American Scientific Research Journal for Engineering, Technology, and Sciences (ASRJETS)
ISSN (Print) 2313-4410, ISSN (Online) 2313-4402

http://asrjetsjournal.org/

Opportunities and Challenges for Integrating Statistical and Spatial Data in East African Countries

Job Chaula*

Ardhi University, Department of Computer Systems and Mathematics, P.O.BOX 35176, Dar es salaam, +255, Tanzania

Email: jobchaula@gmail.com

Abstract

Statistical and spatial data integration for effective implementation of Sustainable Development Goals (SDGs) while addressing emerging socio economic challenges and opportunities in developing countries is globally accepted for business, policy making and data driven decisions making process. However in the EAC countries despite of a considerable increase in birth of geographical data producers, vendors and users in EACs whom do not know about each other's data holding or data needs little is known on the current opportunities and challenges for spatial and statistical data integration. In the recent there are numerous opportunities for integrating statistical and spatial data in EAC countries including availability of geographical data producers, vendors and users in EACs, availability of Geo-network portal, Region STAT and AQUASTAT program and fast emerging trends in technology development such as Volunteered Geographic Information (VGI). Among the challenges confronting the spatial and statistical data integration includes unwatched production of spatial data result much into more duplication of data, poor data quality control, inefficient use of the available data resources, suppression of the geo-information market and frustrated data users, absence of clearing house for EACs countries, lack of common geographic boundaries and data management systems, lack of adequate expertise and skilled labor for data integration, shortage of integrated infrastructure to support the integration initiatives of spatial statistical data, inconsistencies among the structure of the data, lack of regional level and well-defined frameworks. The EAC country Clearing house and Volunteered Geographic Information (VGI) programs at local need to be promoted for fast tracking the data integration in EAC countries while addressing the current challenges confronting the spatial and statistical data integration .

Keywords: Statistical and spatial data integration; statistical data; spatial data integration.

---

* Corresponding author.

1. Introduction

Data integration is the process of making the datasets compatible so that it can reasonably be displayed in the same map while analyzing their relationship [1]. Thus, data integration is the process of combining information from multiple sources to provide a more complete view of information. The overall goal of data integration is to connect and combine business information that was initially collected and placed into separate silos managed by unique stakeholders. Data integration may take place in a physical or virtual data warehouse, utilizing software that hosts large repositories of data from internal and external resources [2]. Other research scholar have lamented that, integration of statistical with spatial statistical analysis in Geographical Information systems (GIS) has the potential to become a powerful analytical toolbox, enabling regional and social scientists to gain fundamental insight into the nature of spatial structures of regional development [2]. GIS play important roles by combining disparate data sets, creation of new/derivative data sets, development and analysis of spatially explicit variables. The UN-CGIM in 2013 also have called for country statistical and spatial data integration for effective implementation of Sustainable Development Goals (SDGs) while addressing emerging socio economic challenges and opportunities in developing countries. In the recently, Reference [3] have lamented that statistical and spatial data integration is a strategic asset in policy is making and data driven decisions making process.Despite of the numerous importance of data integration for business, policy making and data driven decisions making process little is known on challenges and opportunity available in the East African Community (EAC) countries. However [4;5] warned on a considerable increase in birth of geographical data producers, vendors and users in EACs whom do not know about each other's data holding or data needs. In each of the EAC country there are some efforts that have been in place to harmonize the spatial data held by different stakeholders through establishment and implementation of National Spatial Data Infrastructure (NSDI). NSDI is a strategic tools towards bring different data producers, vendors and users and may defined as is a framework of policies, institutional arrangements, technologies, data, and people that makes it possible to share and use effectively geospatial information [1]. According to [6] NSDI also supports implementation of wide range of activities, including technical matters such as data, technologies, standards and delivery mechanisms as well as institutional matters related to organizational responsibilities and policies, need of financial and human resources. With regard to EAC countries, each country has reached different stages on development of SDI with except of Kenya and Rwanda that have full developed their NSDI. However full function NSDI in each of EAC country is vital for challenges related to unemployment, reduced agricultural production resources, effects of climate changes, increased urbanizations, growing populations and economies, changing landscapes, emerging trends of smart cities and internet. Early research studies have reported on different stages reached by each country in developing a full mechanism for integrating, coordination and accessing the statistical and spatial data. However little has is known on the challenges and opportunities for integration of spatial and statistical data in these African countries. Statistical and spatial data integrations is useful in addressing rapid emerging challenges related to environmental and climate change, land use land cover (LULC) change, urbanization and town sprawl, declining forest, biodiversity, agricultural production and disaster management. Earlier research work by [2] anticipated the dependence of future prosperity of the community in statistical and spatial integrations for applications on environmental planning, land registration, disaster response, public health programs, agricultural production and marketing and biodiversity conservation.

2. Methodology

2.1 Geographical location and description of EAC

Figure (1) the geographical location of EAC countries comprising of Republics of Kenya, Uganda, and United Republic of Tanzania, South Sudan, Republic of Rwanda and Burundi and their head quarter Nairobi, Kampala, Dodoma, Juba, Kigali and Burunjumbura, correspondingly. In the context of spatial and statistical data, these countries have different mechanism of sharing, accessing and disseminating of spatial and statistical data.

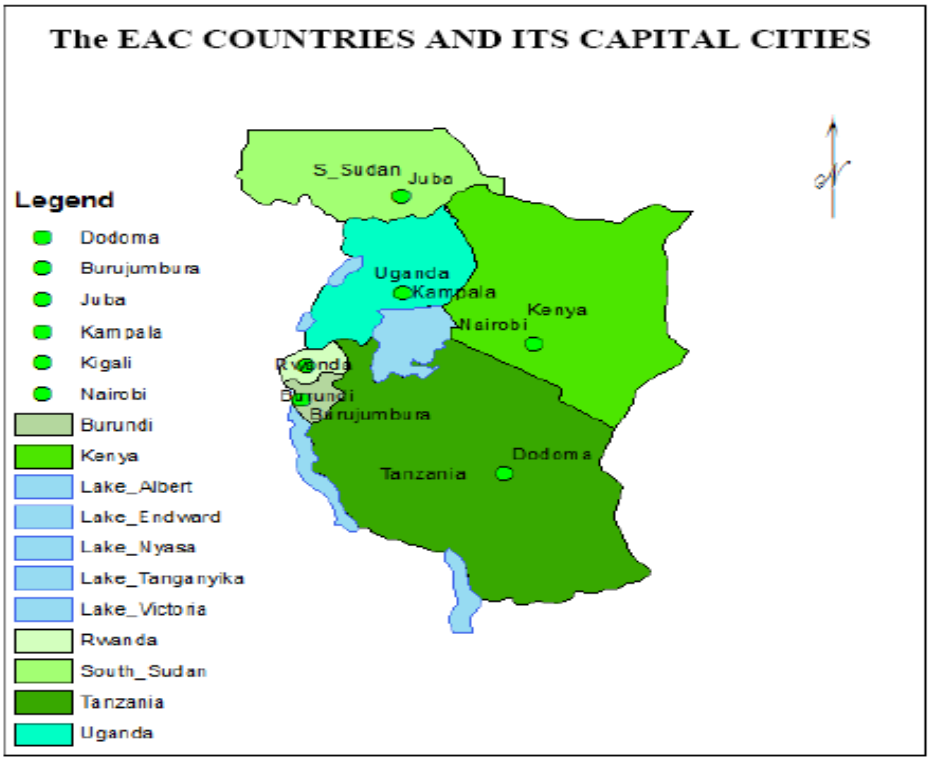

Figure 1: The geographical location of EAC countries and their head quarters

2.2 Methods

Literature review on the importance of data integration for business, policy making and data driven decisions making process in the East African Community (EAC) countries was carried out using available reports. Then, challenges and opportunities available for integration to facilitate for business, policy making and data driven decisions making process were summarized as per chapter 3.

3. Results and Discussion

3.1 Opportunities for Integrating Spatial and Statistical Data in EAC Countries

3.1.1 Availability of government and private agencies with spatial datasets

In the EAC countries, several government and private agencies collects and provides spatial and non spatial data including the National Bureau of Statistics, Ministries, Research Institutions, Census and other form of surveys. UN-GGIM (2014) narrated the constituents of data from National Statistical Office (NSO) which should comprise of dataset from census, demographic, labor force, buildings and agriculture. Other important data also may be compiled from other income and business tax, health services (medical care, pharmaceuticals, work force, and disease epidemiology), land valuation and land use, electoral roles, social welfare services (unemployment, disability and family support). Besides, FAO in implementation its three main goals which are: (i) the eradication of hunger, food insecurity and malnutrition; (ii) the elimination of poverty and the driving forward of economic and social progress for all; and (iii) the sustainable management and utilization of natural resources, including land, water, air, climate and genetic resources for the benefit of present and future generations have developed the geo-network for EAC countries available at GeoNetwork@fao.org (FAO, 2018a). The purpose GeoNetwork's is to improve access to and integrated use of spatial data and information, to support decision making, to promote multidisciplinary approaches to sustainable development and to enhance understanding of the benefits of geographic information. Thus, GeoNetwork open source allows to easily sharing geographically referenced thematic information between different organizations. Another opportunity available for integrating spatial and statistical data is AQUASTAT program which started with the aim to contribute to FAO's goals through the collection, analysis and dissemination of information related to water resources, water uses and agricultural water management, with an emphasis on countries in Africa Asia, Latin America, and the Caribbean (FAO, 2018b). AQUASTAT offers a Standardized data and information to provide decision makers with evidence to measure progress and to substantiate your decisions, tools to allow you to generate your own analysis and conclusions, articles and presentations and Capacity development to improve understanding and monitoring of water resources, uses, and irrigation management (FAO, 2017b). FAO also have developed a RegionSTAT which is a statistical framework and applied information system for analysis and policy-making at the regional level which was designed in order to organize, integrate and disseminate statistical data and metadata on food and agriculture coming from different sources. RegionSTAT gathers and harmonizes scattered institutional statistical data so that information is aggregated to the regional level and becomes compatible with data at the international level. The main objectives are to facilitate decision-makers access to information and to bind data sources that are currently spread throughout the different institutions (FAO, 2018c).

3.1.2 Fast emerging trends in technology development

Ongoing advancement of internet, mobile devices, emerging of location based services and intelligent information processing technologies, modeling which provide easier integration and access to a wide range of different services and applications [7]. However in EAC countries despite of the presence of these emerging technologies that facilitates spatial and statistical data integration there is no clear mechanism for facilitating data integration. Volunteered Geographic Information (VGI) is one type of user generated GI [7], where volunteers use the web and mobile devices to create, assemble and disseminate spatial information. Among the most well known VGI platforms are OpenStreetMap (OSM); and Wikimapia, but there are many others, covering a range of fields such as conservation, planning, and crisis management. Thus, there is a potential for VGI to become an important source of information that could benefit the EAC countries.VGI, on the other hand,

is captured unofficially by volunteers, often using cheap devices, e.g. a handheld GPS or Smartphone's; hence the data quality is usually limited and the data collection is not based on strict standards [3].

3.1.3 The Fast growing human population in EAC countries

Rapid population growth combined with changing habitations patterns, overgrazing, bio-piracy, deforestation, pollution and the unsustainable exploitation of natural resources is a growing concern in East Africa. Figure (2) shows the population and its growth rate in Each of EAC countries. United Republic of Tanzania show to have high population approximating to about 53 millions of people, followed by Kenya with about 46 millions of people, Uganda with about 39 millions of people, Rwanda with about 12, South Sudan with about 12 millions of people, and Burundi with about 11 millions of people. In the context of population growth rate, Rwanda were found to have low population growth rate of about 2% and South Sudan were found to have highest population growth rate of about 4% while Tanzania, Uganda, Kenya and Burundi were found to have highest population growth rate of about 3% per annual [8]. Making timely and data driven decision on managing natural resources, with aim to sustain water supply, food security and agricultural productivity to meet demand of these increasing populations require data integration. Thus, in future data integration will be used in several development programs such as:

- Poverty reduction programs: integrating statistic and geospatial data will help to better planning of various natural resources and land use thereby helping in balanced development in poverty reduction programs in developing countries.
- Disaster management: integrating statistic and geospatial data will help to address issue related to disaster management in developing countries through playing important role in various measures for prediction, mitigation and management of disaster such as erosion, solid waste, floods and drought to mention few.
- Agriculture: integrating statistic and geospatial data will help to address the agricultural constraints through providing suitable tools to various agriculture crops, yield forecasting and monitoring agriculture areas.
- Forestry: integrating statistic and geospatial data will assist in better inventory forest resources, sustainable use of forest resources, reforestation activities and the management of community forests.
- Biodiversity: integrating statistic and geospatial data will help to through better mapping of resources, and strategies for biodiversity conservation.
- Tourism: integrating statistic and geospatial data will assist the tourism industry with better planning of tourism infrastructure, generation and dissemination of such by virtual reality, better quality maps.
- Health and Medicine: integrating statistic and geospatial data will be useful in monitoring and management of areas under herbs, designing herbs plantation areas, monitoring disease and epidemics to mention few.

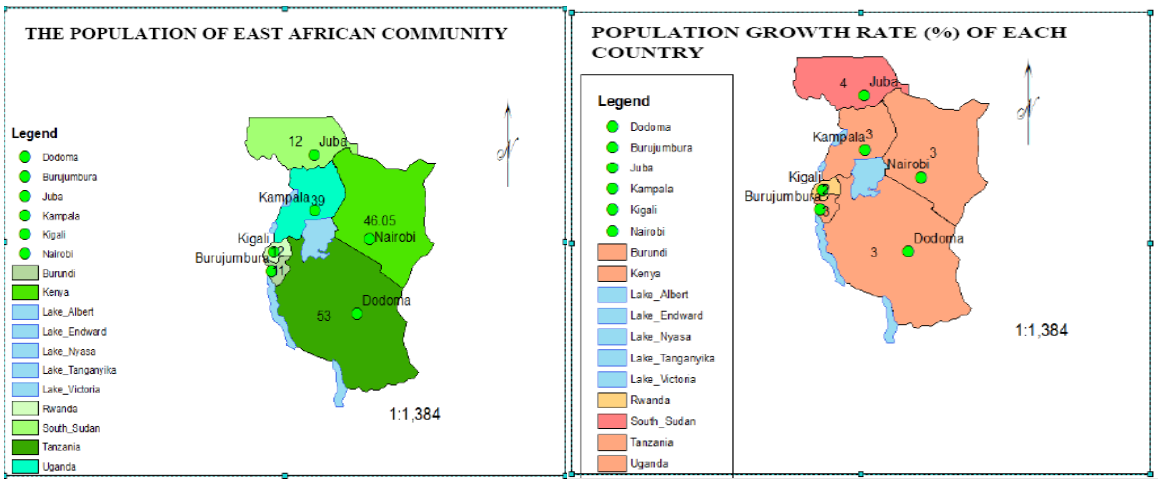

Figure 2: The population of EAC countries and its growth rate, correspondingly

3.2 Challenges for spatial and statistical data integrations in EACs countries

Unwatched production of spatial data result much into more duplication of data, poor data quality control, inefficient use of the available data resources, suppression of the geo-information market and frustrated data users [4]. In the previous, lack of common frame work and standards were reported to be among the challenges hindering data integrations however in the recent a GIS framework has been developed for the most seven priority layers. In connection to that, in the previous issue of metadata interoperability was considered as among of constraint toward data integrations however the GIS framework have also highlighted issue related to metadata. Thus, updating the existing spatial data according to the GIS framework will fast truck the pace towards the spatial and statistical data integrations in EACs countries. Absence of clearing house for EACs countries is another challenge; clearing house is the form of geo-portals and catalog consisting of collection of information resources. In the EACs countries despite of millions of geographical data producers, vendors and users there is no data clearing house from either of the country. Elsewhere in Europe using these emerging technologies, INSPIRE was developed to create a European Union spatial data infrastructure for the purposes of EU environmental policies and policies or activities which may have an impact on the environment. This European Spatial Data Infrastructure enables the sharing of environmental spatial information among public sector organizations, facilitate public access to spatial information across Europe and assist in policy-making across boundaries [9]. Thus, a lesson learned from the European Union Spatial Data Infrastructures which was established and operated by the Member States of the European Union, calls attention to researchers, government organization and private stakeholders toward developing a Spatial Data Infrastructures for the EAC country members. Lack of common geographic boundaries and data management systems also has hindered the statistical and spatial data integration in EACs countries. In context of geographic boundaries, the EACs countries have different geographic boundaries. Whilst data management in the EACs countries differs, there is need for establishing geo-coding which is a process is the process of assigning locations to addresses to that they can be placed as points on a map, similar to putting pins on a paper map and analyzed with other spatial data. Establishing geo-coding in the EACs countries in addition for data integration also it will facilitate easy accessibility and sharing within and between countries. Lack of common standards & guidelines of integration

of statistic and geospatial data is another devastating challenge in implementation of spatial and statistical data integration. Although there is well established GIS common framework consisting of the seven thematic layers, still there is not well established guidelines for integration of statistic and geospatial data in the EAC countries. Thus during the implementation of NSDI each country government should collaborate effectively in collaboratively form policies that will accommodate all matters related to spatial data standards and guidelines covering aspects related to confidentiality and privacy, data quality, analysis, copy right, pricing, dissemination and visualization. On the other hand, lack of adequate expertise and skilled labor for data integration, shortage of integrated infrastructure to support the integration initiatives of spatial statistical data, inconsistencies among the structure of the data, lack of regional level and well-defined frameworks that are governing type of data integration are also among of the major challenges of data integration in developing countries.

4. Conclusion and Recommendations

Basing on the literature review of the available documents, both opportunity and challenges confronting the spatial and statistical data integration do exist in the EAC countries. However, to address the challenges collaborative effort from public and private sectors is deemed necessary to tape the available opportunities for integrating spatial and statistical data in EAC countries. In particularly the EAC country governments need to support development of their own NSDI with subsequent focus of developing the EAC Country Spatial Data which are interoperable to country statistical data. Volunteered Geographic Information (VGI) where volunteers use the web and mobile devices to create, assemble and disseminate spatial information should be well promoted among geographical data producers, vendors and users in EACs.

References

[1]. I. Masser. “GIS Worlds: Creating Spatial Data Infrastructures”. ESRI Press, 2005. p.17.

[2]. A.R.Stevens, H.J.Onsrud and M.Rao. “Global Spatial Data Infrastructure (GSDI): Encouraging SDI Development Internationally”, presented at ISPRS Workshop on Service and Application of Spatial Data Infrastructure, Hangzhou, China, 2005.

[3]. G, Foody L.See, S.Fritz, P. Mooney, A. M, Olteanu-Raimond, C.C. Fonte, and V. Antoniou. (eds.) “Integrating Spatial Data Infrastructures (SDIs) with Volunteered Geographic Information (VGI) for creating a Global GIS platform.” Mapping and the Citizen Sensor, 2017. pp. 273–297.

[4]. G. C. Mulaku and N. C Siriba. “Building Geospatial Data for Multiple Applications: The Role of Standards”. In Proc 5th AARSE CONFERENCE, October, 2004.

[5]. W. Kalande and J.D.Ondulo. “Policy and Strategy: Geo-information Policy in East Africa” 2006.

[6]. J.Okuku, A.Bregt and L.Grus. “Assessing the Development of Kenya National Spatial Data Infrastructure (KNSDI)”. South African Journal of Geomatics, 2014, Vol. 3.

[7]. M.F. Goodchild. “Citizens as sensors: the world of volunteered geography”. GeoJournal,2017 Vol 69, pp 211–221.

[8]. Worldometers. Population in East Africa. Internet: http://www.worldometers.info/world population/rwanda-population. 25th December, 2017.

[9]. INSPIRE. Data integration. Internet: https://inspire.ec.europa.eu/about-inspire/563. 19th January, 2018.