Digital Soil Assessments and Beyond – Minasny, Malone & McBratney (eds)

The challenges of collating legacy data for digital mapping of Nigerian soils

Inakwu O.A. Odeh
Faculty of Agriculture and Environment, The University of Sydney, Australia

Johan Leenaars
ISRIC, Wageningen University and Research, Wageningen The Netherlands

Alfred Hartemink
Department of Soil Science, University of Wisconsin, Madison, US

Ishaku Amapu
Department of Soil Science, Ahmadu Bello University, Zaria, Nigeria

ABSTRACT: It is well known that the world is rich with soil data resulting from the efforts of past soil surveys. However, the majority of these data is largely unused. These legacy soil data come in the form of soil maps, soil survey reports, soil profile descriptions, and/or in the format of antiquated "card catalogue". These data provide many benefits including soil information preservation or as a major component of national environmental monitoring or as the only source of meaningful soil information for Digital Soil Mapping (DSM) in many of data-scarce countries of the developing world. These benefits call for the rejuvenation of these data. In Nigeria, as one of the data-poor countries in terms of modern soil digital soil databases, there is a preponderance of largely untapped legacy soil data as a result of her rich history of soil surveys. This paper provides an account of the challenges faced in the course of the renewal of the Nigerian legacy soil data for the purpose of DSM. In telling this account, we briefly review the processes involved. We then explore the historical soil surveys in Nigeria leading to the synthesis and discussion of the challenges of the processes used for the capture and renewal of the legacy soil data including how the challenges were overcome.

1 INTRODUCTION

Evidently the world is rich with soil data resulting from the efforts of past generations of soil surveyors (Mayr et al., 2008), and majority of these data is largely unused (Rossiter, 2008). These legacy soil data come in the form of soil maps, soil survey reports, soil profile descriptions, or in antiquated "card catalogue". These data have many benefits to offer, if not for the sake of mere soil information preservation (Rossiter, 2008). Legacy soil data is a major component of national environmental monitoring programs based on modelling temporal trends of soil and soil processes (Baxter, et al., 2006), therefore providing new vistas to learn from the past and to work sustainably in the present and predict the future for better management of soil resource.

In the last 30 years or so, new technologies have been developed to quantitatively map the soil in the form of soil classes or soil properties. This approach is now widely known among soil scientists as digital soil mapping (DSM). Based on the definition by Lagacherie & McBratney (2007), DSM constitutes the processes involved with "the creation and population of spatial soil information systems by the use of field observational methods coupled with spatial and non-spatial soil inference systems". As has been demonstrated by many studies (e.g., Bui & Moran 2001; McBratney, et al, 2003; Mayr et al., 2008), soil legacy data have a major role to play in DSM, especially in the case of the soil data-poor countries.

Nigeria, as one the data-poor countries, is rich in soil surveys since the colonial time, especially post-colonial times (FDALR, 1990). There are currently known hundreds of soil surveys at various scales—ranging from farm levels to national scale. The challenge is how they can be captured and transformed into databases that may be used for temporal trend analysis and for a digital soil mapping program. To achieve these goals is challenging because of inadequate or a lack of georeferencing of soil records and the inconsistency in the laboratory and mapping methods, taxonomy, legends, soil survey reports and the deterioration of paper hardcopy soil maps. The aim of this paper is to provide an account of the processes and challenges faced in the course of renewal of the Nigerian legacy soil data for the purpose of DSM. In telling this story, we start by briefly reviewing the processes involved with capturing and renewing the legacy data.

2 PROCESSES INVOLVED IN LEGACY SOIL

Rossiter (2008) categorized the main processes leading to rejuvenation of historical soil data as i) *data archaeology* which involves effort to locate and catalogue the paper hardcopy historical soil records, including soil reports and maps; ii) *data rescue* involving scanning of the historical records and storing them in appropriate, easily-accessible media such as DVD or Web publications; iii) *data renewal* or *data resurrection* (Dent and Ahmed, 1995), which entails the transformation of the rescued data into usable digital (GIS) formats.

Although data archaeology and rescue are important precursor to data rejuvenation, the focus here is on data renewal. The latter comprises several sub-processes (Rossiter 2008), among which are: i) the development of soil information systems and databases populated by soil site observations, including profile description and physical and chemical soil properties—the reliability of this process is of course linked with the accuracy of georeferencing of the site observations, which can be challenging in some situations; ii) the creation of a relational database of soil class polygons (mapping units) linked with their generic values of soil properties (Mayr et al., 2008). While all of the processes listed are important, this paper is focused on the challenges of tackling (i) for the Nigerian scenarios. But first we explore historical soil surveys in Nigeria.

3 THE LEGACY OF SOIL RESOURCE INVENTORIES IN NIGERIA

Nigeria is relatively rich in soil surveys—with their gestation started in the late colonial to early post-independent period. During this period, a few farm level and regional-scale soil surveys were motivated mainly by economic incentives of cash crop returns, with less of them initiated by scientific and/or environmental interests (Okoye, 1992). The first provisional soil map of Nigeria was published by Survey Department, Lagos in 1952. This was followed by the publication of the CCTA soil map of Africa (FAO, 1964) and FAO Soil Map of the World (FAO-UNESCO, 1974). Since then there was hardly any serious efforts towards a national-scale soil inventory until the late 1980s. After independence in 1960, succeeding governments became increasingly interested in national food production self-sufficiency; although regional efforts towards providing soil suitability for various crops did not start until in the mid-1960s. During this period there were projects geared towards large agricultural projects such as the Sokoto-Rima Basin irrigation projects (published by the Northern Nigerian Govt. and UNDF/FAO in 1969). Other post-independence surveys were at the instances of foreign aids as exemplified by United Kingdom's Overseas Development Administration. It was not until the late 1970s that the Federal Department of Agricultural Land Resources, with assistance from USDA initiated a national soil inventory project to produce "a systematic and correlated soil map of the country which would provide a good guide for agriculture and other land development" (FDALR, 1990). Thus broadly, historical soil surveys in Nigeria can be grouped into five main categories based on their scales:

1. Very detailed or detailed soil survey, at the scale of 1:≤25,000;
2. Semi-detailed soil survey, usually at a range of scales between 1:25,000 to 1:<50,000;
3. River basin land system survey, at scales between 1:50,000 and l1:<100,000;
4. Reconnaissance soil survey at scales 1:100,000 and usually up to 1:500,000;
5. Exploratory soil survey at scales smaller than 1:500,000 and smaller than 1:5000,000.

Of these groups of soil surveys/maps in Nigeria as documented by ISRIC (per. comm., 2011), over a third (35%) are very detailed or detailed, with a further 15% semi-detailed. Another third (about 33%) are at an exploratory or a regional level (i.e., at scales smaller than 1: 500,000). A further 16% were at the exploratory level—at 1:>1:500,000, covering very large political regions or at the national level. The most recent national soil map, at the scale of 1:3,000,000, was compiled by Sonneveld (1996) from the pre-1985 soil maps. However, it did not incorporate the soil profile data from those surveys, required to be exploited for DSM. Most importantly, the richness of soil survey outputs also decreases as the scale of the soil survey becomes smaller. For example, while most of the very detailed to semi-detailed soil surveys have many soil morphological observations and a large number of point laboratory measurements most important for DSM, these becomes limiting as the scale of the surveys becomes smaller. This is with the exception of the national-scale FDALR survey reports and maps (FDALR, 1990), which provide a wealth of legacy soil data and information.

A relevant observation indicates an apparent temporal trend with most of the very detailed to semi-detailed surveys carried out pre-1985.

4 THE CHALLENGES IN ACQUIRING AND RENEWING LEGACY SOIL DATA FOR NIGERIA

The need to capture and rejuvenate the relevant legacy soils data for Nigeria in digital form is part of the African Soil Information Services (*AfSIS*) project, a component of a global consortium- the *GlobalSoilMap.Net*. The latter is composed of the seven or so nodes, of which *AfSIS* is one. Thus, legacy soil data are identified as one of the scientific priorities for the operational production of *GlobalSoilMap.net* outputs (published by *GlobalSoilMap.net* Consortium in 2011- held at ISRIC). On top of the priorities of the *AfSIS* activities are the sourcing, capturing and rejuvenation or renewal of legacy soil data which are standardized and/or transformed for DSM. In 2009, a working plan was formulated for the implementation of these activities for the African node. Nigeria, the most populated nation in Africa, is one of the first targeted countries for the actualization of these activities.

The work plan set out the following tangible goals with each having targeted outcomes: i) *Define data and information requirements* involving development of target data model, metadata and database to meet these requirements; ii) *Data rescue*- identification of data sources and data cataloging; iii) *Data renewal or resurrection*- transformation of the rescued data into digital form useful for DSM through their collation and harmonization into standardized database for *DSM*. Most of these goals variably have their limitations and challenges which are highlighted below, vis-à-vis the Nigerian experience.

4.1 *Define data requirements for DSM*

The data requirements for DSM are documented in the *GlobalSoilMap.net* specifications. The most pertinent requirement for the legacy soil data is the vertical dimension or depth. This specification requires that the "depth of soil for which data will be reported" is 2 m. In practice, most soil surveys rarely provide data below 1.0–1.5 m hence there will be the need to extrapolate below these depths. The other requirement is the need to map nine primary and four derived soil properties to be mapped at each of the six depth intervals. This implies that the legacy data for the nine (non-derived) properties should be available in at least three horizons to the depth of not less than 30 cm for them to be useful for DSM. As the soil depth specification requires profile data to bedrock or 2 m, whichever is shallower, the requirement poses some challenges which need to be tackled.

4.2 *Develop target database model*

The database model needs to be defined to serve as a guide and to facilitate a smooth workflow for the consistent compilation of the legacy data. Prior to the development and implementation of a database model, we carried out a preliminary study of data models of existing major legacy soil profile data including: CANSIS, USDA NCRS SSL, CSIRO ASRIS, NATSOILS, AFSIS 1st template Objective 2, INRA DONESOL, IRD SOLTUNESIE, EU SPADE, FAO SDB, ISRIC ISIS, ISRIC WISE, ISRIC SOTER, ISRIC SOTERML and ISO. Evidence indicates that these models are very diverse in content, schemas and their setup. Thus they pose considerable challenges to be standardized. It was therefore decided to proceed as pragmatically as possible and to create a simple database and metadata models. A critical assumption at this juncture was that the 'pragmatic' database would just be another database to be reconciled later with the data and information standards still in development. The schema of data tables was organized to avoid data redundancy and very much mirrors the principles that were assumed when developing the data entry template, reflecting the process of data entry as: i) source inventory (digital and/or analogue); ii) profile inventory; iii) profile layer property value entry; and v) profile layer property value harmonization and quality control; iv) metadata entry, aggregated at the appropriate level. This simplified database and the metadata models provide the basis for the compilation of the Nigerian legacy soil data as a subset of the *AfSIS* database.

4.3 *Data rescue: identification of data sources and data cataloging*

Having created a simplified database model, it was necessary to populate it with the legacy data. To do this would require rescuing the data from wherever they can be found. Finding and locating the sources of legacy data is the most challenging in the process leading to data renewal (Rossiter, 2008). This is even more challenging for Nigeria because as stated above there are hundreds, if not thousands, of soil survey reports in different locations of the country. These reports were produced by different institutions and organizations, including: i) Departments within a number of

Federal Ministries- variably include Agriculture, Environment and Natural Resources- example are the FDALR and National Program for Food Security. both now within the Federal Ministry of Agriculture; ii) Overseas Development Aid Organizations such as the British owned Land Resource Division of the Overseas Development Administration; iii) Former Regional Government Ministries of Agriculture and affiliated Institutions; iv) River Basin Development Authorities of which there are currently 12; v) Government-funded Research Institutions within or outside Universities, such as Institute for Agricultural Research(IAR), Ahmadu Bello University in Samaru Zaria, and the Institute of Agricultural Research and Training (IAR&T) at Obafemi Awolowo University Moor Plantation, Ibadan; vi) Individuals within Faculties of Agricultures in various Universities and Colleges of Agriculture, e.g., Federal College of Agriculture Umudike located near Umuahia; vii) Private sectors- large farm projects; reports indicate that only a limited number in this category have some form of soil assessment for their enterprises.

This long list of organizations/agencies as sources of legacy soil data epitomizes the widespread and dispersive nature of soil survey resources across the country. As reported by Okoye (1992), the whole of Nigeria has been covered by these surveys, although not as uniformly of adequate intensity as would have been optimal (Figure 1). Additionally, it would be difficult to extract the legacy data because different systems of survey were applied due to varied nature of the purposes of the surveys. This problem is exacerbated by the spread of data-holding institutions throughout the country as there are no central-holding institution(s). Therefore to identify and locate the data-holding institutions in Nigeria was not going to be an easy task. Fortunately most of the regional and reconnaissance survey reports and maps have been captured and catalogued. Unfortunately, majority of the legacy data from the more detailed soil surveys are still being held by the survey institutions scattered across the country. To locate these sources of legacy data has been most challenging. So how was this problem partially overcome?

The problem was solved by the combined exploration of the online holdings by various organizations- including ISRIC, FAO, and consultations with our Nigerian contacts- who could easily identify data holdings in their jurisdictions. An inventory of all known data holdings for Nigeria was first made and catalogued. We then consulted with our Nigerian counterparts to compare this list with the list of soil surveys in their area of jurisdictions. To eliminate duplications and minimize search for those data-holdings, we identified the soil surveys missing from the known list. Then the search for the missing surveys back in Nigeria began in earnest, through which 100 s of detailed soil surveys were identified. The analogue reports and maps of these surveys were collected and brought back to Wageningen for cataloguing. Thus the surveys currently catalogued at ISRIC are a combination of the previously held ISRIC database updated by our new effort. Through this effort, the number of soil profile data and observations was increased by ten folds, as will be explained later.

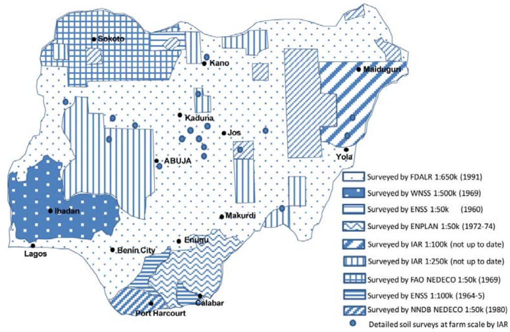

Figure 1. Soil survey coverage in Nigeria up until 1990s (Adopted and modified from Okoye, 1992).

4.4 *Data renewal or resurrection- transformation of the rescued data into digital form useful for DSM*

From the perspective of data transformation for DSM, there are different types of legacy soil data (Rossiter, 2008): i) polygon maps of soil types consisting of one-named soil class or multi-named soil classes; ii) point observations including descriptions of soil morphology, the immediate surface cover and location; iii) chemical and physical data obtained by laboratory analyses; iv) continuous field maps of soil classes or soil properties and v) geophysical surveys/maps. Each of these was available as analogue or digital or as point observations with physical/chemical data georeferenced. Thus the form of the legacy data would determine the degree of difficulty of transforming them into a usable digital form for DSM. In the case of Nigeria, soil data exist as in all five types, although not all of them were accessible and captured. Since the focus of this paper is on soil profile observations and chemical/physical point data, we describe here how these legacy data types were transformed for DSM. In most cases, the data may be accompanied by some form of variable georeferenced information useful for their renewal. The digitally-available legacy data are almost always georeferenced, and therefore are easier to deal with than the analogue data, because the former only need reformatting and transformation to match with the coordinate system of the project.

Thus, there are different kinds of legacy data defined based on the degree and accuracy of georeerenced information contained in the metadata or reports accompanying them:

i. Existing digital geodatabases- most of which were routinely georeferenced with known projections and datum;
ii. Analogue-to-digital legacy data that were georeferenced with known projections and datum;
iii. Analogue-to-digital legacy data that were georeferenced with unknown projections and datum not defined;
iv. Analogue-to-digital legacy data that were not georeferenced but have scaled soil survey maps, which may include soil sampling map accompanying the soil survey reports;
v. Analogue-to-digital legacy data that were not georeferenced, but have descriptive location information that could be used to gauge and refine the georeferencing;
vi. Analogue legacy data that have no georeferenced information or no accompanied sampling or soil scaled maps at all.

The first three types were the easiest to transform and only needed to be re-projected or converted from one coordinate system to the appropriate coordinate system and projection. However, types (ii–iii) require analogue-to-digital transformation before re-projection. Types (iv) and (v) required some additional effort during the data transformation process. The non-georeferenced data listed as (vi) were impossible to rescue, and hence were captured but not for DSM.

In each case of types (i) to (v), the metadata or the soil survey reports were the source of information for deciphering how useful a given set of soil profile data could be entered into the database or not. In deciding this there are number issues that needed to be resolved.

In the case of existing digital databases of profile data, rescue involved accessing and reading the digital data sets from CDs, DVDs or older media (e.g., tape drives, floppy drives, old computers) and transferring these digital files onto a new reliable central database.

Rescue of existing digital or paper data sets is not complete with just scanning. Rescue also must include the collection and input of metadata that describe the source and nature of the rescued data (report, map, database, legend) and the approximate extent and geographic location of the data.

In the case of analogue maps, data tables and maps rescue can be accomplished by means of scanning existing paper copies into a generic PDF

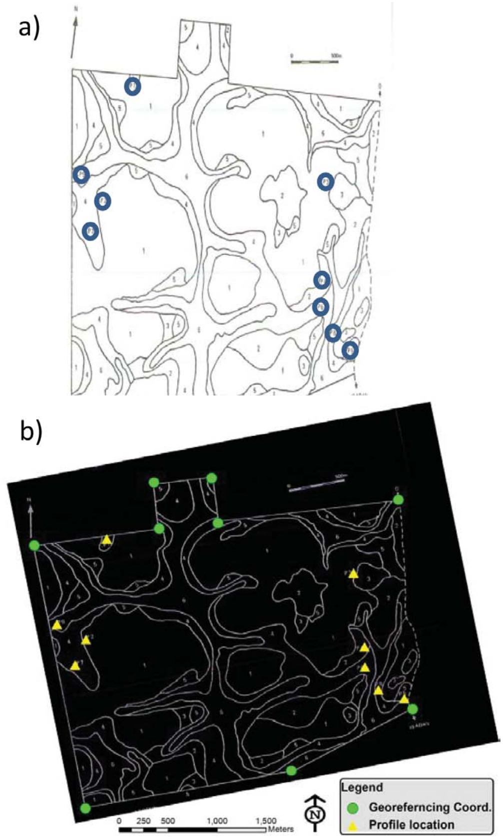

Figure 2. An example of georeferenced profile locations (open circles) from a scaled soil map (a) by the combined use of Google Earth and georectified scanned image of the soil map (b).

or image format (JPG or TIFF). Depending on the level of georeferencing as listed in (ii) to (v) above, we adopted any of the following methods to obtain the georefereing before conversion to the required coordinate system:

i. The character recognition software were used to convert metadata and data from scanned (JPG, TIFF or PDF) into machine readable information;
ii. Where necessary and the soil maps are scaled, we used Google Earth coordinates and GIS rectification and georeferencing tool to obtain the profile locations (see Figure 2);
iii. In the case of descriptive location obtained from soil survey reports, it was a lot more challenging to extract the approximate profile locations if the description is ambiguous, and in some cases vague. In situations where the description is less vague- such as directional

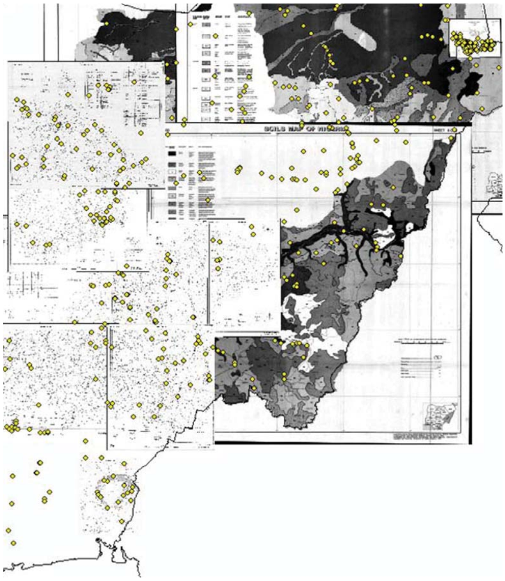

Figure 3. An example of georeferenced profile locations in northeastern Nigeria by combined use of Google Earth and mapping unit characteristics as auxiliary information. Note that the outcome here consists of several soil maps, produced by different agencies and organizations, which had been mosaicked.

description- for example, "12 km from Samaru along Samaru-Funtua road", we used the Google Earth to determine the approximate location of the profile;

iv. In the case of an ambiguous description of a representative soil profile of a mapping unit, we also used the combination of Google Earth and the soil map containing the mapping unit to obtain the approximate locations of the representative profiles (Figure 3) from the morphological description. This may be highly uncertain, but better than not having any data at all.

In spite of the steps above, there were some situations which did not lend themselves to an easy solution. For example, a directional description, such as "12 km from Samaru road" is difficult to decipher, as it is difficult to decide which side of the road and at what point along the road. We normally ignored using such profiles for DSM, as it would significantly increase the locational inaccuracy.

5 CONCLUDING REMARKS

In conclusion, the following salient points are worth noting:

- Through the concerted efforts of renewing legacy soil data for Nigeria, a total of 1220 soil profiles were built into a extractable database;
- Out of the 1220 profiles, there are between 600 and 1100 profiles are useful (depending on soil property) for DSM;
- Data on estimated >10,000 profiles may be available in an analogue format at various locations in Nigeria to be catalogued but the main constraints are bureaucracy, corruption and logistics in accessing the data in locations scattered across the country;
- The practicality of legacy data renewal through various stages and processes is highly labor-intensive. To achieve the renewal of even a fraction of what is available in Nigeria would require serious investments.

REFERENCES

Bui, E.N. & Moran, C.J. 2001. Disaggregation of polygons of surfacial geology and soil maps using spatial modeling and legacy data. Geoderma 103, 79–94.

Dent, D.L. & Ahmed, F.B. 1995. Resurrection of soil surveys: a case study of the acid sulphate soils of The Gambia. I. Data validation, taxonomy and mapping units. Soil Use and Management 11, 69–76.

FDALR, 1990. Soil map of Nigeria. Federal Department of Agricultural Land Resources, Kaduna, Kaduna State Nigeria.

FAO, 1964. Soil Map of Africa. C.C.T.A., Land and Water Development Division, FAO, Rome.

FAO-UNESCO, 1974. Legend of the Soil Map of the World. FAO, Rome, Italy.

Mayr, T.R., Palmer, R.C. & Cooke, H.J. 2008. Digital Soil Mapping using Legacy Data in the Eden Valley, UK.. In: A.E. Hartemink, A. McBratney & M.L. Mendonça-Santos (Editors), Digital Soil Mapping with Limited Data. Springer, pp. 291–301.

McBratney, A.B. & Lagacherie, P. 2007. Spatial soil information systems and spatial soil inference systems: perpectives for digital soil mapping.. In: Lagacherie P., A.B. McBratney & M. Voltz (eds.) *Digital Soil Mapping*. Elsevier, Amsterdam: pp. 3–22.

McBratney, A.B., Mendonça-Santos, M.L. & Minasny, B. 2003. On digital soil mapping. Geoderma 117, 3–52.

Okoye, E.O.U. 1992. Nigeria. In: Soil survey; perspectives and strategies for the 21st century. An International Workshop for Heads of National Soil Survey Organisations, November 23–15, 1992, Organized by ITC and FAO Enschede, The Netherlands.

Rossiter, D.G. 2008. Digital Soil Mapping as a Component of Data Renewal for Areas with Sparse Soil Data Infrastructures. In: A.E. Hartemink, A. McBratney & M.L. Mendonça-Santos (Editors), Digital Soil Mapping with Limited Data. Springer, pp. 69–80.

Sonneveld, B.G.J.S. 1996. Compilation of a soil map for Nigeria: a nation-wide soil resource and land form inventory In: Proceedings of the thirteenth West and Central African Soil Correlation Sub-Committee Meeting, Kumasi, Ghana, 11–15 November 1996.