundary edges conformed and the common features such as roads, rivers, etc. were joined at the border, in order to produce a two-nation boundary file that was seamless at the border. There is not space in this paper to discuss the algorithms, but they consisted of heuristic estimates of best fit, with more dependence on the U.S. side, since the U.S. bsaed boundary file is considered more accurate. The result was a merged binational boundary file (see Figure 1). This file contains the geographic key field for the U.S. and Mexico, and the shape parameters for the polygons that form the small areas. Finally, the merged binational attribute file and merged binational boundary file were ready for input into the ArcView GIS software, which joined the two files together for mapping based on the geographic key field (see Fig. 1).

An example of the twin city output is the distribution of poverty in El Paso and Ciudad Juarez, shown in Figure 2. El Paso is to the north and east of the winding boundary line, and Ciudad Juarez is to the west and south. This example indicates that poverty is located mostly in the city center of El Paso, whereas it is in the peripheral areas of Ciudad Juarez. Such information can be useful in joint binational planning between the two city governments. Also, in Ciudad Juárez, there is a distinctive pattern of rings around a central business district with much higher poverty in the periphery, which contrasts to an urban structure of many wedge shaped sectors on the El Paso side. The urban structure of poverty would be useful to public planners on both sides in understanding the distribution of poverty in a twin city context. Another planning illustration is shown for home ownership in Figure 3. Home ownership is defined identically on the two sides of the border as the proportion of owner occupied housing units to all housing units. Although both sides of the border show similar patterns, with lower home ownership in the center and higher home ownership in the periphery, the interpretation is opposite and revealing. In the culture of Mexico, high home ownership connotes economic privation, while rental status i.e. low home ownership tends to connote prosperity and wealth. On the other hand, on the U.S. side, low home ownership connotes deprivation and high home ownership indicates prosperity. This unusual culturally-reversed situation is useful to public and private sector planners. For instance, it might effect the choice of plant locations by multinational companies that have facilities on both sides of the border. Home builders operating on both sides of the border would likewise find this pattern useful.

The project also utilized the dual-census data for spatial analysis and modeling; one example was applying cluster analysis to study the urban structure and then constructing an index of binationality based on the cluster analysis to determine the extent of similarity or difference between the two sides. In this case, clustering was done for the variables in the merged binational attribute file. This provided a set of distinctive geographic clusters for the twin city. Each cluster was characterized and interpreted. The binational index measured the extent to which the clusters were located separately on one or the other side of the border versus extending across both sides of the border. The index indicates for a given characteristic, whether it is mixed on both sides or distinctive. This information can help public administrators on both sides to better understand similarities and disparities that underlie the subtleties of shared governance of the twin cities.

The binational design offers many other possibilities for analysis and modeling. Regression-based forecasting models predict future values for small area units. Examples are regression-based projection of future land use or of educational levels throughout the twin city. Optimization models can be run to determine, for example, the areas of the city that have the maximal potential for housing improvement programs. The *binational* aspect of the analysis opens up new realms for analysis and offers new support for policy setting and decision making.

There is a wealth of information from the Twin Cities GIS Project that can benefit planners in both the U.S. and Mexico. Even though there are constraints on which attributes can be used and on the matching of variables, there are many new perspectives in viewing these crucially important and often large "portal" cities. The challenge shifts from technical to political and governmental. Once technical barriers are overcome, the twin city governments need to have the will to cooperate and share data, knowledge, and expertise on an ongoing basis in order to understand each other's cities and cityspaces better? Beyond the cities' willpower, do the national governments provide support and encouragement for this cooperation. There needs to be a real governmental interest and motivation in order to have the most use and benefit of such a border twin cities GIS.

In the context of Europe, this example is important. European cooperation is increasing in many arenas, and dual-nation GISs can assist public administration in cooperative planning. In order to achieve dual nation GIS's, public administration agencies at the levels of cities, states, national governments, and the European Union need to encourage systematic censal collection and standardized design frameworks such as the one illustrated to put binational GIS's into effect. The benefit could be great, since Europe has numerous cities located at national boundaries and large amounts of population that is close to borders. Furthermore, many public administration problems and challenges of national borders remain, even as some of the barriers are being lowered. Also, there are many borders of Europe that are not within European Union or trading blocks and may adjoin instead non-European neighbor nations.

IV. Conclusion

This paper has identified four key areas of design issues for a dual-census GIS. They are: (1) attribute definitional matching, (2) data-base design, (3) boundary file matching, and (4) Modeling/analysis with dual-census data. All of them can be addressed utilizing a variety of methods and approaches. Attribute definitional matching can be done exactly and/or within certain threshold tolerances that are acceptable for the project at hand. We recommend a data-base design in which relational tables are applied to reconstitute standard census tables as search

relational ones through use of pointers. This is done separately based on data from each side of the border. Utilizing a common geographic key field, the data can be joined together into a merged binational attribute file. Likewise, boundary files that are often particularly incompatible at their borders can be adjusted by applying systematic algorithms; the adjusted dual files are merged into a uniform and geographically keyed file. Modeling and analysis is based on the merged binational attribute file. Models need to be developed based on assumptions that apply to both sides of the border.

A dual census GIS has a number of constraints. A large proportion of the variables of similar concept may be incompatible and not able to be adjusted to be matched on definition. Another problem is that census tabular structure may be so contorted that conversion to a relational file is too costly. Another type of constraint is the extent of differences in boundary files. Some censuses do not have detailed and accurate base maps, but rather are utilizing base maps from old manually constructed maps or worse yet maps that compromise a lot though interpolation of line segments.

Perhaps the largest challenge for effecting dual-census GIS's is in the will and determination of governments on both sides of national borders to cooperate in order to design, develop, collect standardized date, utilize, and maintain such a GIS. There is often distrust and a history of deficient cooperation in binational GIS-based planning situations. Building up and making use of a dual-census GIS requires leadership and vision to overcome the obstacles outlined in this paper and see through to the new potentials to be realized.

Acknowledgement

Appreciation is expressed to the Ford Foundation, which sponsored the research project.

References

[1] Star, J. and J.E. Estes. 1990. *Geographic Information Systems: Developments and Applications*. Saddle River, New Jersey: Prentice Hall.

[2] Huxhold, William E. 1991. *An Introduction to Urban Geographic Information Systems*. New York: Oxford University Press.

[3] Murphy, Lisa. 1995. "Geographic Information Systems: Are They Decision Support Systems? Proceedings of the 28th Annual Hawaii International Conference on System Sciences, 131-140.

[4] Turban, Efraim and Jay Aronson. 1998. *Decision Support and Expert Systems*. 5th Ed. Upper Saddle River, New Jersey: Prentice Hall.

[5] Fleck, Susan and Constacne Sorrention. 1994. "Employment and Unemployment in Mexico's Labor Force." Monthly Labor Review 117(11), pp. 3-31.

[6] United Nations. Various Years. *U.N. Demographic Yearbook*, New York: United Nations.

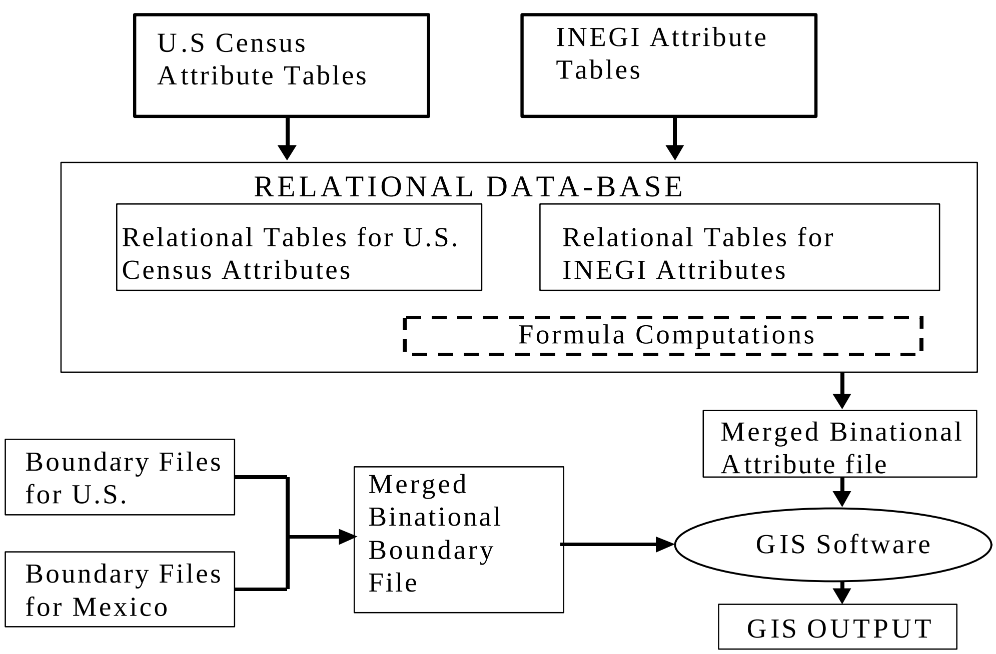
[7] World Bank, The. 1999. *World Development Indicators*. Washington, D.C.: The World Bank.

[8] U.S. Bureau of the Census. 1999. Census Website http://www.census.gov.

[9] Clarke, Keith C. 1997. *Getting Started with Geographic Information Systems*. Upper Saddle River, New Jersey: Prentice Hall.

[10] Peach, James and James Williams. 1999. "Population Projections for the U.S.-Mexico Border Region." Paper Presented at the Meeting of the Association for Border Studies, Fort Worth, Texas, April.

Figure 1. Design of Binational Twin City GIS



search

Figure 2. Poverty in the Border Twin City
of El Paso, Texas, USA, and Ciudad Juarez, Mexico

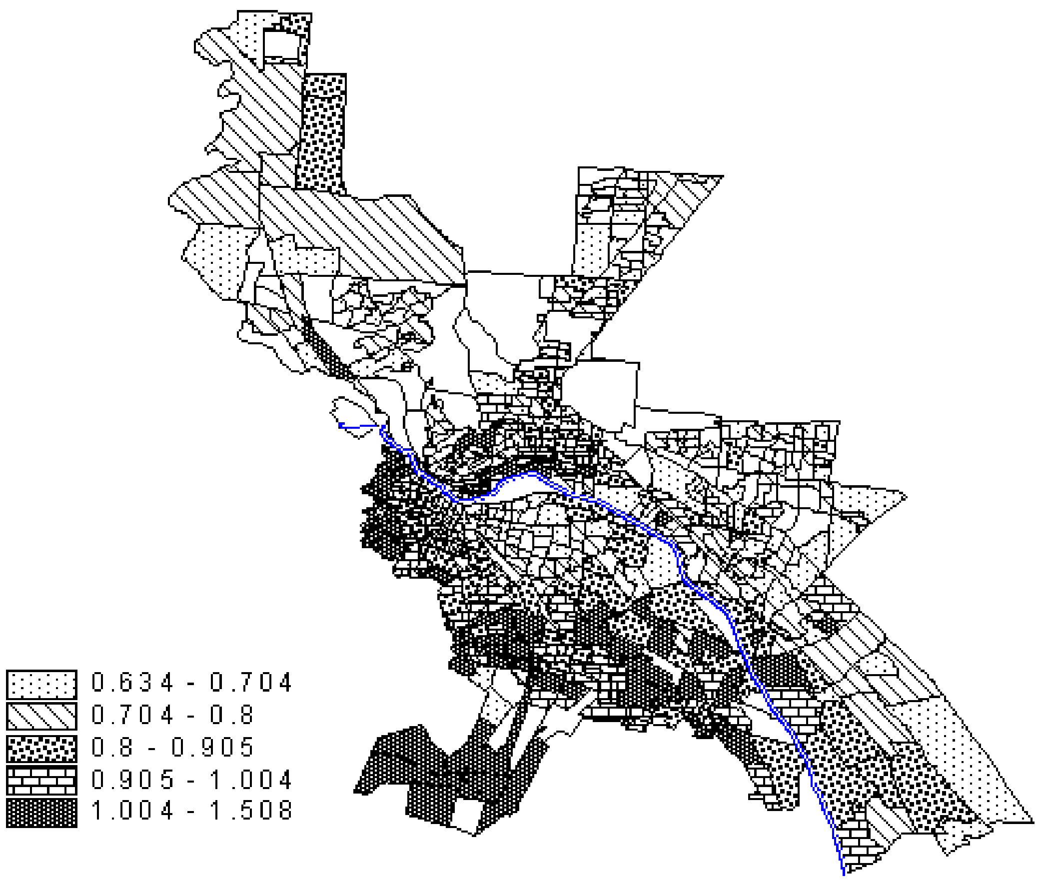


Figure 3. Home Ownership in the Border Twin City of Brownsville, Texas, USA and Matamoros, Mexico