International Journal of Science and Research (IJSR)
ISSN (Online): 2319-7064
Index Copernicus Value (2013): 6.14 | Impact Factor (2014): 5.611

Developing an Organizational Geospatial Data Framework: A Case Study of Kenya National Bureau of Statistics

Hellen Wanyoike[1], David Kuria[2]

Institute of Geomatics, GIS & Remote Sensing, Dedan Kimathi University of Technology, P.O. Box 657-10100, Nyeri, Kenya

Abstract: ***Many organizations across the world duplicate geospatial data due to lack of a platform where they can identify this data. The main objective of this study was to develop an organizational geospatial data framework for the Kenya National Bureau of Statistics (KNBS). The study addresses components of the framework namely information, technical, operational and business contexts and the institutional roles. This involved interviews and consultations with various officers to come up with representative framework components. The data standards were generated and documented guided by the Federal Geographic Data Committee (FGDC) Data Standards. Using the framework, the data was processed in a GIS and a geoportal. The results demonstrate how KNBS eight most commonly used geospatial data themes and one non- geospatial theme can be represented and shared across the organization. Two Standards are provided for the framework; the data standards documented in the KNBS Map production and Specification Manual, and the metadata standard, which is an ISO 19139 Geographic Information Metadata - XML Schema Implementation. A geoportal has been developed allowing users access the KNBS geospatial data. This KNBS geospatial data framework is a case study from which other organizations can learn from and come up with their own frameworks.***

Keywords: Geospatial Data, Geospatial Data Framework, Framework Components, Statistical Data

1. Introduction

Spatial data is data that identifies the geographical location of features and boundaries on the Earth surface. The data plays a central role in supporting economies and promoting business in both private and public sectors all over the world [1]. Production of geospatial data is very expensive but many organizations that use and produce it spend a lot of resources duplication the same due to lack of a platform for discovering, accessing and sharing it. A geospatial data framework reduces data duplication by providing a reliable and standardized source for commonly used geographic data themes. It can be used as a data resource into which geographical data producers can contribute data and continually change and improve this data. Its elements include procedures, guidelines, and technology to enable users build, integrate, maintain, distribute, and use framework data. It also features institutional relationships and business practices.

The geospatial data framework is one of the key building blocks and forms the data backbone of the Spatial Data Infrastructure (SDI). SDIs enable the sharing of geospatial information at organizational, national, regional, or global level, and so link people to data.They reduce duplication in spatial data creation and maintenance [2]. The first generation SDI development started in the 1980s in USA and Australia to promote economic development, stimulate better governance and foster environmental sustainability [3]. In Africa the initiative started in 2000 but the implementation is being done at a seemingly slow pace and is still in its infancy [4].

The Kenya National Spatial Data Infrastructure (KNSDI) was conceptualized in the year 2001. There is a lot of geospatial data that is produced and used by both private and public organizations but there is no systematic way of accessing and sharing it mainly due to lack of appropriate policy, institutional and legal arrangements on data sharing, access and exchange [5]. It is therefore difficult to know what data is available in any organization, and this has led to duplication of efforts and wastage of resources.

This research focused on creating a spatial data framework for the Kenya National Bureau of Statistics (KNBS). The organization collects spatial data after every ten years during the census cartographic mapping in order to create Enumeration Areas (EA) for census undertaking. Having a Geospatial Data Framework in the organization will help avoid or minimize duplication and enhance sharing and dissemination of collected data.

KNBS is a semi-autonomous government agency under the Ministry of Devolution and Planning. It is mandated to carry out the Kenya population and housing censuses and since 1969 it has carried out censuses after every ten years, with the next one to be anticipated in 2019. The organization has offices in all the 47 counties, which are manned by County Statistical Officers (CSO). KNBS headquarters are situated in Herufi House in Nairobi. The organization has other offices in Nyayo house on 16th, 17th and 18th floors, where the directorate of Population and Social Statistics is situated. The organization has six Directorates namely: Population and Social Statistics, Finance and Administration, Information and Communication Technology, Strategy and Development, Micro Economics Statistics and Production Statistics all of which are headed by directors.

The study was confined to the directorate of Population and Social Statistics which has three divisions namely: Population, Social Statistics and Sampling, Cartography\GIS and Field Administration. The directorateoverseesthe population and housing census exercise, generate population

and social sector statistics (population, demographic, health, education, governance, gender, vital), develop and maintain the National Sample Surveys and Evaluation Programme (NASSEP) frames carry out cartographic mapping and the co-ordination of activities in all KNBS offices in the 47 counties across the country. A geospatial data framework is therefore necessary to avoid duplication of data as the directorate executes its mandate.

2. Study Area

Nyeri County, the study area (figure 1), is located in the Central region of Kenya and to the south western side of Mount Kenya. It has an area of 3356 $km^{2.}$ It borders Laikipia County to the North, Kirinyaga to the East, Murang'a to the South and Nyandarua to the west. It has 8 sub counties, 6 constituencies and 30 wards. The county had 2 districts namely Nyeri North and Nyeri South, 196 sub-locations and a total population of 693,558 based on the 2009 census.

Figure 1: Nyeri County - Study Area

3. Methodology

Components of the framework were identified through interviews and working together with KNBS GIS staff. The most commonly used geospatial data were documented together with their specifications for use during capturing and processing of the data and during map composition. The specifications that formed the framework data standards implemented the Federal Geographic Data Committee (FGDC) data standards, being structured to reflect the needs of the organization. These specifications included (i) the features specifications which detailed the name, type and width of the features to be used, (ii) the administrative unit specifications consisting of the field name, type and width of the codes and names of both the administrative and political boundaries, and (iii) the EA map specifications and symbol specifications specifying the symbol type, colour, font type, font size and width of all the symbols used to represent the features used in creation of maps. These standards guide the digitization of physical features and boundaries, creation of attribute tables and the creation of EA maps to be used in census enumeration to ensure standardization.

Data to test the framework was collected in-house and comprised of both geospatial data in form of shape files and statistical data provided in excel tables. ArcGIS software was used to process the data to the set framework standards and in integration of geospatial and statistical data sets. A geoportal was developed, providing access to the KNBS geospatial data. This geoportal is based on the OpenGeo Suite opens source software. Figure 2 provides an overview of the entire process of undertaking the research.

3.1 Data

All the data used was collected from KNBS and consisted of both geospatial and non-spatial data sets. They were collected from Cartography/GIS, Population sections and Nyeri county Office, all within the Directorate of Population and Social Statistics. These data sources are tabulated in table 1

Table 1: Data Source

Data	Type	Source
Transport	Geospatial	Cartography/GIS
Hydrography		
Imagery		
Boundary		
Locational		
Population	Statistical	Population
Education	Statistical	Nyeri County Office

Transport data included roads and railway lines used to delineate, identify and access EAs. Hydrographic data included rivers, lakes, ocean, dams and water holes used to delineate and identify EAs. Imagery data was used to easily delineate EA Boundaries which comprised of both administrative and political boundaries. The boundaries were to ensure enumerators and research assistants worked within their scope and also for geospatial analysis of the statistical data. Locational data such as schools, health facilities and markets were landmarks used in identification while the statistical data was to be integrated with the geospatial data for geospatial analysis.

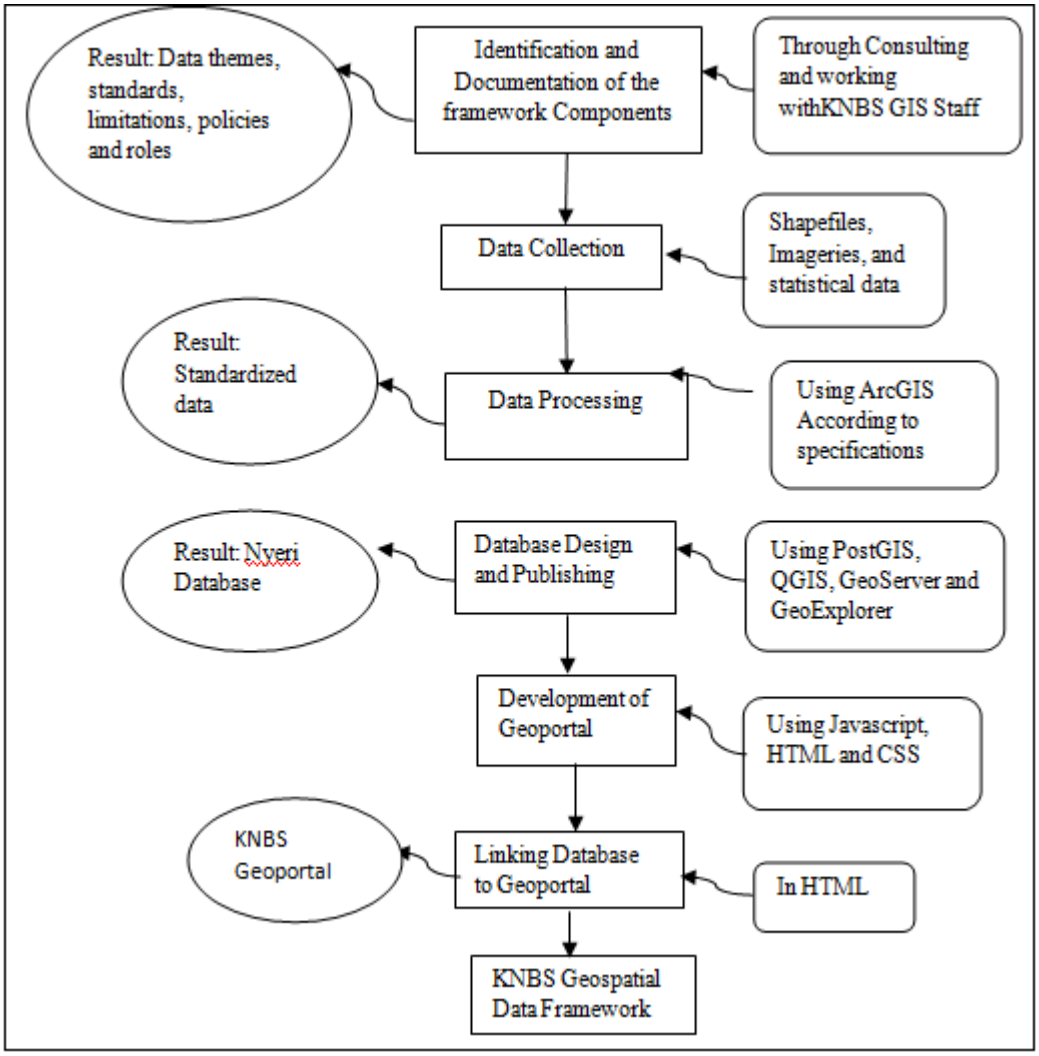

Figure 2: Research Strategy

4. Results

The main objective of the study was to develop a geospatial data framework. The framework defines the different types of data themes, their standards, status, the coordinate system and the metadata standard used. The framework comprises of nine data themes eight of which are geospatial and onenon-geospatial. The eight geospatial data themes include boundary, transportation, hydrography, place names, locational data, forests and national parks, imageries and maps while the non-geospatial theme is statistical data theme.A major updating of the geospatial data themes was to be done during the census cartographic mapping exercise done after every ten years.

4.1 KNBS Spatial Data Framework

The document was created on October, 2015. It contains 8 geospatial data themes that add value to the statistical data (statistical theme) collected by KNBS by showing where the data was collected. The data is also available to other users who include students, government and private organizations. The themes, their description, purposes, status and standards are shown from figures 3 -10

4.1.1 Boundaries

1. Description

These are census boundaries created at an Enumeration Area level with the help of the local administration. Base maps used during this exercise include topographic maps procured from Survey of Kenya and aerial photographs and satellite imageries. They are supplemented with maps used during previous censuses. An Enumeration Area (EA) comprises of part, whole or combination of several villages and it is identified with the County, Sub-County, Division, Location, Sub-Location and Village as well as the constituency and ward it falls in. According to the 2009 Population Census Kenya had 47 Counties, 158 District, 635 Divisions, 2728 Locations, 7149 Sub locations and 96252 Enumeration Areas.

2. Datasets:

Administrative boundaries: Counties, Sub-Counties, Divisions, Locations, Sub-Locations and Enumeration Areas.

Political Boundaries: Constituencies and wards

3. Purposes

The purposes of these boundaries include:

- Ensuring complete coverage of the country during census enumeration

- Budgeting and planning of the census enumeration
- Analysis of census data and other statistics

4. Status
The boundaries are updated after every ten years prior to the census enumeration.

5. Data Standard
KNBS WI-54-7-1: Map Production and Specification Manual.

6. Metadata Standard
ISO 19139 Geographic Information Metadata - XML Schema Implementation

7. Coordinate System
Geographic Arc 1960 coordinate system

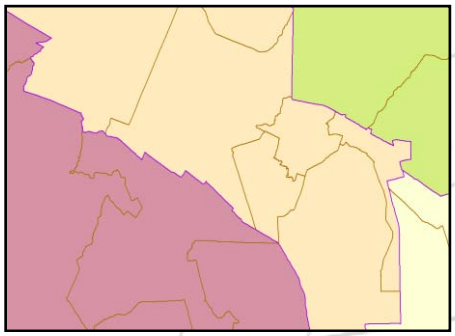

Figure 3: Boundaries

4.1.2 Imageries
1. Description
This is data captured using aircrafts and space borne vehicles. They are used to identify physical features both natural and man-made.
2. Datasets
These include aerial photographs and satellite imageries that are most current in order to accurately portray the situation as it is on the ground during delineation of EAs.
3. Purposes
To guide in delineation of EAs for they show the exact situation of features on the ground
4. Status
The imageries used are of high resolution so that much detail can be detected
5. Standards
Those who acquire this information have to be licensed.
6. Coordinate System
Geographic Arc 1960

Figure 4: Imageries

4.1.3Place Names

1. Description:
These are names of areas and features and are given by the local administration
2. Datasets:
Include geographical names of places and the names of physical features such as mountains, forests, schools among others
3. Purposes
They are important for cultural identity and heritage. They also help in identification of location of places and features
4. Status
The names are as provided by the locals and administration.
5. Data Standard
KNBS WI-54-7-1: Map Production and Specification Manual.
6. Metadata Standard
ISO 19139 Geographic Information Metadata - XML Schema Implementation
7. Coordinate System
Geographic Arc 1960 coordinate system

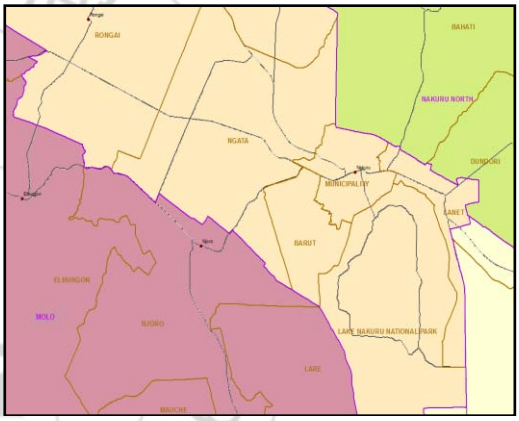

Figure 5: Place Names

4.1.4 Transport Data
1. Description
Transport data include means used to move people and goods. It contains connectivity and characteristics such types and names
2. Datasets
Include data on roads, railway lines, ports and airports
3. Purposes
They are used to delineate, access and identify the EAs
4. Status
They are provided by the Transport ministry and are updated during the census cartographic mapping exercise.
5. Data Standard
KNBS WI-54-7-1: Maps Production and Specification Manual
6. Metadata Standard
ISO 19139 Geographic Information Metadata - XML Schema Implementation
7. Coordinate System
Geographic Arc 1960

Figure 6: Road Network

4.1.5 Hydrography

1. **Description:**

These are data on water bodies

2. **Datasets:**

Include data on rivers, dams, lakes, ocean and coastline

3. **Purposes**

They are used to delineate and identify the EAs

4. **Status**

They are captured from imageries and topographic maps and are updated during the census cartographic mapping exercise. They are described by their characteristics such as type.

5. **Data Standard**

KNBS WI-54-7-1: Map Production and Specification Manual.

6. **Metadata Standard**

ISO 19139 Geographic Information Metadata - XML Schema Implementation

7. **Coordinate System**

Geographic Arc 1960 coordinate system

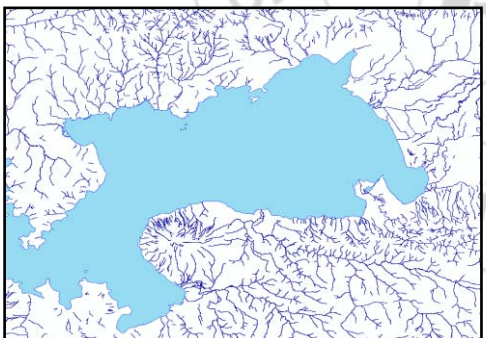

Figure 7: Imageries

4.1.6Locational data

1. **Description:**

These are XY data collected using GPSs and show locations of features.

2. **Datasets:**

Include locational data on schools, markets, health facilities, cattle dips and buildings.

3. **Purposes**

They are used to identify the EAs

4. **Status**

They are updated during the census cartographic mapping exercise. They are described by their characteristics such as type.

5. **Standards**

KNBS WI-54-7-1: Map Production and Specification Manual.

6. **Metadata**

The data uses ISO 19139 Geographic Information Metadata - XML Schema Implementation.

7. **Coordinate System**

The locational data are picked using Geographic Arc 1960 coordinate system

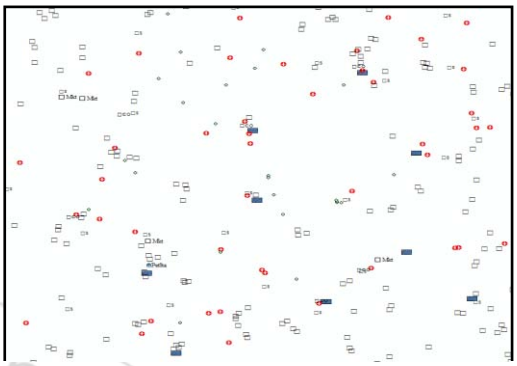

Figure 8: Locational Data

4.1.7 Forests and National Parks

1. **Description:**

These are protected areas that are clearly defined and legally recognized for conservation of nature.

2. **Datasets:**

Include data on forests, national parks and national reserves.

3. **Purposes**

To identify whether there are people residing there so that they are not left out during the census enumeration.

4. **Status**

They are captured from topographic maps and imageries and area updated during the census cartographic mapping exercise.

5. **Standards**

KNBS WI-54-7-1: Map Production and Specification Manual

6. **Metadata**

It is created using the ISO 19139 Geographic Information Metadata - XML Schema Implementation

7. **Coordinate System**

The data is in Geographic Arc 1960 coordinate system

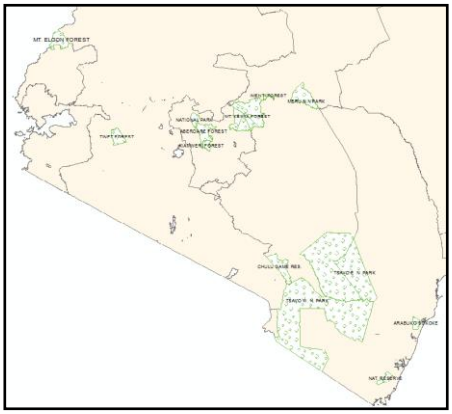

Figure 9: Imageries

4.1.8Enumeration Area Map

1. Description:

An EA map is a map of part or whole sub\location that shows the number of EAs it represents .It is the main product of KNBS spatial data. The map shows the layout of boundaries and features within that area. It is available in both hard and soft copies.

2. Datasets:

The datasets used in the production of this map are various levels boundaries and features.

3. Purposes

Identification of the EA each individual will work in during the census enumeration and also during household based surveys. It guides during planning of the work.

4. Status

They are created after every ten years and updated when carrying out household based survey.

5. Standards

KNBS WI-54-7-1: Map Production and Specification Manual.

6. Metadata

It is created using the ISO 19139 Geographic Information Metadata - XML Schema Implementation

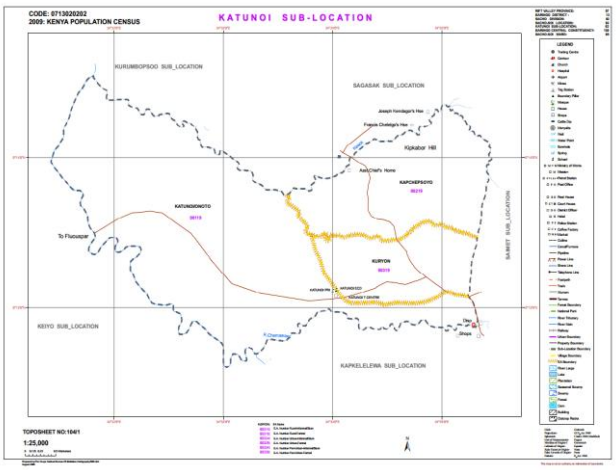

Figure 10: Imageries

4.1.9 Statistical Data

1. Description

This is data collected during census enumeration or other surveys using a questionnaire structured in such a way that it can be integrated with geospatial data. It is availed in table format as an excel worksheet for easy integration with the spatial data.

2. Datasets:

It includes population, demographic and social statistics collected based on geographical locations.

3. Purposes

This data is used for national planning and used by other stakeholders.

4. Status

The population data is collected after every ten years while other statistical data sets are intercensal therefore collected between census dates. They are either programmed like Kenya Demographic Health Survey (KDHS) or collected when need arises.

5. Standards

Statistical data are collected guided by the KNBS Work Instruction (WI) manuals and administrative codes are used to integrate it with the geospatial data.

Table 2: Disability by Type

Sub Location	*Visual*	*Physical*
NARO MORU	6	6
NDIRITI	3	25

4.2 Data Standards

Two standards were adopted for the framework, (i) the KNBS data standards comprising of features specifications, EA map specifications, administrative units' specifications and symbol specifications, and (ii) the metadata standard, "ISO 19139 Geographic Information Metadata XML Schema Implementation". Included in the data standards are the framework's coordinate system based on the Arc 1960 ellipsoid, and the unique identification code, which is the administrative unit's codes. The data standards are documented in the "KNBS WI 54-7-1 Map Production and Specification Manual". The metadata standard describes the themes, their limitations of use and gives credit to those involved with the production of the themes. Tables 3 – 6 give the specifications for the various information pieces that are required in the framework.

Table 3: Features specifications

	Type	**Width**	**Dec**
Forest layer	Text string	25	
i) Area	Number/Short Integer	7	2

Table 4: Administrative Units specifications

Field Name	*Type*	*Width*	*Decimal Places*
COUNT CODE	Number/Short Integer	**3**	0

Table 5: Symbols Specifications

	Type	*Width*	*Dec*
Forest layer	Text string	25	
i) Area	Number/Short Integer	7	2

Table 6: EA Map Specifications

Item	*Colour*	*Text Size*
Neighbouring Sub Counties	Black	10

The existing KNBS Policy which detailed the data sets found in KNBS was adopted for the study. It can be accessed from the organization's portal *(http://www.knbs.or.ke)*. Since KNBS had no geospatial data framework where geospatial data could be identified, this has been glossed over in the policy. It gets only a mention in the scope of the policy.

4.3 Maps

Various data sets of Nyeri County were collected to test the framework. These datasets were converted in accordance with the framework allowing easy integration and overlay. After conversion of the data, administrative, political and thematic maps of Nyeri were produced. Some of these maps are given in figure 11.

4.4Geoportal

A prototype geoportal has been developed in this work. All the data was put into a geo-database from which the portal reads the information and presents it to the end users. The geoportal has various capabilities: querying the data, printing it, map view, interactive manipulation and measurement capabilities. Using the identifying tool the user gets information concerning a particular feature, its name and other details attached to it. The print option allows the user to select the printer and orientation of the map to be printed. Using the query tool one can query data by attributes. All these capabilities are accessible to any user who accesses the geoportal irrespective of the internet browser they may be using. The capability to edit or modify the data is limited to only those with the requisite authentication credentials. Figure 12 shows the landing site of the geoportal.

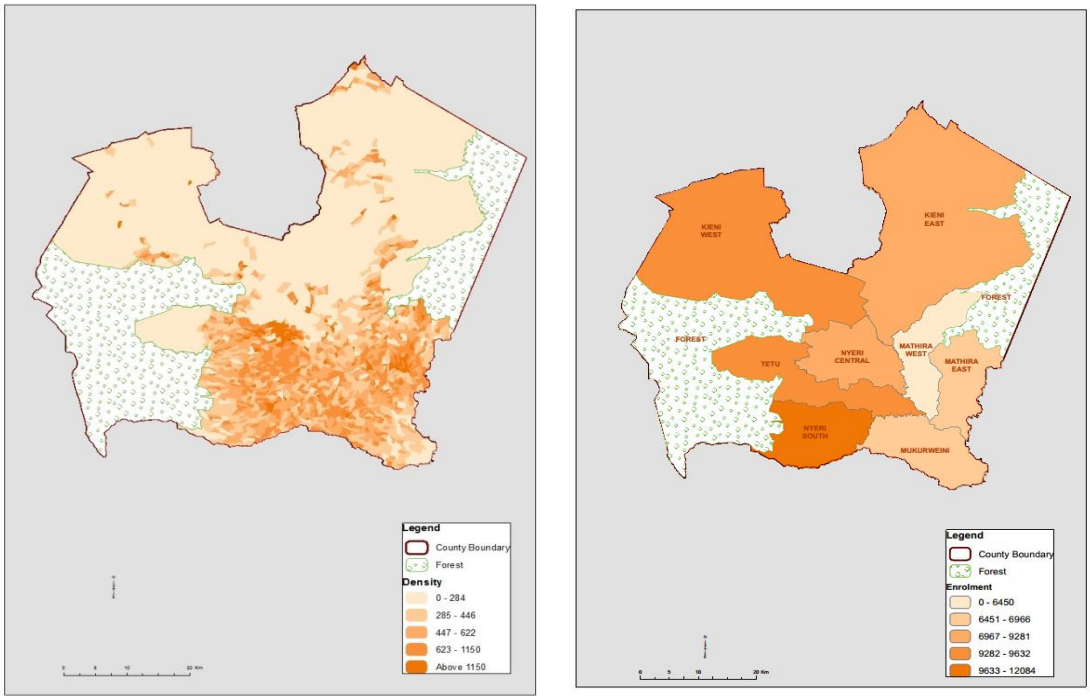

Figure 11: Thematic maps of Nyeri County (a) Population density and (b) Secondary school enrolment

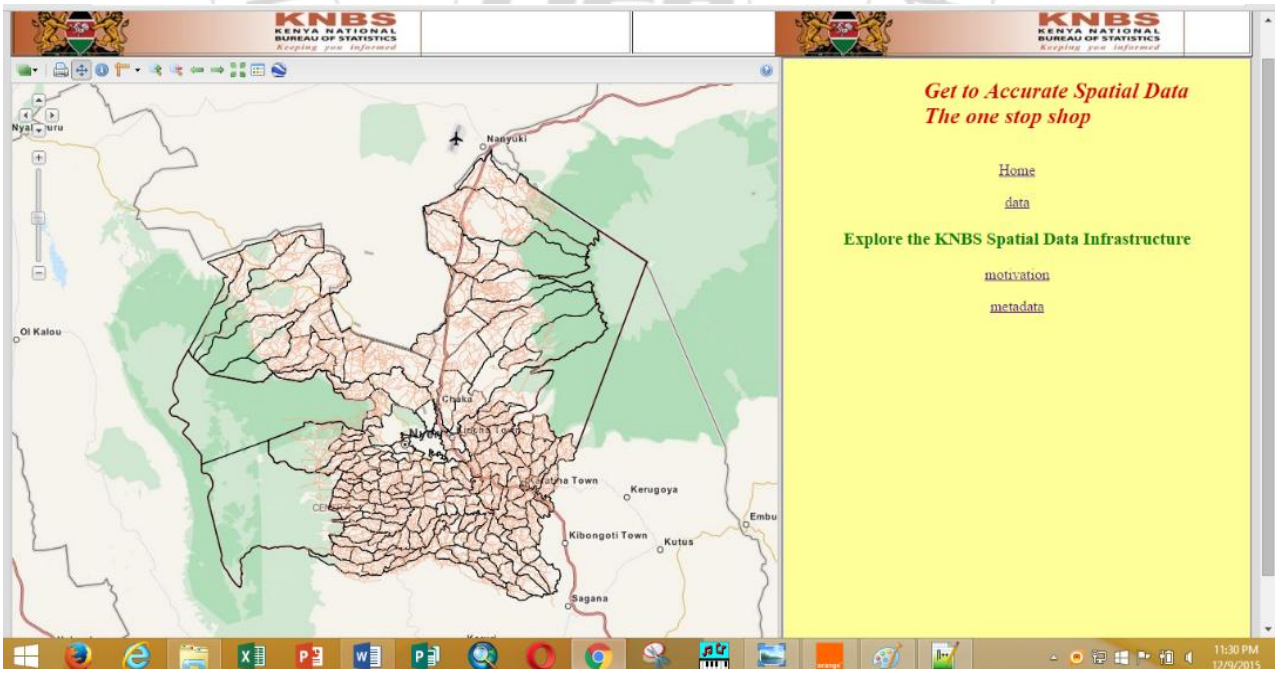

Figure 4: KNBS Geoportal

5. Conclusion

The result showsnine (9) data themes that are mostly commonly used in KNBS. These data themes are majorly updated after every ten years. They are common in other organizations and therefore sharing will reduce duplication. A geoportal has been developed in this work, allowing users access the KNBS geospatial data. The Framework will help the organization identify what is already available and therefore only work at updating it rather than duplicating it,

thereby minimizing cost of data collection and maintenance. It will also help the organization enforce the standards of the data it requires when acquiring it from other organizations. Other organizations will benefit from the Framework since they will know what is available in KNBS and its status and thereby avoid duplication. The existence of the KNBS Framework will benefit other organizations, by being a reference benchmark as they embark on the process of developing their own Geospatial Data Frameworks, and it will lay a good foundation upon which a National Spatial Data Framework can be built.

6. Recommendations

Other organizations that deal with geospatial data should come up with their own frameworks which will help identify what exist and in what standards. This will help the country come up with common standards of data which will be important in integration of various data sets

References

[1] Genovese, E., Roche, S., Caro, C., (2009). The value Chain Approach to Evaluate the Economic Impact of Geographic Information
[2] U.S. Federal Geographic Data Committee, (1995). Development of a National Geospatial Data Framework
[3] Vandebroucke, D., & Jansse, K. (2008). INSPIRE State of Play: Generic approach to assess NSDI. In Crompvoets, J., A. Rajabifard, B. van Loenen and T. Delgada (Eds.). A Multi-View Framework to Assess National Spatial Data Infrastructures (ISBN: 978-0-7325-1623-9), Melbourne University Press, Australia.
[4] Makanga, P., & Smit, J., (2008). A review of the status of Spatial Data Infrastructure Implementation in Africa. Free and open source software for Geospatial (Foss4g) conference, (pp. 357 - 365). Cape Town, South Africa.
[5] Masser, I., (1999). All shapes and sizes: the first generation of National Spatial Data Infrastructures. In International Journal of Geographical Information Science, 13, 67–84

Authors Profile

Hellen Wanyoike works with Kenya National Bureau of Statistics. She received her BA in Geography from the University of Nairobi and is currently pursuing her M.Sc. Degree at Dedan Kimathi University.

David Kuria is an Associate Professor of the Dedan Kimathi University. He has a B.Sc. Degree in Surveying from Nairobi University, a M.Sc. in Photogrammetry and Geoinformatics from Stuttgart University of Applied Sciences and a Ph.D. in Civil Engineering from the University of Tokyo.