DESIGNING GEODATABASES FOR THE GENERAL AUTHORITY FOR STATISTICS OF THE KINGDOM OF SAUDI ARABIA

by

ALGHAMDI, KHALID ABDULLAH

(Under the Direction of Yuxia (Lucy) Huang)

ABSTRACT

Developing a new system of both statistical surveys and geographic entities for the General Authority for Statistics (GaStat) in Saudi Arabia is needed to respond to fast growing statistical data as well as to be more relevant with users and related resources through providing a diversity of datasets. This project developed a new geodatabase conceptual model for the GaStat by adopting the methods used by the United States Census Bureau. The new model consists of two main components: statistical surveys and geographic entities. First, statistical surveys use the methods of field surveys, partnership agreements, and self-response to feed the database in the Information Bank, which is a data warehouse and will be used in GaStat. More specifically, additional types of statistical surveys are identified and included in the proposed model. These types include the Saudi Community Survey, Saudi Housing Survey, Saudi Income Survey, Saudi Spending Survey, Saudi Economic Survey, Saudi Industry Survey, Saudi Agricultural Survey, and the Saudi Employment Survey. Further, a series of formatting time methods including quarters, one year, three years, and five years is used in the statistical systems in order to provide up-to-date information. Second, similar to the geographic entities used in U.S. Census Bureau, geographic entities for the GaStat are classified into two groups: legal and administrative entities, and statistical entities based on their corresponding geographic subdivision, and they are further organized in a hierarchical structure. This design should provide a powerful tool for collecting information and creating a

standard set for the GaStat in order to retrieve the requests from users. More specifically, the legal

and administrative entities include the country and its provinces, governorates, holy areas,

economic zones, ZIP code areas, school districts and voting districts. Statistical entities include

regions, statistical tracts, statistical block groups, statistical blocks, urban areas, urban growth

areas, places, sample-data areas and governorates subdivisions. Additionally, a unique reference

numeric code is used to integrate between the data from both

statistical surveys and geographic entities. The proposed geodatabase model is expected to address

the limitations of the current systems in the GaStat.

INDEX WORDS:

Geographic Entities, Statistical Surveys, Geodatabases

DESIGNING GEODATABASES FOR THE GENERAL AUTHORITY FOR STATISTICS OF THE KINGDOM OF SAUDI ARABIA

by

ALGHAMDI, KHALID ABDULLAH

BACHELOR OF SCIENCE IN MANAGEMENT INFORMATION SYSTEMS, YANBU UNIVERSITY COLLEGE IN ROYAL COMMISSION FOR JUBAIL & YANBU, SAUDI ARABIA, 2012.

A project research Submitted to the Graduate Faculty of The Texas A&M University-Corpus Christi in Partial Fulfillment of the Requirements for the Degree

MASTER OF GEOSPATIAL SURVEYING ENGINEERING

CORPUS CHRISTI, TEXAS

2016

© 2016

ALGHAMDI, KHALID ABDULLAH S

All Rights Reserved

DESIGNING GEODATABASES FOR THE GENERAL AUTHORITY FOR STATISTICS OF THE KINGDOM OF SAUDI ARABIA

by

ALGHAMDI, KHALID ABDULLAH

Major Professor: Dr. Yuxia (Lucy) Huang

Committee: Dr. Gary Jeffess

Dr. Richard Smith

Electronic Version Approved:

JoAnn Canales Dean of the College of Graduate Studies Texas A&M University – Corpus Christi May 2016

TABLE OF CONTENTS

| | Page |
|-----------|--|
| LIST OF | ΓABLESvi |
| LIST OF I | FIGURESvii |
| CHAPTEI | 3 |
| 1 | INTRODUCTION1 |
| | 1.1 Introduction |
| | 1.2 Purpose of Research |
| | 1.3 Significance of the Study4 |
| | 1.4 Definitions of Terms5 |
| 2 | LITERATURE REVIEW6 |
| | 2.1 The Saudi Arabia General Authority for Statistics |
| | 2.2 The Current System of Geographic Entities for the General |
| | Authority for Statistics |
| | 2.3 The System of Geographic Entities |
| | 2.4 The United States Census Bureau |
| | 2.5 GeoID |
| | 2.6 Comparison of Datasets in the US Census Bureau and Datasets in the Saudi |
| | Arabia General Authority for Statistics |
| | 2.7 Michael Bauer Research (MB-research) |
| 3 | METHODOLOGIES |
| | 3.1 Introduction |

| | | 3.2 Concepts of Statistical Surveys | 26 |
|------|--|---|----|
| | | 3.3 Concepts of Geographic Entities | 27 |
| | | 3.4 The New System of Geographic Entities for the General Authority | |
| | | for Statistics | 30 |
| | | 3.5 The Geographic Entities for the USA Census Bureau and the KSA General | |
| | | Authority for Statistics | 31 |
| | | 3.6 Concepts of Coding for Geographic Entities | 33 |
| | | | |
| | 4 | RESULTS | 34 |
| | | 4.1 Finding Types of Statistical Surveys | 34 |
| | | 4.2 Theoretical Geographic Entities of the Saudi Arabia General Authority | |
| | | for Statistics | 35 |
| | | 4.3 Applied of Coding for Geographic Entities | 42 |
| | 5 | CONCLUSIONS | 44 |
| REFI | 4.1 Finding Types of Statistical Surveys | | |
| APPI | END | ICES | 51 |
| | A | Map of General Authority for Statistics Geographic Entities | 51 |
| | | | |

LIST OF TABLES

| | Page |
|--|------|
| Table 2-1: Geographic Entities of KSA | 24 |
| Table 2-2: Geographic Entities of USA | 24 |
| Table 3-1: The Current System of Geographic Entities | 28 |
| Table 3-2: The New System of Geographic Entities | 28 |
| Table 4-1: GeoID for the New System of Geographic Entities | 42 |

LIST OF FIGURES

| Pa | ige |
|---|-----|
| Figure 2-1: Standard Hierarchy of Census Geographic Entities | 9 |
| Figure 2-2. Esri Demographics and Business Data List | 16 |
| Figure 2-3A. Esri Demographics and Business Tables from Census 2010 | 18 |
| Figure 2-3B. Esri Demographics and Business Tables from Census 2010 | .19 |
| Figure 2-4A. Esri Demographics and Business tables from American Community Survey | .20 |
| Figure 2-4B. Esri Demographics and Business tables from American Community Survey | 21 |
| Figure 2-5. Esri Demographics and Business Tables from General Authority for Statistics | .22 |
| Figure 2-6. Topics Use by Esri | .23 |
| Figure 4-1. Theoretical Standard Hierarchy of General Authority for Statistics Geographic | |
| Entities | .35 |
| Figure A-1. The Kingdom of Saudi Arabia | .51 |
| Figure A-2. Governorates of the Kingdom of Saudi Arabia | .52 |
| Figure A-3. Al-Riyadh Province, Governorate and City | .53 |
| Figure A-4. The Three Statistical Units of Geographic Entities System | 54 |
| Figure A-5. Economic Zones and Urban Areas Geographic Entities. | .55 |
| Figure A-6. Four Types of Geographic Entities | 56 |
| Figure A-7. The Kingdom of Saudi Arabia Map Seven. | .57 |

CHAPTER 1

1.1 Introduction

The Ancient civilization of Egypt was one of the first societies that developed the statistical methods to build data to measure population, agriculture and water resources (The people, 2008). Later, the Industrial Revolution became an influential movement in the statistical surveys, censuses, and third part records to collect the information in order to design statistical data (Rothwell & Rustemeyer, 1979). Moreover, a powerful analysis of statistical data is needed through taking advantage of the spatial data ability to collect, visualize, and analyze millions of records.

In the next five years, the Council for Economic and Development Affairs in Saudi Arabia will be leading significant economic changes in Saudi Arabia by gradually eliminating most services offered by the government or products that have a program of subsidies. For example, subsidies for water, electricity, and housing since 2016 will be decreased until it becomes eliminated totally. In addition, education and health care sector will become private sector. Also, tax programs will be added progressively in the next five years (The Economist, 2016). Lastly, these changes mentioned above require credible sources of statistical data from an authoritative agency for providing information successfully in order to implement the desired economic results.

The Central Department of Statistics & Information in Saudi Arabia, which became the General Authority for Statistics (GaStat) at the beginning of 2016, is responsible for collecting information about the Kingdom, such as the economic situation of its people, household information, and specific population facts. The GaStat was established to gather information through surveys, censuses, and third-party records to design the datasets in order of the Kingdom's leaders use statistical data to implement the changes in Saudi Arabia (GaStat, 2016). Moreover, the GaStat is expected to develop at all levels of internal sections to achieve the vision of the Saudi government and provide new products to supply users with different data categories by designing a data warehouse described as an Information Bank (GaStat, 2016).

However, current datasets in the GaStat are not able to support the goal of feeding the Information Bank because they are based on limited topics in the community statistics; for example, age groups, gender, and disability are not included aspects in the current surveys. Moreover, the majority of existing surveys follow a time format based on quarters or annually. Similarly, the current geographic entities included eight units, which are the country as a whole, and its provinces, governorates, urban areas, urban clusters, neighborhoods, neighborhood sectors, and city blocks. These geographic entities are not entirely reliable in current datasets because they exist through of the country, provinces, and governorates. Although the other geographic entities are available, they are designed to use for internal purposes and not available to the public sector.

Therefore, the GaStat needs to develop a new geodatabase conceptual model to implement the changes in Saudi Arabia as well as to respond to the fast growing of the statistical data. In this study, a new model is proposed by adopting the methods used by the United States Census Bureau, a common data source in the USA for government, universities research centers, and public users. The geographic hierarchy of Census Bureau has excellent geographic references because it contains a diversity of geographic entities. The GaStat needs to develop statistical surveys for interest areas such as community survey and housing survey, which include a variety of topics. Additionally, data should be updated frequently and be available to users at the different levels of geographic entities.

1.2 Purpose of Research

The purpose of this study is to develop a new geodatabase conceptual model for the GaStat in order to implement the changes in Saudi Arabia such as economic and social changes as well as to respond to fast-growing statistical data. This study also designs a theoretical hierarchy structure for the geographic entities in order to provide users with spatial and non-spatial data at different levels of geographic areas. Understanding the basic concepts in surveys and GIS is important in order to design such systems of statistical surveys and geographic entities.

1.3 Significance of the Study

The General Authority for Statistics (GaStat), which is under the Ministry of Economy and Planning in Saudi Arabia, is the authoritative agency of collecting information and designing the datasets currently used and is responsible for providing valuable data concerning community activities. The Ministry of Economy and Planning is a part of the Council for Economic and Development Affairs. The statistical data and geographic entities are main components in the GaStat, Unfortunately, the current statistical data and geographic entities are limited and are not able to support the vision of the government in economic changes.

In fact, the Saudi Arabian government is undergoing economic changes in order to prevent the instabilities because of in oil price instabilities affect the primary income for the government budget. While making the necessary changes, the Saudi Arabian government needs information regarding economic and community circumstances in order to plan and prepare a better future with sustainable sources. Also, other users are using new software and methods. For example, business analytics uses Excel and Power Map to integrate the statistical data with geographic entities through visualization.

Finally, this study will aid in the statistical data integration with legal and administrative geographic entities and statistical entities, which would achieve the vision of the government in economic and social changes and investment chances to improve future Saudi Arabia.

1.4 Definitions of Terms

> Statistical surveys

It is an ongoing assessment that provides essential information on a short term about a country, and its economy and people by using a method of samples based on a small percentage of the population (Lynn, 2008).

> Geographic entity

The geographic entity is also known as a geographic subdivision or geographic level. It is a feature that occupies a position in space with data describing the attributes and location of recorded information. Furthermore, the geographic entity is not included parcels information if it is commercial, industrial, agricultural or residential (The Pcmag, Encyclopedia; Census Bureau 2010).

> Spatial Data

It is the information about a physical object representing location, size, and shape of an area by numerical values in a geographic coordinate system such as cities and towns (Esri, 2010).

> GeoID

The GeoID is a coding method that links both non-spatial data and spatial data. The GeoID must be available on both sides of non-spatial data and spatial data in order to join them. (Census Bureau; Esri, 2010).

CHAPTER 2

LITERATURE REVIEW

2.1 The Saudi Arabia General Authority for Statistics

Kingdom of Saudi Arabia is a member of the Gulf Cooperation Council (GCC) with estimated population over 30,000,000 in the last census conducted in 2014. The Kingdom population density is a fifteen people per square kilometer, and it covers an area of 2,149,690 square kilometers (GaStat, 2016). In addition, the countries that are members of the Gulf Cooperation Council (GCC) as well members of the Gulf Statistical Center (GSC) such as the General Authority for Statistics (GaStat), Saudi Arabia. The Central Department of Statistics & Information (CDSI), which became the General Authority for Statistics (GaStat) in 2016, provides statistical data since the 70's. The methods used for recording datasets are decennial census, surveys and third-party records (CDSI, 2013). In addition, the GaStat also uses a geographic information system (GIS) to provide spatial data through geographic entities. There are a total of eight units of geographic entities, including the country as a whole, and its provinces, governorates, urban areas, urban clusters, neighborhoods, neighborhood sectors, and city blocks. These geographic entities are organized in a hierarchy level as shown in Table 3-1. The first level is the country itself, and it is exhibited as the Kingdom of Saudi Arabia. The second level is the provinces and contained thirteen provinces are divided the Kingdom. Also, the provinces are numbered uniquely with a two-digit. The third level is the governorates, which are divided the provinces into 118 governorates and governorates are numbered uniquely with a four-digit.

Also, the first two-digit of the governorate number identifies the province. The fourth level is the urban areas, which is included cities or villages and urban areas are numbered uniquely with a three-digit. The fifth level is the urban clusters, which is used to divide urban areas. Also, urban clusters are numbered uniquely with a two-digit number. The sixth level is the neighborhoods, which is divided urban areas and urban clusters. Also, neighborhoods are numbered uniquely with a three-digit designation. The seventh level is the neighborhood sectors, which is divided the neighborhood into sectors, also neighborhood sectors are numbered uniquely with a two-digit number. Finally, the eighth level is the city blocks, which nest within neighborhoods and neighborhood sectors, and city blocks are small in area. Also, city blocks are numbered uniquely with a three-digit designation. However, country, provinces, governorates levels are selected to use for publishing datasets to the public sector and remaining levels, which are mentioned above used for internal purposes. (CDSI, 2010).

2.2 The Current System of Geographic Entities for the General Authority for Statistics

The current system of geographic entities is inefficient in response to the programmatic or analytical needs of data users. For instance, Michael Bauer Research GmbH uses the current legal and administrative geographic entities, but the statistical entities are not released to users (see Table 2-1). The current statistical geographic entities utilize the urban area units to identify the composition and statistical boundary units. One of the issues regarding using the urban area units is personal information, which can be insecure and affect the people who provide information to the Saudi Arabia General Authority for Statistics (GaStat).

For example, neighborhoods are used as a geographic entity, which includes the names, locations, and boundaries of the neighborhoods and it can be identified easily and determine the people who are living in it. For this reason, the current system of geographic entities is not released to the public sector to avoid this problem. Furthermore, the current statistical geographic entities need to include new units that already have boundaries such as economic zones, school districts, and ZIP code areas to meet the needs of data presentations.

2.3 The system of Geographic Entities

The system of geographic entities is used in over 130 countries around the world to design census or statistical units and boundaries geodatabases (Ersi,2014). Although the system of geographic entities is designed different from a country to another, it is usually classified into legal and administration entities, and then further into statistics entities and irrespective of their relative density importance. For example, In United States, United Kingdom, Australia, Canada and New Zealand, the system of geographic entities is classified into legal and administration entities, and statistics entities. Moreover, blocks are often used as the basic geographic entity in many counties, such as the USA and the Saudi Arabia (CDSI, 2010; Census Bureau 2010).

2.4 The United States Census Bureau

The United States Census Bureau is the responsible department to collect data by using decennial census, statistical surveys, and third-party records. The datasets are available in different periods, such as decennial census, one year, three years, and five years. There are a total of 18 census geographic entities and they are organized in a hierarchy level as shown in Figure 2-1 (Census Bureau, 2010).

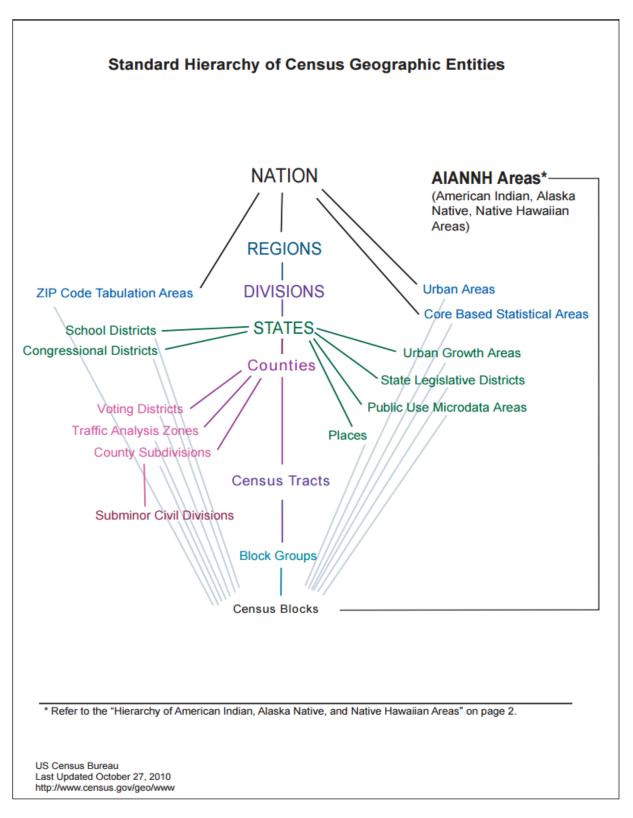


Figure 2-1. Standard hierarchy of census geographic entities and the second page is not recommended to discuss in this project because it is about race in the U.S. (Census Bureau, 2010).

> Nation Geographic Entity

The nation entity is the geographic level that has all of the entities under it because the nation entity covers the boundary of the United States of America and any lower level entities cannot cross it (Census Bureau, 2010).

> Regions Geographic Entity

The regions are the second geographic level, and classified into four types: Northeast, Midwest, South and West regions. (Census Bureau, 2010).

Divisions Geographic Entity

The third level is called divisions, which is divided into New England Division, Middle Atlantic Division, East North Central Division, West North Central Division, South Atlantic Division, East South Central Division, West South Central Division, Mountain Division, and Pacific Division. Also, two or more of the division entity come under one region (Census Bureau, 2010).

> States Geographic Entity

States are the fourth geographic level, and they classify the data with every state boundary, which is presented in all states in the USA. (Census Bureau, 2010).

Counties Geographic Entity

The states in the United States are divided by counties as a primary legal boundary. There are a total of 3,144 counties in the USA. Texas has 254 counties the maximum number of counties in the country, and Delaware has three counties with the minimum number of counties in the country (Census Bureau, 2010).

Census Tracts Geographic Entity

Census Tracts are geographic level designed for statistical subdivisions of the counties. Census tracts contain 1,200 to 8,000 residents, and the boundaries must not cross county or state boundaries. There are a total of 73,057 Census Tracts in the USA (Census Bureau, 2010).

▶ Block Groups Geographic Entity

After census tracts are used as statistical subdivisions of a county, Block groups are utilized to be a statistical division of census tracts. The block groups contain a minimum of 600 residents or 240 housing units and a maximum of 3,000 residents or 1,200 housing units. There are a total of 217,740 block group divisions in the USA (Census Bureau, 2010).

> Census Blocks Geographic Entity

Census Blocks are the lowest statistical areas of collecting the information. There are a total of 11,078,297 census blocks in the USA (Census Bureau, 2010).

> Urban Areas Geographic Entity

Urban Areas classified urbanized areas and urban clusters than the rural areas in the USA. Urban Areas is represented the densely developed area include residential, commercial, and other urban land uses. Urban Areas are the places that have high population density, and there are a total of 3,573 Urban Areas in the USA (Census Bureau, 2010).

➤ Core Based Statistical Areas Geographic Entity

The Core Based Statistical Areas consists of equivalents or counties entities associated with a minimum of one core and core area holding a large population center, together with nearby communities having a high degree of social and economic integration with that same core (Census Bureau, 2010).

> ZIP Code Tabulation Areas Geographic Entity

The ZIP Code designation is used and designed by the United States Postal Service (USPS) ZIP Code service areas, and it uses for mail delivery routes (Census Bureau, 2010).

> Urban Growth Areas Geographic Entity

This is an index of urban area growth that is updated yearly (Census Bureau, 2010).

> State Legislative Districts Geographic Entity

These are selected areas from which members are elected to be in state legislatures, and is used for small-scale thematic mapping for presenting to state legislative districts (Census Bureau, 2010).

Public Use Microdata Areas Geographic Entity

These are selected areas used as a sample method based on a small percentage of the population (Census Bureau, 2010).

> Places Geographic Entity

This is the latest updating of boundaries selected by the respective states for important places like city, town, village, or borough, and it selected to provide services or administer an area without necessarily looking to the population size (Census Bureau, 2010).

> School Districts Geographic Entity

This entity consists of the boundaries of school districts that provide by public educational services, Elementary, high school, secondary and Unified (Census Bureau, 2010).

> Congressional Districts Geographic Entity

There are a total of 435 areas of congressional districts that comprise the House of Representatives in the USA (Census Bureau, 2010).

> Voting Districts Geographic Entity

After every census (like 2010) state governments have to reestablish the new boundaries for voting districts based on population. (Census Bureau, 2010).

> Traffic Analysis Zones Geographic Entity

This is used to present the place where there are high traffic zones (Census Bureau, 2010).

> County Subdivisions Geographic Entity

This includes county divisions, census subareas, unorganized territories and minor civil divisions and it is a primary division of the counties and equivalent (Census Bureau, 2010).

2.5 GeoID

An instance of GeoID used by the United States Census Bureau includes:

- 01 The United States
- 48 Texas State
- 355 Nueces County
- 0018.02 Census Tract
- 18021 Census block Group1
- 1030 Block

An example of GeoID used by the Saudi Arabia General Authority for Statistics:

- 1 Saudi Arabia
- 01 Al-Riyadh province
- 00 Al-Riyadh governorate
- 001 Urban Area
- 01 Urban Cluster
- 001 Neighborhood Sector

The GeoID code employed by the United States Census Bureau uses Federal Information Processing Standards (FIPS) to identify legal geographic entities, but the GeoID for statistical geographic entities code is designed by the Census Bureau. On the other hand, the Saudi Arabia General Authority for Statistics has developed GeoID code for legal and administrative entities, and statistical entities. (Census Bureau, 2010; CDSI, 2010).

2.6 Comparison of Datasets in the US Census Bureau and Datasets in the Saudi Arabia General Authority for Statistics

According to the download center of the Fact Finder, which is under the United States Census Bureau, the US Census Bureau provides more than 51,501 tabulate statistical data for public sector to download. The data includes for example, Summary File1, Summary File2, Summary File3, Summary File4 and the ACS (American Community Survey) data. Summary File1 (SF1) and Summary File2 (SF2) include topics on race, Hispanic/Latino origin, households, families, age, sex, housing units, and owner/renter status. Also, Summary File3 (SF3) and Summary File4 (SF4) include topics on foreign birth, commuting, occupation, ancestry, disability, income, education, household financial arrangements, year housing structure built and many other populations and housing subjects (Census Bureau, 2016). American Community Survey is a program that collects information about the community like education, age, and gender and follows time formats based on one, three, and five years. There are 130 programs in the US Census Bureau of surveys and censuses that collecting information about the USA. However, 19 programs are published to the public sector, such as the American Community Survey by one, three, and five-year estimates. Nevertheless, the three-year data estimate from American Community Survey was eliminated in 2015 because of difficulty in the budget (Census Bureau, 2016). On the other hand, there are approximately of 2,000 tabulate statistical data in the Saudi Arabia General Authority for Statistics, which are published to the public sector since 2000 to 2015 by quarters or one-year estimate (GaStat, 2016).

2.7 Michael Bauer Research GMBH (MB-Research)

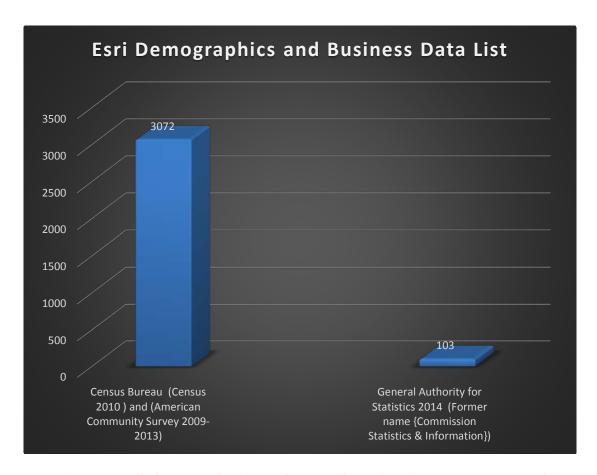


Figure 2-2. Summary of both statistical data lists (attributes) used by Esri from the USA, Census Bureau and the KSA, General Authority for Statistics (Esri, 2016).

Michael Bauer Research GmbH Company (MB-research) offers latest international market data and digital boundaries, and it is the official partner of the Environmental Systems Research Institute (Esri), a leading GIS company. MB-research is responsible for providing demographics and business datasets from around the world to Esri.

Esri provides Geographic Information System software as a first international supplier, geodatabase management applications, and web GIS. MB-research has provided statistical data for Saudi Arabia, United States and others 134 countries (Ersi; MB-research, 2016). MB-research is a client for both the United States Census Bureau and the Saudi Arabia General Authority for Statistics (GaStat). An analysis of statistical data and geographic entities that are available for the Saudi Arabia GaStat and the United States Census Bureau has provided a picture of what needs to improve on the GaStat systems (see Figures 2-3A, 2-3B, 2-4A, 2-4B and 2-5). The US Census Bureau has designed the datasets with different kinds of programs methods such as decennial census and the American Community Survey. On the other hand, the GaStat only provides decennial census, but statistical surveys such as Saudi Community Survey have not been used in a current system. Although there are specialized surveys available for the community, they are limited. For example, employment, income, and spending are the current specialized surveys available for community data. However, age group, gender, and education are not the available topics in the current surveys. They are only available in the decennial census. MB-research takes the advantage of data design in the US Census Bureau to deliver more than 3,000 attributes to Esri. Basically, it is showing a high level of performance by the US Census Bureau if it links to what MB-research could have from General Authority for Statistics which are 103 attributes (see Figure 2-2). Statistical surveys and excellent geographic references lead to make MB-research deliverable more data from the United States than the Saudi Arabia GaStat. The community can have access to all they share in order back to provide better services in the future. Esri uses statistical data to help the new businesses, schools and universities and this is growing and expanding.



Census 2010

Census 2010 Population by Age and Sex

(68 Attributes) Example: 2010 Median Male Age

2010 Median Female Age

Census 2010 Population by Single Year Age Increments (< 22 Years) and Sex

(66 Attributes) Example

2010 Male Population: Age 9 2010 Female Population: Age 9

Census 2010 Population by Race

(68 Attributes) Example:

2010 Population of 1 Race: White 2010 Population of 1 Race: Asian

Census 2010 Hispanic Population by Race

(72 Attributes)

Example:

2010 Hispanic Population of 1 Race: White 2010 Hispanic Population of 2 Races: White-Pacific

Census 2010 White Population by Age and Sex

(126 Attributes) Example:

2010 Median White Age 2010 Median White Male Age 2010 Median White Female Age

Census 2010 Black/African American Population by Age and Sex

(63 Attributes) Example: 2010 Median Black Age 2010 Median Black Male Age 2010 Median Black Female Age

Census 2010 American Indian/Alaska Native Population by Age and Sex

(63 Attributes)

Example:

2010 Median American Indian/Alaska Native Age 2010 Median American Indian/Alaska Native Male

2010 Median American Indian/Alaska Native

Census 2010 Asian Population by Age and Sex

(63 Attributes) Example:

2010 Median Asian Age 2010 Median Asian Male Age 2010 Median Asian Female Age

Census 2010 Pacific Islander Population by Age and Sex

(63 Attributes) Example: 2010 Median Pacific Islander Age 2010 Median Pacific Islander Male Age 2010 Median Pacific Islander Female Age

Census 2010 Other Race Population by Age and Sex

(63 Attributes) Example:

2010 Median Other Race Age 2010 Median Other Race Male Age 2010 Median Other Race Female Age

Census 2010 Multiple Races Population by Age and Sex

(63 Attributes)

Example:

2010 Median Multiple Races Age 2010 Median Multiple Races Male Age 2010 Median Multiple Races Female Age

Census 2010 Hispanic Population by Age and Sex

(63 Attributes) Example: 2010 Median Hispanic Age 2010 Median Hispanic Male Age 2010 Median Hispanic Female Age

Census 2010 White Non-Hispanic Population by Age and Sex

(40 Attributes)

Example

2010 White Non-Hispanic Male Population 2010 White Non-Hispanic Female Population

Census 2010 Population in Households by Race/Hispanic Origin

(19 Attributes)

Example:

2010 Population in Households 2010 Asian Population in Households 2010 Hispanic Population in Households

Census 2010 Population in Families by Race/ Hispanic Origin

(10 Attributes)

Example:

2010 Population in Families

2010 Population in Families w/Asian Householder

Census 2010 Population in Households by Type

(8 Attributes)

2010 Population in Husband-wife Family Households 2010 Total Population in Nonfamily Households

Census 2010 Population in Households by Relationship

(18 Attributes)

Example:

2010 Stepchild in Family Households 2010 Parent in Family Households

Census 2010 Population Age <18 Detail

Example:

2010 Population < 18 in Households: Own Child 2010 Population < 18 in Households: Grandchild

Census 2010 Population Age 65+ Detail

(27 Attributes)

Example

2010 Population 65+ in Family Households 2010 Population 65+ in Nonfamily Households

Census 2010 Group Quarters Population

(10 Attributes)

Example:

2010 Institutionalized Population in Group Quarters 2010 GQ Institutionalized Population: Other Institutional

Census 2010 Population in Households by Tenure Detail

(22 Attributes)

Example:

2010 Population in Occupied HUs: Owned Free and Clear 2010 Population in Renter-occupied Housing Units

Census 2010 Households by Type

(18 Attributes)

Example:

2010 Total Family Households 2010 Total Nonfamily Households

Census 2010 Households by Race/Hispanic Origin of Householder

(32 Attributes)

Example:

2010 Households w/White Householder 2010 Households w/Asian Householder



Census 2010

Census 2010 Family Households by Race/ Hispanic Origin of Householder

(18 Attributes)

Example:

2010 Family Households w/White Householder 2010 Family Households w/Asian Householder

Census 2010 Husband-wife Family Households by Race/Hispanic Origin of Householder

(10 Attributes)

Example:

2010 Husband-wife Family Households w/White Householder

2010 Husband-wife Family Households w/Asian Householder

Census 2010 Other Family Households by Race/Hispanic Origin of Householder

(10 Attributes)

Example:

2010 Other Family Households w/White Householder 2010 Other Family Households w/Asian Householder

Census 2010 Nonfamily Households by Race/Hispanic Origin of Householder

(10 Attributes)

Example:

2010 Nonfamily Households w/White Householder 2010 Nonfamily Households w/Asian Householder

Census 2010 Family Households by Size

(7 Attributes)

Example:

2010 Family Households: 2-Person 2010 Family Households: 7+ Person

Census 2010 Nonfamily Households by Size

(8 Attributes)

Example:

2010 Nonfamily Households: 1-Person 2010 Nonfamily Households: 7+ Person

Census 2010 Nonfamily Households by Size

(8 Attributes)

Example:

2010 Nonfamily Households: 1-Person 2010 Nonfamily Households: 7+ Person

Census 2010 Family Households by Age of Householder

(9 Attributes)

Example:

2010 Family Households w/Householder 15-24 2010 Family Households w/Householder 35-44

Census 2010 Nonfamily Households by Age of Householder

(9 Attributes)

Example:

2010 Nonfamily Households w/Householder 15-24 2010 Nonfamily Households w/Householder 35-44

Census 2010 Households by Type and Presence of Children

(70 Attributes)

Example:

2010 Husband-wife Families: Own Children <18 2010 Other Family HHs: No Own Children <18 2010 Family HHs w/Pacific Islander HHr: Own Children

Census 2010 Households by Type and Presence of Population Age 65+

(6 Attributes)

Example:

2010 Households with Population 65+: 1-Person Nonfamily HHs

2010 Households with No Population 65+: Family House holds

Census 2010 Households by Tenure and Household Type

(9 Attributes)

Example:

2010 Occupied Housing Units: Owned Free and Clear 2010 Renter-occupied Housing Units

Census 2010 Owner-occupied Housing Units by Race/Hispanic Origin

(9 Attributes)

Example:

2010 Owner-occupied Housing Units w/White

Householder

2010 Owner-occupied Housing Units w/Asian House holder

(9 Attributes) Example:

Race/Hispanic Origin

2010 Renter-occupied Housing Units w/White Householder 2010 Renter-occupied Housing Units w/Asian Householder

Census 2010 Renter-occupied Housing Units by

Census 2010 Owner-occupied Housing Units by Size

(8 Attributes)

Example:

2010 Owner-occupied Housing Units: 1-Person 2010 Owner-occupied Housing Units: 7+ Person

Census 2010 Renter-occupied Housing Units by

(8 Attributes)

Example:

2010 Renter-occupied Housing Units: 1-Person 2010 Renter-occupied Housing Units: 7+ Person

Census 2010 Owner-occupied Housing Units by Age of Householder

(8 Attributes)

Example:

2010 Owner-occupied Housing Units w/Householder 15-24 2010 Owner-occupied Housing Units w/Householder 35-44

Census 2010 Renter-occupied Housing Units by Age of Householder

(8 Attributes)

Example:

2010 Renter-occupied Housing Units w/Householder 15-24 2010 Renter-occupied Housing Units w/Householder 35-44

Census 2010 Housing

(10 Attributes)

Example:

2010 Vacant Housing Units

2010 Vacant Housing Units: For Rent

2000 Data in 2010 Geography

(8 Attributes)

Example:

2000 Population in Families 2000 Total Family Households

Figure 2-3B. Esri has selected 43 tables with 1,274 attributes from the United States Census 2010 manipulated and sorted by Michael Bauer Research (MB-research), (Esri, 2016).



American Community Survey (ACS)

2009-2013 ACS Key Demographic Indicators

(9 Attributes)

Example:

2009-2013 ACS Total Population 2009-2013 ACS Total Households

2009-2013 ACS Population by Marital Status

(15 Attributes)

Example:

2009-2013 ACS Population Age 15 or Older: Never

Married

2009-2013 ACS Population Age 15 or Older:

Widowed

2009-2013 ACS Population by School Enrollment

(72Attributes)

Example:

2009-2013 ACS Population Age 3 or Older Enrolled in School

2009-2013 ACS Population Age 3 or Older Enrolled in Kindergarten

2009-2013 ACS Population by Educational Attainment

(45 Attributes)

Example:

2009-2013 ACS Pop 25+ by Educ Attainment: No

Schooling

2009-2013 ACS Pop 25+ by Educ Attainment:

Kindergarten

2009-2013 ACS Population by Language Spoken at Home

(156 Attributes)

Example:

2009-2013 ACS Population Age 5-17 who speak only

English

2009-2013 ACS Reliability Population Age 5-17 who

speak Spanish

2009-2013 ACS Population by Industry and Occupation

(129 Attributes)

Example:

2009-2013 ACS Industry: Employed in Construction 2009-2013 ACS Industry: Employed in Wholesale

Trade

2009-2013 ACS Females by Age of Own Children and Employment Status

(39 Attributes)

Example:

2009-2013 ACS Females Age 20-64 with Own Children Age less than 6 Only

2009-2013 ACS Females Age 20-64 with Own Children Age 6-17

2009-2013 ACS Place of Work

(12 Attributes)

Example:

2009-2013 ACS Workers Age 16 or Older 2009-2013 ACS Reliability Workers Age 16 or Older

2009-2013 ACS Means of Transportation to Work

(42 Attributes)

Example:

2009-2013 ACS Workers Age 16 or Older: Drove Alone to

2009-2013 ACS Workers Age 16 or Older: Bus/Trolley Bus to

2009-2013 ACS Travel Time to Work

(42 Attributes)

Example:

2009-2013 ACS Workers Age 16 or Older by Travel Time to Work Base

2009-2013 ACS Workers Age 16 or Older by Travel Time to Work: 5-9 minutes

2009-2013 ACS Means of Transportation to Work by Sex (Available at the Tract Level and Above)

(48 Attributes)

Example:

2009-2013 ACS Male Workers Age 16 or Older 2009-2013 ACS Male Workers Age 16 or Older: Bicycle to Work

2009-2013 ACS Workers by Sex and Vehicles Available (Available at the Tract Level and Above)

(21 Attributes)

Example

2009-2013 ACS Workers Age 16 or Older in Households 2009-2013 ACS Male Workers Age 16 or Older in

Households: 1 Vehicle Available

2009-2013 ACS Population by Ratio of Income to Poverty

(24 Attributes)

Example:

2009-2013 ACS Population by Ratio of Income to

Poverty Base

2009-2013 ACS Population by Ratio of Income to Poverty: 1.00-1.24

2009-2013 ACS Population by Poverty Status and Educational Attainment (Available at the Tract Level and Above)

(57 Attributes)

Example:

2009-2013 ACS Population Age 25 or Older for whom Poverty Status is Determined

2009-2013 ACS Population Age 25 or Older with Income Below Poverty Level

2009-2013 ACS Population by Poverty Status and Labor Force Status (Available at the Tract Level and Above)

(57 Attributes)

Example:

2009-2013 ACS Civilian Population Age 16 or Older for whom Poverty Status is Determined 2009-2013 ACS Civilian Population Age 16 or Older with Income Below Poverty Level

2009-2013 ACS Population by Poverty Status and Race (Available at the Tract Level and Above)

(81 Attributes)

Example:

2009-2013 ACS White Population for whom Poverty Status is Determined

2009-2013 ACS White Population with Income at/

2009-2013 ACS Population by Veteran Status

(26 Attributes)

Example:

2009-2013 ACS Civilian Population Age 18 or Older 2009-2013 ACS Civilian Population Age 18 or Older:

Figure 2-4A. Esri has selected 34 tables with 1,274 attributes from the American Community Survey (ACS) manipulated and sorted by Michael Bauer Research (MB-research), (Esri, 2016).



American Community Survey (ACS)

2009-2013 ACS Population by Period of Military Service

(45 Attributes)

Example:

2009-2013 ACS Civilian Veterans Age 18 or Older: Served in Gulf War (9/2001 or Later) but not Vietnam 2009-2013 ACS Civilian Veterans Age 18 or Older: Served in Gulf War (8/1990-8/01 & Later) and Vietnam

2009-2013 ACS Population by Types of Health Insurance Coverage by Age

(120 Attributes)

Example:

ACS 2009-2013 Civilian Noninstitutionalized Population ACS 2009-2013 Civilian Noninstitutionalized Population

2009-2013 ACS Householder Age 65 or Older

(57 Attributes)

Example

2009-2013 ACS Household Income and Householder Age 65 or Older Base

2009-2013 ACS Household Income \$35,000-\$39,999 and Householder Age 65 or Older

2009-2013 ACS Households by Poverty Status

(39 Attributes)

Example:

2009-2013 ACS Total Households

2009-2013 ACS Households with Income at/Above

Poverty Level

2009-2013 ACS Households by Poverty Status and Receipt of Food Stamps/SNAP (Available at the Tract Level and Above)

(42 Attributes)

Example:

2009-2013 ACS Household Received Food Stamps/

2009-2013 ACS Household Received Food Stamps/

SNAP: Income at/Above Poverty Level

2009-2013 ACS Home Value

(81 Attributes)

2009-2013 ACS Owner Households

2009-2013 ACS Home Value less than \$10,000

2009-2013 ACS Households by Mortgage Status

(27 Attributes)

Example:

2009-2013 ACS Owner Households

2009-2013 ACS Reliability Owner Households

2009-2013 ACS Households by Rent

(84 Attributes)

Example

ACS 2009-2013 Renter Households

2009-2013 ACS Renter Households Paying Contract

Rent: \$100-\$149

2009-2013 ACS Households by Year Householder Moved into Unit

(42 Attributes)

Example:

2009-2013 ACS Owner Households by Year Householder Moved In: 2010 or Later 2009-2013 ACS Owner Households by Year Householder Moved In: 2000 to 2009

2009-2013 ACS Households by Heating Fuel

(30 Attributes)

Example

2009-2013 ACS Households by House Heating Fuel: Utility Gas

2009-2013 ACS Households by House Heating Fuel: Electricity

2009-2013 ACS Households by Vehicles Available

(42 Attributes)

Example:

2009-2013 ACS Owner Households by Vehicles

2009-2013 ACS Owner Households by Vehicles

2009-2013 ACS Housing by Units in Structure

(33 Attributes)

Example:

2009-2013 ACS Housing Units by Units in Structure: 1-

Detached

2009-2013 ACS Housing Units by Units in Structure: 1-

Attached

2009-2013 ACS Housing by Year Structure Built

(32 Attributes)

Example:

2009-2013 ACS Housing Units by Year Structure Built:

2000-2009

2009-2013 ACS Housing Units by Year Structure Built:

1980-1989

2009-2013 ACS Income

(78 Attributes)

2009-2013 ACS Total Households

2009-2013 ACS Household Income \$10,000-\$14,999

2009-2013 ACS Householder Age 15-24

(57 Attributes)

Example:

2009-2013 ACS Household Income and Householder

Age 15-24 Base

2009-2013 ACS Household Income \$10,000-\$14,999 and Householder Age 15-24

2009-2013 ACS Householder Age 25-44

(57 Attributes)

Example:

2009-2013 ACS Household Income and Householder Age 25-44 Base

2009-2013 ACS Household Income \$10,000-\$14,999 and Householder Age 25-44

2009-2013 ACS Householder Age 45-64

(57 Attributes)

Example

2009-2013 ACS Household Income and Householder

Age 45-64 Base 2009-2013 ACS Household Income \$75,000-\$99,999

and Householder Age 45-6



Population and Community (General Authority for Statistics(GaStat))

GaStat-2014 Population

(4 Attributes) Example: 2014 Total Female Population (MBR) 2014 Total Male Population (MBR)

GaStat-2014 Households

(2 Attributes) Example: 2014 Average Household Size (MBR) 2014 Total Households (MBR)

GaStat-2014 Education

(9 Attributes) Example:

2014 Pop 10+/Education: Primary (MBR)

2014 Pop 10+/Education: Secondary/Equivalent (MBR)

GaStat-2014 Marital Status

(4 Attributes) Example: 2014 Marital Status: Divorced (MBR) 2014 Marital Status: Married (MBR)

GaStat-2014 Spending

(84 Attributes) Example:

2014 Food & non-alcoholic beverages: Total (MBR)

2014 Footwear: Total (MBR)

2014 Glassware/tableware/utensils: Total (MBR)

Figure 2-5. Esri has selected 5 tables with 103 attributes from the General Authority for Statistics. Saudi Arabia manipulated and sorted by Michael Bauer Research (MB-research), (Esri, 2016).

MB-research has provided a clear picture that the United States Census Bureau provides a diversity of statistical data. Alternatively, the Saudi Arabia GaStat has limited topics about the community date, so Michael Bauer Research (MB-research) would not be able to have as many statistical data points as the US Census Bureau. As a consequence, the data used by Esri for the Saudi Arabia GaStat reflects that the Saudi Arabia GaStat needs to develop a new method to design statistical data, and eliminate the limitations in the current statistical data. (see Figure 2-6) (Esri, 2016).

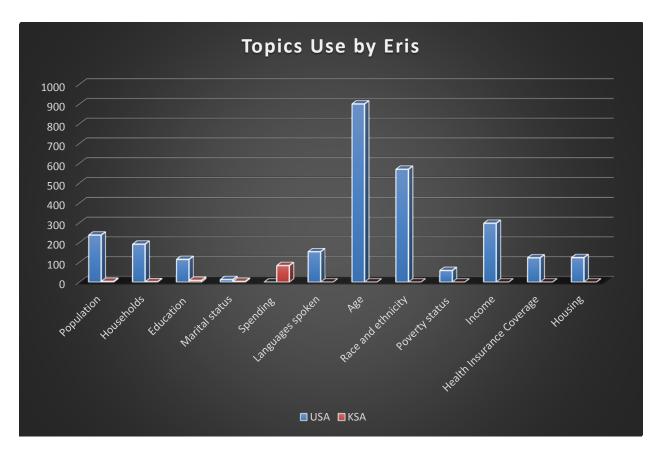


Figure 2-6. A total twelve topics used by Esri, including 11 topics provided by the US Census Bureau and five topics by the Saudi Arabia GaStat (Esri, 2016).

Esri has used twelve topics related to communities from the United States and the Saudi Arabia GaStat in order to design a geodatabase. The twelve topics are population, households, education, marital status, spending, languages spoken, age, race and ethnicity, poverty status, income, health insurance coverage and housing (see Figure 2-6). The US Census Bureau provides all topics except for spending. However, MB-research selected five topics from the Saudi Arabia GaSta because of data design issues (MB-research, 2016), which are population, households, education, marital status and spending. The Saudi Arabia GaStat has released statistical data through the country and its provinces and governorates geographic entities. In the end, Esri uses geographic entities from the United States and the Saudi Arabia GaStat to present statistical data (see Table 2-1 and 2-2).

Table 2-1. Geographic Entities of KSA

| Name | Count |
|---------------------------------|-------|
| Saudi Arabia | 1 |
| Region (official name Province) | 13 |
| Province (official name | 118 |
| Governorate) | |

(Esri, 2016)

Table 2-2. Geographic Entities of USA

| Name | Count |
|--------------------------------|---------|
| United States (US) | 1 |
| State (ST) | 51 |
| County (CY) | 3,143 |
| Census Tract (TR) | 73,057 |
| Block Group (BG) | 217,740 |
| Place (PL) | 29,261 |
| ZIP Code (ZP) | 31,719 |
| County | 35,703 |
| Subdivision (CS) | |
| CBSA (CB) | 917 |
| DMA (DM) | 210 |
| Congressional District (CD) | 436 |

(Esri, 2016)

As a result, developing new systems of statistical surveys and geographic entities for the General Authority for Statistics (GaStat) in Saudi Arabia are needed before the 2020 census to respond to the fast growing in the statistical data as well as to be more activated with users and related resources through providing a diversity of datasets.

CHAPTER 3

METHODOLOGIES

3.1 Introduction

The General Authority for Statistics of Saudi Arabia has been developed to satisfy the needs of users through providing more statistical data (GaStat, 2016). Indeed, The Information Bank is a data warehouse that will be released in the last quarter of 2016 by Saudi Arabia GaStat. However, it needs to have additional informational resources in order to provide statistical data that can be capable of designing a database. Also, it needs tools to specify and classify datasets by types, times and locations, the three main factors needed to use as a standard of designing a database for the Information Bank. The design used in this study includes statistical surveys and geographic entities in order to propose a new geodatabase model, which is expected to address the limitations of the current systems in the GaStat. First, types of information classify by statistical surveys. Also, a series of formatting time methods of quarters, one year, three years and five years use in the statistical systems in order to provide up-to-date information. Second, the system of geographic entities classifies into two groups: legal, administrative entities, and statistical entities based on their corresponding geographic subdivision and they organize in a hierarchical structure. Also, the unique reference numeric code utilizes to integrate between spatial and non-spatial data of the geographic entities. Finally, a new geodatabase conceptual model design to the Saudi Arabia GaStat by following the methods mentioned above.

3.2 Concepts of Statistical surveys

The statistical surveys need to design specific topics to obtain up-to-date information. Indeed, the current community data designed from the decennial census collected annually several topics from a sample target area or third party data such as employment, education, marital, income, and spending. Additionally, the decennial censuses are primary sources of data in the Saudi Arabia GaStat and it needs more resources such as statistical surveys. For example, Saudi Community Survey, Saudi Housing Survey and other sections need to have a specialized program of statistical survey. The statistical survey technique provides information on various topics as the same level of available topics in decennial censuses or more. The program management and systems engineering design types of information, questionnaire form and warehouse system to record the data. Also, designing a systemic method to integrate between statistical surveys and geographic entities to feed the Information Bank.

First, partnership agreements are needed in order to design statistical surveys. For example, National Information Center, Saudi Post, Ministry of Social Affairs and Ministry of Housing are partnerships for Saudi Community Survey and Saudi Housing Survey. The Partnership program obtain third-party data, which is a one of the information sources for statistical surveys. The second method is to use a self-response form that people can access online to provide important information for the statistical surveys. Advertisements, universities, partnership programs and other ways can be used to inform the people to respond to digital form. The digital form is a responsive web page that has multiple modes and can work on any device.

The third method is field surveys, which are major resources of gathering information. The field surveys can be Non-Interview or Interview to household units. Non-Interview uses the contact strategy to send the questionnaire form by mail or via a phone call. The people who use Saudi Post mailing address is the target of Non-Interview method. The interview method consists of visiting housing units to target other people that they do not have the mailing address services.

Moreover, the serial of time of quarters, one year, three years and five years, can be used to have up-to-date information. The reasons for the existence of this four period model is to provide statistical data given the fact-that the Kingdom of Saudi Arabia is a developing country and is evolving rapidly with increasing in the population as well as the needs of users (The Economist, 2016). Another reasons are very high youth percentage in Saudi Arabia population and to measure statistical and temporal frequency as well have a balance in geographic resolution is a challenge (GaStat, 2016). In the end, program management and system engineering have to integrate with a deep understanding of the concepts needed to design several programs of statistical surveys and improve the efficiency and effectiveness in the database before upcoming 2020 census. Statistical surveys are used to re-engineering field operations and record information more than any time before in the Saudi Arabia GaStat history.

3.3 Concepts of Geographic Entities

The system of geographic entities is classified into two categories depending on the design features: legal and administrative entities and statistical entities (see Table 3-1 and 3-2). These two categories provide statistical data to accommodate a variety of user-requests.

Table 3-1.
The Current System of Geographic Entities

| Geographic Entities | Status |
|-----------------------|-----------------------|
| Country | Legal |
| Provinces | Legal/ Administrative |
| Governorates | Legal/ Administrative |
| | |
| Urban Areas | Statistical |
| Urban Clusters | Statistical |
| Neighborhoods | Statistical |
| Neighborhood Sectors | Statistical |
| City Blocks | Statistical |

(GaStat, 2016)

Table 3-2.
The New System of Geographic Entities

| Geographic Entities | Status |
|---------------------------------|-----------------------|
| Country | Legal |
| Regions | Statistical |
| Provinces | Legal/ Administrative |
| Governorates | Legal/ Administrative |
| Statistical Tracts | Statistical |
| Statistical Block Groups | Statistical |
| Statistical Blocks | Statistical |
| Holy Areas | Legal |
| Urban Areas | Statistical |
| Economic Zones | Administrative |
| ZIP Code Areas | Administrative |
| Urban Growth Areas | Statistical |
| School Districts | Administrative |
| Places | Statistical |
| Sample-Data Areas | Statistical |
| Voting Districts | Legal/ Administrative |
| Governorate Subdivisions | Statistical |

➤ Legal and Administrative Geographic Entities

Legal and administrative entities usually proceed from legal actions, court decisions, statutes, treaties, resolutions, and ordinances. The boundaries of legal and administrative entities are used such as exist according to the related sources. For example, the boundaries of the provinces are obtained through the Ministry of Interior or General Authority of Survey (GAS), Saudi Arabia. Another example, the boundaries of the ZIP Code Areas are taken by Saudi Postal Corporation (Saudi Post) which is the government-operated postal service. In fact, government system determines from geographical entities: legal and administrative entities (see Table 3-1 and 3-2). However, these geographic entities need a second level, which are statistical entities to meet the needs of statistical data (Census Bureau, 2010).

> Statistical Geographic Entities

Statistical entities are designed in circumstances where legal areas are inconsistent over time, inadequate, incomplete, inconsistent, or nonexistent. Additionally, the boundaries used by statistical entities do not prescribe in contrast with legal and administrative entities. In general, statistical geographic entities are a result from usage, need, practice or custom, and census. Saudi Arabia GaStat develops the foundations and guidelines for their classification and delineation and creates statistical units to respond to the programmatic or analytical needs of users. Furthermore, cooperation with the state, local, and tribal officials is relevant for establishing a standard set of statistical entities. In doing so, it determines size, structure, and boundaries to provide statistical data (Census Bureau, 2010).

3.4 The New System of Geographic Entities for the General Authority for Statistics

The new system of geographic entities is integrated between legal and administrative entities and statistical entities. For instance, voting districts and holy areas are legal geographic entities included in the new system. Also, the new statistical units are used where legal entities are not available, which are statistical tracts, statistical block groups, statistical blocks units. Moreover, the new units of geographic entities work as basic connectedness and interdependence with the three legal and administrative units, which are country, provinces, and governorates (see Table 3-2).

In addition, the new units, which were mentioned previously earlier, can avoid the problem of personal information, to secure statistical data. The new units do not contain names, locations and boundaries of the neighborhoods. The three new units use a method to cover the entire of the country rather than cover only the urban areas as on the current system.

Besides, design size, structure, and boundaries for the new statistical units are determined after project of updating of housing units and population measurement, which it starts in first quarter of 2016 and will finish in the second quarter of 2016 (GaStat, 2016). Also, other entities are needed for recording development plans and the needs of users such as places and governorate subdivisions (see Table 3-2). Finally, the administrative units, such as economic zones and ZIP code areas, have been added in the new geographic entities system (see Table 3-2).

3.5 The Geographic Entities for the USA Census Bureau and the KSA General Authority for Statistics

> Similarities

The United States Census Bureau and The Saudi Arabia General Authority for Statistics (GaStat) use the same method to classify the geographic entities into legal and administrative entities and statistical entities, in the current systems. The legal and administrative main levels are available with three geographic entities in the US Census Bureau, which are the nation, states and counties while Saudi Arabia GaStat have three, which are the country, provinces, and governorates. Also, urban areas and urban clusters are available with the both the US Census Bureau and The Saudi Arabia GaStat.

Additionally, the new system of the Saudi Arabia GaStat includes the voting districts, ZIP code areas, school districts, statistical tracts, statistical block groups, statistical blocks regions, urban areas, urban growth areas, places, public use, sample-data areas and governorates which are similar to what is available in the US Census Bureau (see Figure 2-1 and Table 3-2). In conclusion, the similarity between the US Census Bureau and the new Saudi Arabia GaStat system of geographic entities existed because of using the same concept and the government systems.

Differences

The US Census Bureau and the Saudi Arabia GaStat have differences relative to the size of the land and government departments. The legal and administrative entities and statistical entities are different in the current design of geographic entities. The Saudi Arabia GaStat system of geographic entities is limited and does not contain all geographic levels that it is needed, in contrast to the US Census Bureau system, which is a varied system by including several geographic levels (see Figure 1 and Table 4).

On the other hand, the new system of geographic entities for Saudi Arabia GaStat has differences in two geographic entities categories. First, the legal and administrative entities do not contain the state legislative districts and congressional districts that are used in the US Census Bureau because they are related to the system used by the USA government. Also, Traffic Analysis Zones entity is not used in the new system for the Saudi Arabia GaStat because it is an administrative entity related to a government department which it is not utilized boundaries system for traffic analysis. Furthermore, holy areas and economic zones are used in the system of geographic entities for General Authority for Statistics but not in Census Bureau because they do not exist in the USA.

The United States does not use holy areas while in Saudi Arabia this is intrinsically related to geographic areas. Also, the economic zones entity is an economic system used to build industrial cities, and it is not utilized in the USA (see Figure 2-1 and Table 3-2). Finally, statistical entities in the Saudi Arabia GaStat do not include Core Based Statistical Areas such as in the US Census Bureau because the economical method used to design this entity is not employed in Saudi Arabia. In conclusion, the differences between the two models are principally related to the size of the land and the system used by the respective governments.

3.6 Concepts of Coding for Geographic Entities

The unique reference numeric code was used to integrate the data between spatial and non-spatial of the geographic entities (Census Bureau, 2010; CDSI, 2010). Furthermore, the new unique numeric code has utilized connectedness and interdependence for legal and administrative entities and statistical entities. The unique reference numeric code is used one, two, three, four and five digits depending on level of geographic entities. Finally, merging these digits to designing series codes in order to work as a geographic identifier.

CHAPTER 4

RESULTS

4.1 Finding Types of Statistical Surveys

Designing statistical surveys is expected to increase the performance of the Saudi Arabia General Authority for Statistics (GaStat) and provides new datasets. The suggestions to the statistical surveys are:

- Saudi Community Survey
- Saudi Housing Survey
- Saudi Income Survey
- Saudi Spending Survey

- Saudi Economic Survey
- Saudi Industry Survey
- Saudi Agricultural Survey
- Saudi Employment Survey

These statistical surveys are utilized to collect information to populate the Information Bank. For example, the Saudi Community Survey program contains the categories: population, population change, education, family, age, gender, disability, poverty and more. Statistical surveys provide varied data categories to support the needs of users such as local governments, profit or non-profit organizations researchers, and decision-makers. Also, it offers estimated data with formatting time using quarter years, one year, three years and five years to release it and measuring the impact of change as well as having timely information that is essential for users. Data users can take these advantages of statistical surveys in order to study, analyze and make a realistic decision. As of the end, the benefit of statistical surveys is having intelligent sources of information and constantly working with updating information.

4.2 Theoretical Geographic Entities of the Saudi Arabia General Authority for Statistics

The advanced system of geographic entities is used to locate statistical data. The new seventeen geographic entities have designed a theoretical standard of the Saudi Arabia General Authority for Statistics (GaStat), which includes legal and administrative entities and statistical entities (see Figure 4-1).

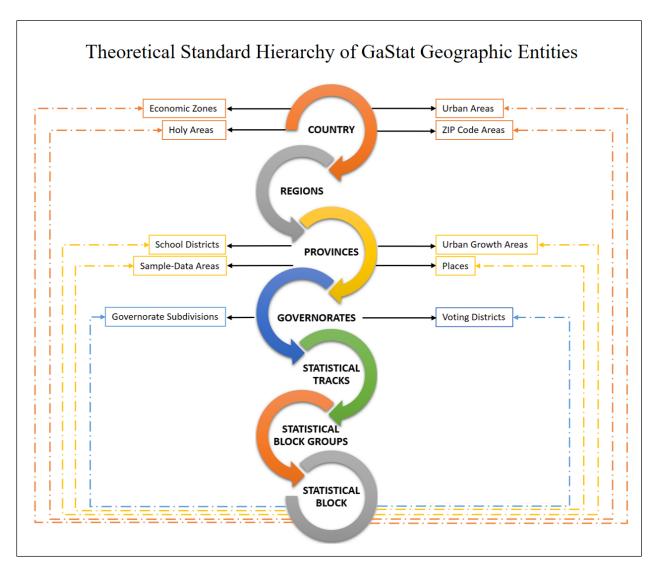


Figure 4-1. Theoretical Standard Hierarchy of General Authority for Statistics Geographic Entities

Country Geographic Entity

The geographic level of the country presents political boundaries between Saudi Arabia and nearby countries. Furthermore, all levels of geographic entities cannot cross-country geographic entity. The country boundaries can be obtained from the Saudi Arabia General Commission for Survey (GCS).

(see Appendix A-1)

> Regions Geographic Entity

This level corresponds to a grouping of provinces to five custom regions, which are East, West, North, Middle and South regions. First, East region has only one province, which is the Eastern province because it covers all the east of Saudi Arabia. Second, West region covers Makkah Al-Mokarramah and Al-Madinah Al-Monawarah provinces. Third, North region, includes four provinces, Northern Borders, Al-Jouf, Hail, and Tabouk. Fourth, Middle region covers Ar-Riyad and Al-Qaseem provinces. Fifth, South region includes four provinces Al-Baha, Aseer, Jazan and Najran. The regions boundaries can be obtained from the Saudi Arabia GaStat. (see Appendix A-1).

> Provinces Geographic Entity

The thirteen provinces in regions geographic entity of Saudi Arabia are divided legally and administratively, depending on connectedness and interdependence of local cultures and tribes. The provinces boundaries can be obtained from the Saudi Arabia GCS. (see Appendix A-1).

➤ Governorates Geographic Entity

There are 118 governorates in the thirteen provinces in Saudi Arabia, where Al-Riyadh province has twenty governorates with the maximum number of governorates, and both Al-Jouf and Northern Borders have three governorates with the minimum number of governorates. The governorates boundaries can be obtained from the Saudi Arabia GCS.

(see Appendix A-2 and A-3)

> Statistical Tracts Geographic Entity

Statistical tracts are statistical subdivisions of governorates entity and they cover the entire country. The purpose of statistical tracts is to provide a geographic unit that can present statistical data when legal and administrative entities are not available to support statistical data. The features used to determine the boundaries are major streets, natural and population size. The suggested size of the population for the statistical tracts entity can be determined after project of updating the general framework of population and housing paves the way for a new labor force survey, which was started in first quarter of 2016 and will be finished in the second quarter of 2016 (GaStat, 2016). The statistical tracts boundary cannot cross governorates, provinces, regions and country, but it may cross the boundaries of the other geographic entity. The statistical tracts boundaries can be obtained from the Saudi Arabia GaStat.

(See Appendix A-3 and A-4).

> Statistical Block Groups Geographic Entity

Block groups are statistical divisions and they cover the entire country. The boundaries are determined by coterminous with statistical tracts boundaries. Also, the boundaries of block groups cannot cross statistical tracts entity, but it might cross the boundaries of other geographic entities. It consists of clusters of blocks, and the suggested size of the population for the block groups can be determined by the current updating project of housing units and population measurement finished which was started in first quarter of 2016 and will be finished in second quarter of 2016 (GaStat, 2016). Block groups boundaries can be obtained from the Saudi Arabia GaStat. (see Appendix A-4).

> Statistical Blocks Geographic Entity

Statistical blocks are the basic statistical unit used for the Saudi Arabia GaStat of hundred-percent data collection from all housing units, rather than a sample of housing units, and it covers the entire country. The boundaries of statistical blocks design by following several different features. The first feature is typical boundaries such as streets, roads and creeks and statistical block may match those boundaries when it available. The second feature is the population and lands sizes and other visible features as specific place need. The statistical blocks can have some blocks with a zero-person. Block boundaries cab be obtained from the Saudi Arabia GaStat. (see Appendix A-4).

➤ Holy Areas Geographic Entity

The holy areas are places that are visited by Muslims from around the world. While Makkah province has the most areas, there are also holy places in Madinah province. The pilgrimage season, which is called Al-Hajj, is one of the resources of statistical data of the Saudi Arabia GaStat. Statistical data can determine the capacity on holy places in Al-Hajj season. From one year to another the Saudi government has built facilities and added many new services with new rules to manage the Al-Hajj season. The holy areas boundaries can be obtained from the Ministry of Hajj, Saudi Arabia.

> Urban Areas Geographic Entity

The high human population density and greater human-built features are called Urban Areas, and it can be cities, towns or conurbations. However, they are classified into two types, which are urbanized areas and urban clusters depending on the size of population. For example, urbanized areas contain a high human population density such as 50,000 people, but urban clusters contain a less human population density such as 5,000 people.

The first type is the urbanized areas which have two classes of urban areas with human-built features and urban areas with limited human-built features. The first class is a designation containing high human population density and greater human-built features. The second class contains high human population but has have limited human-built features. For example, water supply or sewage are not available in these urban areas.

The second type is the urban clusters, which also has two classes as similarly described above. The Urban Areas boundaries can be obtained from the Saudi Arabia GaStat.

(see Appendix A-5)

Economic Zones Geographic Entity

The Royal Commission for Jubail and Yanbu (RCJY) and other economic cities are not reported to the local government, but they connect directly with the Saudi Arabia Council of Ministers. It has different rules and lifestyle from local cities. RCJY and other economic cities need to have specific data to make a decision for better future of the Kingdom of Saudi Arabia. The country depends on the economies of its cities as the largest sources of income. The boundaries of statistical tracts, statistical block groups, and blocks design in these areas must not cross this entity, because of the independence of the systems of these cities. Also, the boundaries can be obtained from the surveying departments that are related to these cities.

(see Appendix A-5)

> ZIP Code Areas Geographic Entity

The Saudi Postal Corporation (Saudi Post) has divided ZIP service areas in the entire country. Saudi Post provides the boundaries of this entity, and it needs to use for e-commerce statistical data.

(see Appendix A-6)

> School Districts Geographic Entity

This is an administrative unit used for education officials to provide educational services. There are three level including elementary, intermediate and secondary. The boundaries can be obtained from the Ministry of Education, Kingdom of Saudi Arabia.

(see Appendix A-6)

> Urban Growth Areas Geographic Entity

This type of entity is designed by the Saudi Arabia GaStat to measure the growing in urban areas around the country. It is used to obtain data about population and human-built features. The General Authority for Statistics develops the actual boundaries.

(see Appendix A-7)

> Places Geographic Entity

Cities, towns, villages, boroughs, and other places such as commercial structures and wildlife sanctuaries can be under this entity.

(see Appendix A-7)

> Sample-Data Areas Entity

This is designed for internal exclusive use, and it does not publish to the public sector. The sampling method needs to select target areas in order to have sample data. The Sample-Data boundaries can be obtained from the Saudi Arabia GaStat.

> Voting Districts Geographic Entity

The municipal elections group the neighborhoods into the constituency and decide electoral seats.

The boundaries are obtained from the Ministry of Municipal and Rural Affairs of Saudi Arabia.

(see Appendix A-6)

➤ Governorates Subdivisions Geographic Entity

This entity divides the cities and the places depending on government administration offices. For example, the area in the south of Al-Riyadh city has the South Al-Riyadh education office. Governorates within the subdivision boundaries can be obtained from the Saudi Arabia GaStat. (see Appendix A-6)

4.3 Applied of Coding for Geographic Entities

Table 4-1. GeoID for the New System of Geographic Entities

| Geographic Entity | Description | |
|-----------------------------|---|--|
| Country | The country geographic entity uses one digit. | |
| Regions | The regions geographic entity uses one digit. | |
| Provinces | The provinces geographic entity uses two digits. | |
| Governorates | The governorates geographic entity uses two digits. | |
| ZIP Code Areas | The ZIP Code Areas geographic entity uses five digits | |
| Other Entities | The other entities use two digits before adding the digit for every entity. Holy areas, urban areas, economic zones, urban growth areas, school districts, places, sample-data areas, voting districts and governorates subdivisions. | |
| Statistical Tracts | The Statistical Tracts geographic entity uses four digits. | |
| Statistical Block Groups | The Statistical Block Groups geographic entity uses four digits. | |
| Statistical Blocks | The Statistical Blocks geographic entity uses four digits. | |

Designing geographic entities are needed to integrate spatial and non-spatial data through a common code. The number of digits are determined depending on the needs of geographic entities and how many elements it contains in every geographic entity. The country and regions use a unique number with one-digit. The provinces use two-digit, the numeric code starts by 01 and ends by 13. The governorates utilize a number with a four-digit, where the first two-digit are the same two-digit of provinces on that governorates are located. Then, for the statistical tracts, statistical block groups, and block groups, each uses a four-digit numeric code. If the four-digit on statistical tracts contain zeros this indicates areas of water. The statistical block groups use the first digit of statistical tracts code, and statistical blocks, following the same of block groups but use first two-digit of block groups code.

In addition, two-digit is used to merge between provinces, governorates, statistical tracts, statistical block groups, block groups and other geographic entities, which will be mentioned below to avoid duplicity. The holy areas and voting districts use two-digit, and economic zones have three-digits. The four-digit numeric code is used in urban areas, urban growth areas, places, school districts, sample-data areas, and governorates subdivisions. The ZIP code areas use a five-digit numeric code (see Table 4). For instance, below is an example of GeoID that can be used for the Saudi Arabia GaStat:

- 1 Saudi Arabia
- 2 West Region
- 02 Makkah Al-Mokarramah Province
- 0202 Makkah Governorate
- 02021111 Statistical Tract
- 020212222 Statistical Block Group
- 0202223333 Statistical Block

CHAPTER 5

CONCLUSIONS

The Saudi Arabian government is currently undergoing structural economic changes in order to prevent instabilities caused by worldwide oil price fluctuations, which affect the government's primary income. While making the necessary changes, the Saudi Arabian government needs better and more timely information regarding economic and community circumstances in order to plan and prepare for a better future with sustainable sources. The General Authority for Statistics (GaStat), which is under the Ministry of Economy and Planning in Saudi Arabia, is the authoritative agency of collecting information and designing the datasets currently used and is responsible for providing valuable data concerning community activities. The Ministry of Economy and Planning is a part of the Council for Economic and Development Affairs. The statistical data and geographic entities are main components in the GaStat. The new design and utilization of advanced statistical data and geographic entities presented in this study are expected to address the limitations of the current systems in the GaStat and to be able to support the vision of the government in economic changes.

First, this study proposed statistical surveys that use the methods of field surveys, partnership agreements, and self-response to feed the database in the Information Bank. More specifically, additional types of statistical surveys are identified and included in the model.

These types include, Saudi Community Survey, Saudi Housing Survey, Saudi Income Survey, and Saudi Spending Survey, Saudi Economic Survey, Saudi Industry Survey, Saudi Agricultural Survey, and Saudi Employment Survey. Further, a serial of formatting time methods of quarters, one year, three years and five years is used in the statistical systems in order to provide up-to-date information.

Second, the geographic entities are classified into legal and administrative entities, and statistical entities based on their corresponding geographic subdivision and they are organized in a hierarchical structure. This design is expected to provide a powerful tool for collecting the information and creating a standard set for the GaStat in order to respond the requests from users. More specifically, legal and administrative entities are in charge of the country and its provinces, governorates, holy areas, economic zones, ZIP code areas, school districts and voting districts. The statistical entities include regions, statistical tracts, statistical block groups, statistical blocks, urban areas, urban growth areas, places, sample-data areas and governorates subdivisions. Additionally, the unique reference numeric code, namely GeoID, is used to integrate spatial and non-spatial data of the geographic entities.

The ultimate goal of this study is to support statistical data integration for the requisite legal, and administrative, geographic and statistical entities within the Kingdom of Saudi Arabia in order to achieve a new vision of governance, which will enhance the proposed economic and social changes designed to benefit the people of Saudi Arabia.

REFERENCES

Central Department of Statistics & Information. (2007). *Demographic Survey 1428H* (2007). Retrieved from http://www.stats.gov.sa/ar/31

Central Department of Statistics & Information. (2013). *Household Expenditure and Income*Survey. Retrieved from http://www.stats.gov.sa/ar/37

Central Department of Statistics & Information Detailed Results of 1425AH - 2004AD. (2004).

*Population & Housing Census (Demographic Characteristics (Regions)). Retrieved from http://www.stats.gov.sa/ar/13

Central Department of Statistics & Information. (2008). *Population and Housing 2008*.

Retrieved from http://www.stats.gov.sa/ar/43

Central Department of Statistics & Information. (2008). *Kingdom of Saudi Arabia Map*.

Retrieved from http://www.stats.gov.sa/ar/10

Environmental Systems Research Institute (Esri). (2011, January). *GIS Best Practices*. Retrieved from https://www.esri.com/library/bestpractices/urban-regional-planning.pdf

Environmental Systems Research Institute (Esri). (n.d.). *Demographics and Business Data List- ACS*. Retrieved

from http://downloads.esri.com/esri_content_doc/dbl/us/2015_US_ACS_Data_List.pdf

Environmental Systems Research Institute (Esri). (n.d.). *Demographics and Business Data List-Census*. Retrieved

from http://downloads.esri.com/esri_content_doc/dbl/us/2013%20Esri%20Demographic%2
0and%20Business%20Data%20List%20-%20Census%202010.pdf

Environmental Systems Research Institute (Esri). (n.d.). *Demographics and Business Data List-Saudi Arabia*. Retrieved from

http://downloads.esri.com/esri_content_doc/dbl/int/Saudi_Arabia_MBR_2014.pdf

Environmental Systems Research Institute (Esri). (n.d.) *Methodological Notes and Usage of MB-Research Market Data*. Retrieved from

http://downloads.esri.com/esri_content_doc/dbl/int/mb-research_notes.pdf

General Authority for Statistics. (2016). *GCC "Energy Statistics" in Riyadh* .Retrieved from http://www.stats.gov.sa/en/news/62#sthash.6s1OJTfg.dpuf

General Authority for Statistics. (2016). *Updating the general framework of population and housing paves the way for a new labor force survey*. Retrieved from http://www.stats.gov.sa/en/news/59#sthash.rWMO7i9M.dpuf

General Authority for Statistics, (2016). Saudi Arabia Participates in the UN Statistical

Commission session. Retrieved

from http://www.stats.gov.sa/en/news/56#sthash.WKECcozu.dpuf

Herodotus. (n.d.). *The people of ancient Egypt*. Reshafim. Retrieved from http://www.reshafim.org.il/ad/egypt/people/index.htm

Kubbara, F. (2002). *Evaluation of GIS use in Kingdom of Saudi Arabia*. Retrieved from http://www.kfu.edu.sa/ar/Deans/Research/Documents/1029.pdf.

Kubbara, F. (2007). *International Digital Address System*. Retrieved from http://www.gisclub.net/library/research/gps/0001.pdf

Lynn, P. (2008). *International Handbook of Survey Methodology*. Retrieved from http://documents.tips/documents/survey-methodology-13.html

Ort, C., Eppler, J. M., Scheidegger, A., Rieckermann, J., Kinzig, M., & Sörgel, F. (2014).

Challenges of surveying wastewater drug loads of small populations and generalizable aspects on optimizing monitoring design. Addiction, 109(3), 472-481 10p. doi:10.1111/add.12405.

Retrieved from

http://eds.a.ebscohost.com.manowar.tamucc.edu/eds/detail/vid=1&sid=da26966d-e9d2-4b24-bed9-

3d6b262b567f%40sessionmgr4004&hid=4205&bdata=JnNpdGU9ZWRzLWxpdmUmc2NvcGU 9c2l0ZQ%3d%3d#AN=104023596&db=ccm

Rothwell, N. D., & Rustemeyer, A. M. (1979). *Studies of Census Mail Questionnaires*. Journal Of Marketing Research (JMR), 16(3), 401-409. Retrieved from

http://eds.a.ebscohost.com.manowar.tamucc.edu/eds/detail/detail?vid=1&sid=de947ba5-fe47-4fe3-8c10-

<u>c54dd5965243%40sessionmgr4001&hid=4205&bdata=JnNpdGU9ZWRzLWxpdmUmc2N</u> veGU9c2l0ZQ%3d%3d#AN=5004631&db=bth

The Economist. (2016). *Saudi Arabia's post-oil future*. Retrieved from http://www.economist.com/news/middle-east-and-africa/21697673-bold-promises-

The PC Magazine Encyclopedia. (n.d.). *Geographic entity*. Retrieved from http://encyclopedia2.thefreedictionary.com/geographic+entity

U.S. Department of Commerce Economics and Statistics Administration Bureau of the Census.
(1994 ,November) .Geographic Areas Reference Manual. Retrieved
from https://www.census.gov/geo/reference/garm.html

U.S. Government Accountability Office. (2014, February10). *Census Bureau Can Improve Use of Leading Practices When Choosing Address and Mapping Source*. Retrieved from http://eds.a.ebscohost.com.manowar.tamucc.edu/eds/detail/detail?vid=3&sid=d0ea6e39-

8a5f-49e8-b2a7-

<u>2138c805b5a3%40sessionmgr4002&hid=4205&bdata=JnNpdGU9ZWRzLWxpdmUmc2NvcGU</u> 9c2l0ZQ%3d%3d#AN=98700355&db=f5h

Vining, G. (2013). Technical Advice: Scientific Method and Approaches for Collecting Data. *Quality Engineering*, 25(2), 194–201. http://doi.org/10.1080/08982112.2013.764228.

Retrieved

5b-a449-4f37-8439-

from http://eds.a.ebscohost.com.manowar.tamucc.edu/eds/detail/detail?vid=1&sid=c892ab
5d-6dd8-4372-bc5c-

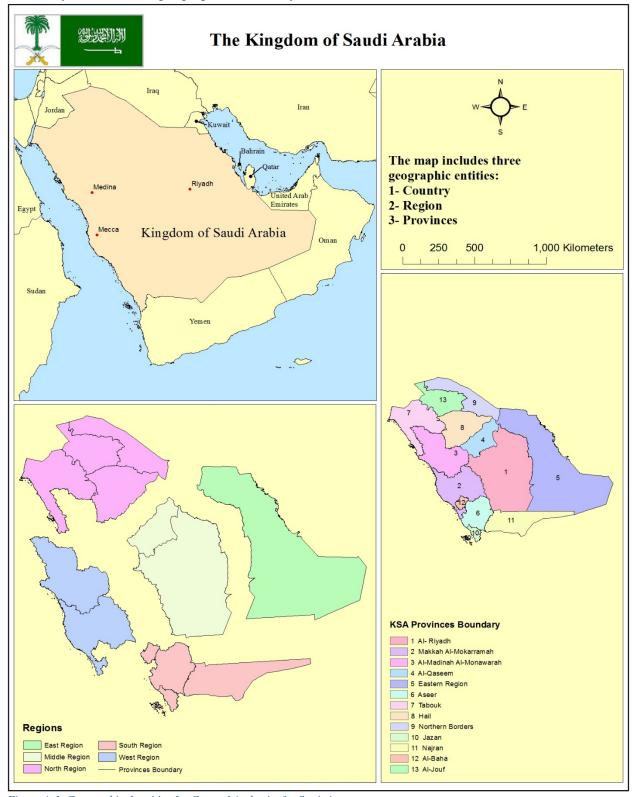
cfc04b508a2a%40sessionmgr4005&hid=4205&bdata=JnNpdGU9ZWRzLWxpdmUmc2N vcGU9c2l0ZQ%3d%3d#AN=85797310&db=bth

2020 census: Bureau needs to improve scheduling practices to enhance ability to meet address list development deadlines: report to the Chairman, Committee on Homeland Security and Governmental Affairs, U.S. Senate. (2013). [Washington, D.C.]: United States Government Accountability Office, 2013. Retrieved from http://eds.b.ebscohost.com.manowar.tamucc.edu/eds/detail/detail?vid=1&sid=1a5730

<u>292d85650511%40sessionmgr103&hid=122&bdata=JnNpdGU9ZWRzLWxpdmUmc2Nvc</u> GU9c2l0ZQ%3d%3d#AN=texasam.b1588709&db=cat00069a

APPENDICES

Appendix includes maps of theoretical standard set hierarchy of the Saudi Arabia General Authority for Statistics geographic entities system.



 $Figure\ A-1.\ Geographical\ entities\ for\ General\ Authority\ for\ Statistics.$

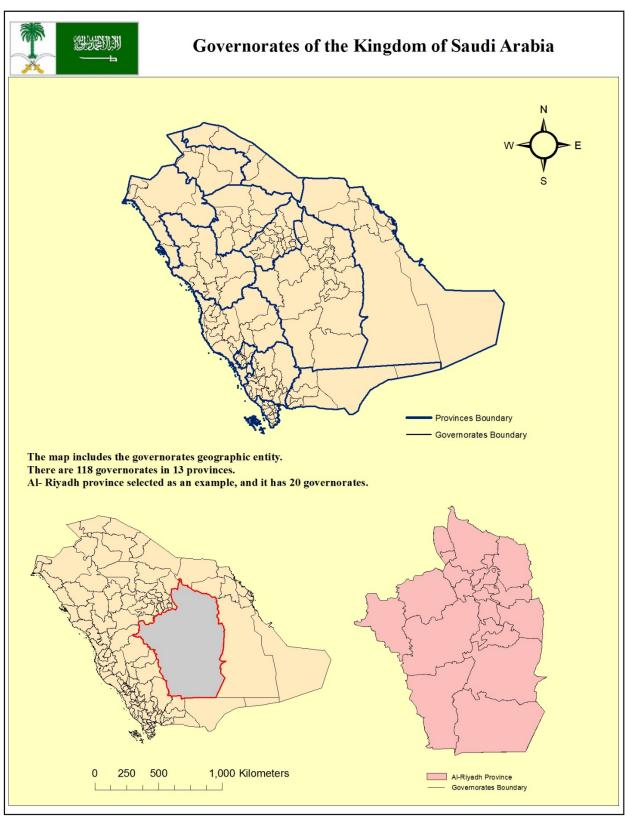


Figure A-2. Geographical entities for General Authority for Statistics.

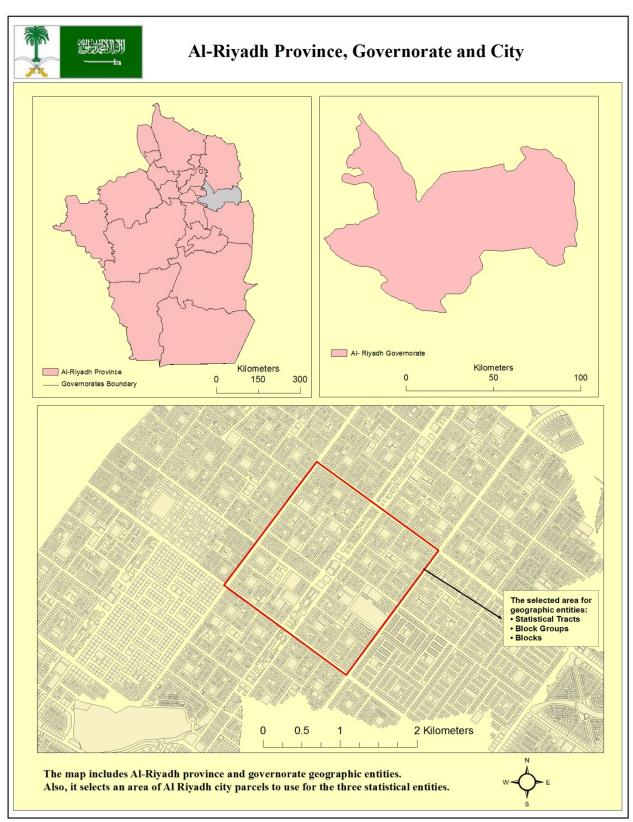


Figure A-3. Geographical entities for General Authority for Statistics.

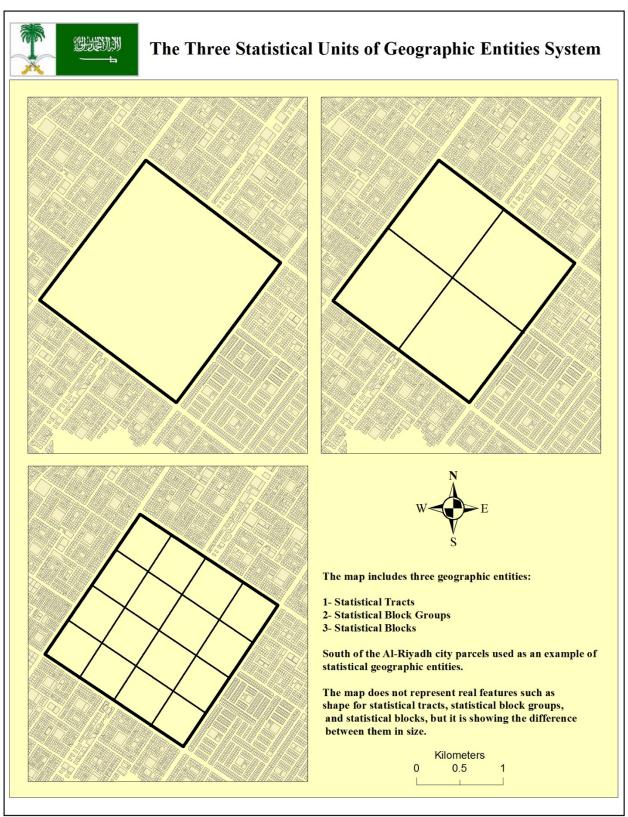


Figure A-4. Geographical entities for General Authority for Statistics.

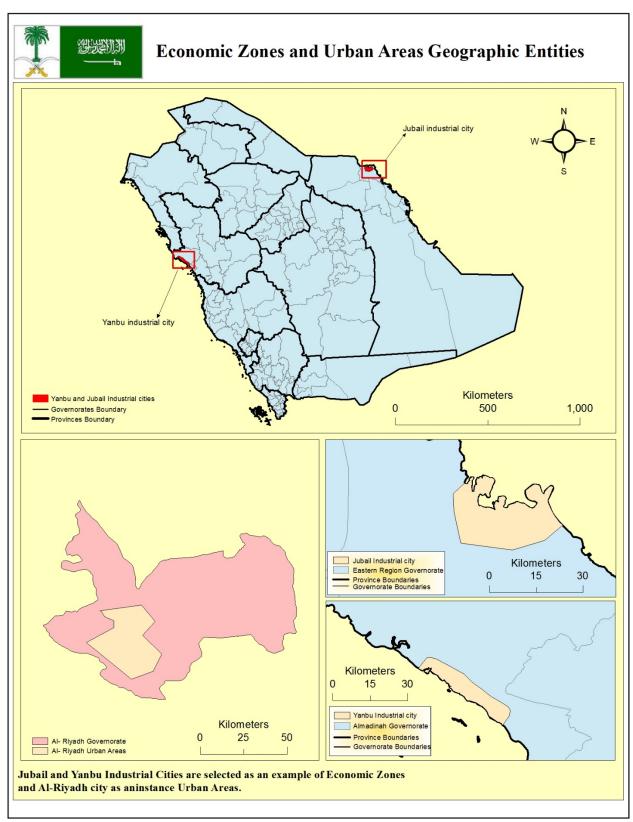
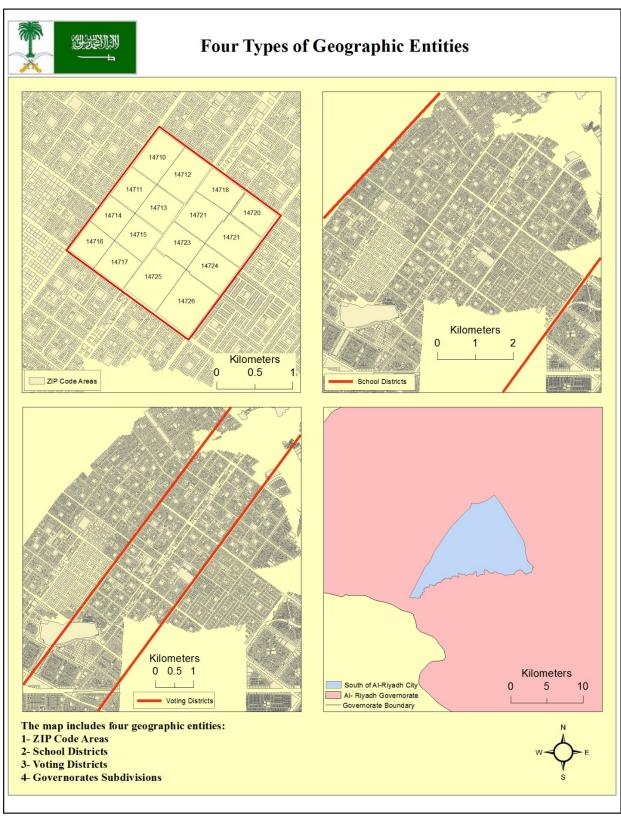


Figure A-5. Geographical entities for General Authority for Statistics.



 $Figure\ A-6.\ Geographical\ entities\ for\ General\ Authority\ for\ Statistics.$

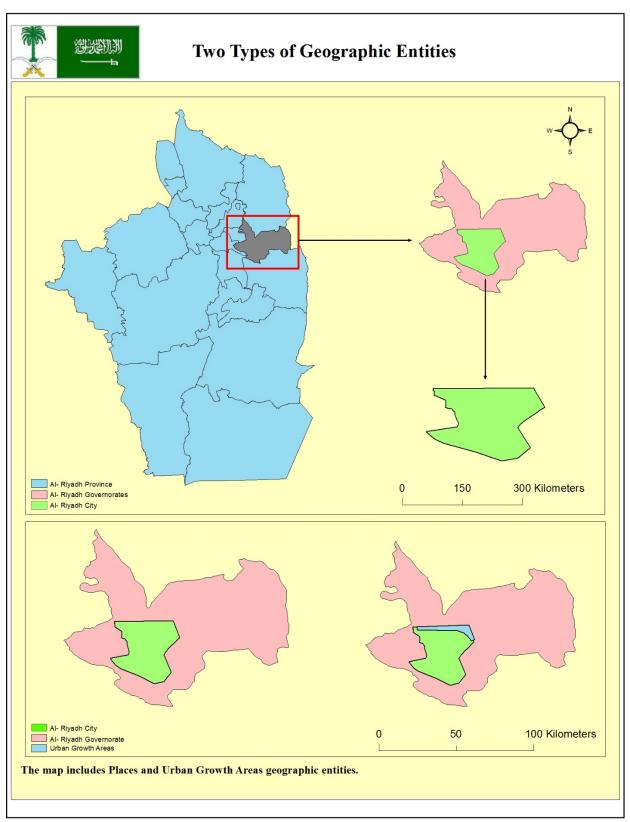


Figure A-7. Geographical entities for General Authority for Statistics.