Journal of Sustainable Development in Africa (Volume 12, No.4, 2010)

ISSN: 1520-5509
Clarion University of Pennsylvania, Clarion, Pennsylvania

Geospatial Quality Data Acquisition Problems in Sub-Saharan Africa

A. Ayanlade, O. Babatimehin, M.O Olawole and Orimogunje, O.O.I

ABSTRACT

This paper discuses problem of data acquisition in Sub-Saharan African countries. The purpose of any Geospatial research work is to provide information which will assist in quick decision making and effective management. For this reasons, effort is now in place to assess in quality of the data acquisition and derived information products in Sub-Saharan Africa. Meanwhile, in spite of these, there is no such thing as an "exact" measurement. All measurements being estimate of the true 'point' value. Many data, especially, rainfall data, therefore contain some degree of error and the extent of data accuracy or precision can only be specified in relative terms. Data acquisition, using some old equipment available in Sub-Saharan Africa, is samples in space and time which are subject to various errors, notably; observer error and instrumental error. The major aim of this paper, therefore, is to depict Geospatial quality data acquisition problems in Sub-Saharan Africa. Generally, the objectives of this paper is to: (1) examine the various aspects of spatial data acquisition problems in Sub-Saharan Africa; (2) discus recent development in Geospatial data acquisition in Sub-Saharan Africa; and (3) explain the concept of metadata in data acquisition reform in Sub-Saharan Africa. Rainfall data acquisition was used as a case study for this region, which was part of the findings from our resent research work. It was observed that some organizations and research institutions are still employing traditional means of data acquisition despite the recent innovation in Geoinformatics and Remote Sensing field.

Key Words: Geospatial Data, GIS data acquisition, Data quality problems, Rainfall Data, Sub-Saharan Africa

INTRODUCTION

The purpose of any Geospatial research work is to provide information that will assist in quick decision-making and effective management. GIS Data acquisition, and the maintenance of databases, remains the most expensive and time consuming aspect of setting up a major GIS facility in sub-Saharan Africa. This typically costs about 62-81% of the overall costs of a GIS project. Since GIS may be used for applications as varied as climate change/variability analysis, biodiversity analysis, archeological analysis, marketing research, and urban planning, among others, the source data can be difficult to inventory and classify comprehensively. Even within a single GIS project, the range of data employed can be daunting. Consider, for example, a list of the sources that were used in Guinea Savanna Ecological Zone of Nigeria to develop the database for the Rainfall Variability Impact on Crop Yield. Meanwhile, in spite of these, there is no such thing as an "exact" measurement, all measurements being estimate of the true 'point' value. Many data, especially, rainfall data in Sub-Sahara Africa therefore contain some degree of error and the extent of data accuracy or precision can only be specified in relative terms. Data acquisition, using some old equipment, are available in Sub-Saharan Africa and are sampled in space and time, which are subject to various errors, notably; observer error and instrumental error.

For this reasons, effort is now in place to assess in the quality of the data acquisition and derived information products in Sub-Saharan Africa. Rainfall data acquisition in Guinea Savanna part of Nigeria was used as a case study for this study, which was part of findings from our resent research work. The major aim of this paper, therefore, is to depict Geospatial quality data acquisition problems in Sub-Saharan Africa. Generally, the objectives of this paper are to: (1) examine the various aspects of spatial data acquisition problems in Sub-Saharan Africa; (2) discus recent development in Geospatial data acquisition in Sub-Saharan Africa; and (3) explain the concept of metadata in data acquisition reform in Sub-Saharan Africa.

Geospatial Data Acquisition in Sub-Saharan Africa: Past and Present

Not so long ago in Sub-Saharan Africa, most Geographical Information Systems (GIS) and Cartographic projects had to rely almost exclusively upon data available only in printed or "paper" format. Although, much of the data available for use is still published on paper, a great deal of information is now distributed in digital formats. The ever-increasing pace of this transformation from paper to digital sources has many repercussions for GIS. Data already produced in digital format will certainly ease the work and speed the process of developing GIS, but only if users learn how to employ these new sources effectively. The facts to note here is that all data sources have strengths and limitations. Digital sources are no different. It is imperative to understand their characteristics, costs, and benefits before using them. Learning a little about commonly employed digital formats will save much work in the long run.

In Nigeria, for instance, local, state, and federal government agencies are major providers of data. A good deal of detective work is sometimes required, however, to find the data you need. This is, perhaps, less the case at the federal level if only because certain key agencies, such as the Federal Office of Statistic (FOS), Federal Ministry of Environment, Nigerian Meteorological Services, Oshodi, Lagos, and Federal Emergency Management Agency (FEMA), provide standard sorts of information for the entire nation. A good example is word-class map that is subsidiary of Classic Systems Ltd in Ile-Ife, Nigeria. Commercial mapmaking firms are among the largest providers, but other firms have, for years, supplied detailed demographic and economic information, such as data on retail trade and marketing trends. Some of this information can be quite expensive to acquire. Also, it is important to check on restrictions that might apply to the use of commercially provided data. In some cases, copyright and licensing restrictions may apply to your intended use and publication of the information.

The digital data revolution also means that users must often search for data in different places. To acquire some digital sources, users must contact the producers directly and work closely with them to gain the necessary data in a useable format. Also, the Internet and Worldwide Web are being used to distribute data to a greater extent and this means knowing where to look and how to search the networks. The Internet is currently the medium to disseminate Geodata and maps, which is part of the effort being made in the Department of Geography, Obafemi Awolowo University, Ile-Ife, Nigeria.

Spatial Data Acquisition Problems in Sub-Saharan Africa

One of the major data acquisition problems in Sub-Saharan Africa includes instrumental errors, which are associated with the functioning of the instruments. An error frequent with the rain gauges is during the siphoning cycle, where the rain persists to enter the rain gauge. Also, it was discovered that up to 0.25 mm of rainfall might be required to moisten the funnel and inside surfaces of an initially dry rain gauge. These values may cumulate annually to more than 25 mm.

Observer errors arise from the limitation of human expertise, especially error in the reading of the rain gauge measuring cylinder and chat. Observational errors are usually small, but may be cumulatively large. For instance, the light amount of rainfall, usually written as trace {'Tr'} in the rainfall recording charts in most of the rainfall stations in Nigeria and may sum up to more than 25mm on an annual basis. Other sources of observer error that is common, but not evident in the rainfall data of Nigeria include: transcription errors, reading of rainfall data to the nearest say 0.1 mm, and personal idiosyncrasies.

The heterogeneity of Georeferenced Digital Libraries (GDL) is a serious problem for researchers who need to query several GDLs to find the best spatial data available for their projects. Differences in content, standards, user interfaces, semantics, database structure, etc. are the rule on the Internet. In addition, users have no tool to help them select the best sources of spatial data once they have clearly defined their needs and have found several potential sources in the GDLs. Data warehousing technology, coupled with a data transformation/integration tool plus a data selection layer running on the Internet, appears to be a promising solution for such problems of heterogeneity and best selection. These facts and statistics, in addition to similar ones found in the survey, clearly show the heterogeneity of GDLs. As a result, in order to retrieve the desired information, users have to understand every interface and process. In a way, such a problem is analogous to the problem of finding information on the Internet using different search engines, but it adds a strong geospatial consideration and new user interface considerations.

Over the past ten years, the internet and web have become the leading means of acquiring primary and secondary data. In some cases, you must still contact providers in person or by mail to obtain data, but this is changing very rapidly. The is, now, the primary means of disseminating data produced by the federal government and is assuming that role for most state governments. At the moment, finding exactly what you are looking for on the Internet can be difficult, but navigation and focused searching of the networks are becoming easier.

METADATA IN DATA ACQUISITION REFORM IN SUB-SAHARAN AFRICA

Metadata (which is defines as Data about data) are very important. Metadata allows us to understand the origin of the data, its geometric characteristics, its attributes, and what type of cartographic, digital image processing, or modeling has already been applied to the data. The Content Standard for Digital Geospatial Metadata is now in place and there are working groups focused on how to improve the standard (FGDC, 1997b; FGDC, 1998). Additional research should continue on: a) how to organize, store, and serve metadata using regional National Geospatial Data Clearinghouse (NGDC) nodes; b) development of improved web-based interfaces for efficiently browsing and downloading metadata; and c) documenting the genealogy (lineage) of all of the operations that have been performed or applied to a dataset (Lanter and Veregin, 1992). A user must have a complete understanding of the content and quality of a digital spatial dataset in order to make maximum use of its information potential.

There are a number of issues which arise when developing a metadata for climate change/variability projects in Guinea Savanna part of Nigeria. The first issue is to determine how the data will be stored, either inform of vector or raster format. Considerations here include:

- the nature of the source data .
- the predominant use to which it will be put
- the potential losses that may occur in transition
- storage space (increasingly less important)
- requirements for data sharing with other systems/software

If documentation is limited, it is important to consider the following questions:

- What is the areal coverage of the data?
- To what map scale was the data digitized?
- What projection, coordinate system, and datum were used in maps?
- What is the age of the data?
- Where did it come from?
- In what medium was it originally produced?
- How accurate are positional and attribute features?
- Does the data seem logical and consistent?
- Do cartographic representations look "clean?"
- Is the data relevant to the project at hand?
- In what format is the data kept?
- How was the data checked?
- Why was the data compiled?

As a general rule, it is best to retain the maximum amount of information in the metadata environment. If the data is available as points, lines, or polygons then it should be kept that way. If a raster approximation of this data is also needed for analytical purposes, then a raster version of the data may be kept in addition to the vector coverage. Many systems provide from quick conversion from vector to raster. The issue of scale is often raised in relation to GIS data base development. It is important to remember that data stored in a GIS does not have a scale. Sometime people refer to a 1:25000 scale data base. What they mean is that the data has been taken from 1:25000 maps or that it has a level of accuracy which is roughly equivalent to that found on 1:25000 scale maps.

Global Positioning System (GPS) Data Collection

Geospatial analysts, the government organizations, NGOs, and surveyors are making increasing use of a GPS to collect x, y, z coordinate information (Lapine, 1989). The former Director of the National Geodetic Survey (NGS) and now Chief of the S. C. Geodetic Survey identifies the concerns that must be addressed by government agencies, private industries, NGOs and research institutes to improve our GPS data collection potential for GIS exterts and other users (Jensen et al, 1998). Real-time differential horizontal (latitude, longitude) data collection can achieve operational goals of 1-3 m for general GIS data collection.

The actual problem is the precision of the vertical (z) measurement. The goal is to acquire vertical values comparative to the classic vertical network of 3 cm using post-processing techniques or 1-3 m in real-time. The current state of the art is about 10 cm with post-processing and 10 m in real-time, thus, this is improper for most GIS works. However, it is possible to postprocess the vertical data to obtain 2-5 cm accuracy using prototype techniques pioneered by NGS. The South Carolina Geodetic Survey is working with the NGS to develop the GPS techniques for obtaining operational procedures for an accuracy of 3 cm. These techniques may provide the solution for improved real-time accuracy as well.

An important new finding is that this same network of differential GPS may be of significant value for real-time weather prediction. The ionosphere refraction measured by the dual frequency GPS receivers is capable of identifying the precipitable water vapor concentration. This is the most significant variable in weather prediction. The 6 and 24 hour forecasts could be improved, significantly, by a more accurate measurement of precipitable water vapor. The receivers may be placed at every airport in the nation, dramatically increasing the precision of our national weather prediction capability and simultaneously providing a more dense network of base stations for the real-time GIS users.

In order to solve the problems of data acquisition quality in Sub-Saharan Africa, the following issues most be properly considered.

Encourage Data Acquisition Standards: Many organizations and data users have developed and promoted standards for spatial data collection and representation. Good examples of such include FGDC, Open GIS Consortium, and ISO. Good summaries are found in NAPA (1998). In the United States, the Federal Geographic Data Committee (FGDC) oversees the development of a National Spatial Data Infrastructure (NSDI). The UCGIS research community endorses the significant strides made by FGDC to establish and implement standards on data content, accuracy, and transfer (FGDC, 1997). FGDC's goal is to provide a consistent means to directly compare the content and positional accuracy of spatial data obtained by different methods for the same point and thereby facilitate interoperability of spatial data. UCGIS and other scientists should assist Sub-Saharan Africa in determining the impact on data collection if and when businesses and organizations implement international environmental standards as prescribed by the International Organization for Standardization (ISO). With the Standards Data Quality Solution, organizations can incorporate data quality business rules across data sources and platforms. End-to-end audit capabilities highlight system-related quality issues by applying data quality algorithms and analytics. Other benefits of Data Acquisition Standardization are discussed below:

- **Easily define data correction rules** to reflect organizational changes and cleanse data. Specialized interfaces enable business users to create and visualize the impact of business rules and data cleansing efforts. State-of-the-art data quality tools enable both business and technical users to cleanse, standardize, integrate, and augment data.
- **Provide decision makers with information they can trust.** Accurate data delivers consistent reporting and analytic results. By integrating data quality automatically into the metadata and data integration processes, data used for business intelligence and analytics will be current and accurate. As a result, researchers and decision makers can trust the information they receive and make decisions with confidence.
- **Internationally capable.** The solution understands the differences in various languages, including names, addresses, organizations, and other areas where data parsing is required.
- **Match code generation** consolidates data from multiple data sources when a complete match is not possible. The solution also creates unique key values by using fuzzy logic to group together information with similar values.

- **Data cleansing, reduplication and standardization** capabilities provide the ability to eliminate or reduce inconsistencies in corporate data.

While important suggestions have been made, there are still unresolved issues that need to be investigated, including: 1) the determination of error evaluation sample size based on map or image scale and other relevant criteria, 2) identification of the most unbiased method of allocating the test sample data throughout the study area (e.g. by line, quadrant, stratified-systematic-unaligned sample, etc.), 3) development of improved methods for reporting the positional accuracy of maps or other spatial data that contain multiple geographic areas of different accuracy, 4) development of more rigorous criteria to identify coordinate 'blunders', and 5) development of improved statistical methods for assessing horizontal and vertical positional accuracy.

CONCLUSION

We have presented a solution to the problems of quality data acquisition in Sub-Saharan Africa. We have discussed the current issues in Geospatial data collection in the area using rainfall data acquisition in Guinea Savanna, part of Nigeria, as a case study. The experience we had during our previous research projects was used as a base for this paper and we have experimented with several issues dealing with spatial metadata, semantic analysis, selecting the best source of data, data quality analysis, data modelling, and GIS. When developing a Geospatial data set for specific purposes, there are a number of design considerations. These include: the physical extent of the data base; the resolution (grid size); the themes to be included; the classifications to be used within the themes; and the appropriateness of scale of input data to the preferred grid size. A lot of time and money can be wasted by seeking to build a data base which incorporates all known information about an area. It is, first, appropriate to determine what questions the GIS will be required to answer and what data is needed to answer those questions. For example, while geological maps of an area may be available they may be of no relevance to the specific decision process. The higher the resolution, the better the approximation of reality - provided the data is good enough to support this resolution. If you assume a certain error in map preparation and digitization, then this translates to certain on ground error. There is little point in making the grid size smaller than the probable error. A smaller than necessary grid size leads to larger files and longer processing times. In the best case, halving the cell size will quadruple the processing time. Experience with raster processing suggests that more than 2 million cells are excessive in most contexts. The size of the pixel must be half of the smallest distance to be represented. It is important to know that the range of analytical or modelling operations, which are available, may be limited by the type of data measure (scale) being used.

REFERENCES

FGDC. (1995). Development of a National Digital Geospatial Data Framework. Washington, DC: Federal Geographic Data Committee, (ftp://www.fgdc.gov/pub/standards /refmod.txt).

FGDC. (1997a). Framework: Introduction and Guide. Washington: Federal Geographic Data Committee, 105.

FGDC. (1997b). Geospatial Metadata. Washington: Federal Geographic Data Committee, 2.

FGDC. (1997c). Draft Geospatial Positioning Accuracy Standards. Washington: Federal Geographic Data Committee, 50.

FGDC. (1998). The Value of METADATA. Reston: FGDC Secretariat, 4.

Goodchild, M. (1997). Rasters. NCGIA Core Curriculum in Geographic

Information Science. Unit 055, URL:
http://www.ncgia.ucsb.edu/giscc/units/u055/u055.html.

Jensen,J., Saalfeld, A., Broome, F., Price, K., Ramsey, D., & Lapine, L., (1998) SPATIAL DATA ACQUISITION AND INTEGRATION, URL: http://www.ucgis.org/priorities/research/research_white/1998%20Papers/data.html

Kufoniyi, O. (1995). Spatial Coincidence Modeling, Automated Database Update and Data Consistency in Vector GIS. (Unpublished Ph.D thesis), Department of surveying and Remote Sensing, Wegeningen Agricultural University, The Netherlands ITC Publication 28, pp 203.

Kufoniyi, O. (1997). TOWARDS 3D GIS FOR EFFICIENT MANAGEMENT OF URBAN ENVIRONMENT.

Kufoniyi, O. (1998). Aspects of Institutional Framework for the Implementation of GIS in Nigeria. Workshop paper presented at the 1998 Survey Training Conference, Bauchi. June 30 – July 2, 1998. 4.

Kufoniyi, O. (2003). Proposed National Geoinformation Policy: Executive Summary, Proceedings of the National Geospatial Data Infrastructure. NARSDA, Abuja, 90.

Kufoniyi, O. (2004). Development of Geospatial Data Infrastructure in Nigeria. A Paper presented at the National Workshop on Satellite Remote Sensing (Nigeria Sat – 1) and GIS: A solution to sustainable National Development Challenges. 15th – 17th June, 2004, Le Meridien Hotel Abuja. 4.

Kufoniyi, O. & Agbaje, G.I. (2005). National Geospatial Data Infrastructure Development in Nigeria: the journey so far. A paper presented at From Pharaohsto Geoinformatics FIG Working Week 2005 and GSDI – 8 Cairo, Egypt. 3.

Lanter, D. P. & Veregin, H. (1992). A Research Paradigm for Propagating Error in Layer-based GIS. Photogrammetric Engineering & Remote Sensing, 58(6),825-835.

Lapine, L., (1989). Correspondence, Columbia: South Carolina Geodetic Survey

NAPA. (1998). Geographic Information for the 21st Century: Building a Strategy for the Nation. Washington: National Academy of Public Administration, 358.

ABOUT THE AUTHORS:

A. Ayanlade, O. Babatimehin, M.O Olawole and Orimogunje, O.O.I
Department of Geography
Obafemi Awolowo University, Ile-ife, Nigeria