Geoderma 192 (2013) 77–85

Contents lists available at SciVerse ScienceDirect

Geoderma

journal homepage: www.elsevier.com/locate/geoderma

Harmonizing legacy soil data for digital soil mapping in Indonesia

Yiyi Sulaeman [a,*], Budiman Minasny [b], Alex B. McBratney [b], Muhrizal Sarwani [a], Atang Sutandi [c]

[a] Indonesian Center for Agricultural Land Resource Research and Development, Jl. Tentara Pelajar 12, Cimanggu, Bogor, West Java, Indonesia
[b] Faculty of Agriculture, Food and Natural Resource, the University of Sydney, New South Wales, Australia
[c] Department of Soil Science and Land Resource, Bogor Agricultural University, Jl. Meranti 1, Kampus IPB Darmaga, Bogor, West Java, Indonesia

ARTICLE INFO

Article history:
Received 14 February 2012
Received in revised form 30 July 2012
Accepted 1 August 2012
Available online 16 November 2012

Keywords:
Legacy soil data
Equal-area spline depth function
Soil database
Soil assessment
GlobalSoilMap.net

ABSTRACT

Soil legacy data are basic input data for digital soil mapping. Any monitoring of global soil change cannot be successful unless soil data availability and characterization at the national level are well understood. Spatial data infrastructure frequently hampers many developing countries to take advantage of their soil legacy data for digital soil mapping at the national and regional level. While the initial purpose for developing a soil database is to store survey data, it cannot be efficiently used to support quantitative, digital soil mapping and assessment. This paper aims to develop a prototype soil observation database for global soil mapping in Indonesia. We will outline the steps needed to prepare legacy data for digital soil mapping, and describe the challenges and the Indonesian responses. The steps cover (i) legacy data identification and collection, (ii) data selection, (iii) database development and population, (iv) data harmonization and display, and (v) dataset integration. Because of uncoordinated and poorly-defined soil databases, we have to resort to a pragmatic approach to build a new and simpler database to support mapping activities. Historical soil survey reports and soil maps were collected, scanned, and summarized. After various considerations, we decided to only use soil profile observations which have clear geographical coordinates. We designed a database for soil profile observations and implemented it at two levels: spatial site data and horizon data. Spatial site data includes site geographical coordinates and attributes, while horizon data includes soil physical and chemical properties. The depths of soil profile database entries were standardized using the equal-area spline. Soil legacy data management for supporting digital soil mapping and GlobalSoilMap.net project are then discussed. These steps and data management are found helpful in Indonesia and this experience may be useful for other countries having similar impediment.

1. Introduction

The amount and type of legacy soil data determines the appropriate techniques that can be used for digital soil mapping (DSM) (McBratney et al., 2003). Legacy data may include soil maps with accompanying legends, or soil observations with site and horizon data. Minasny and McBratney (2010) suggested that the appropriate methods that can be used to develop digital soil maps and to obtain new soil samples depend on the availability and the type of legacy data. For example, Bui and Moran (2003) Kempen et al. (2009) developed different techniques to derive soil datasets for digital soil mapping purposes. Hence, the assessment of soil legacy data for a given region at any scale (global, continental, national) is required before legacy data can be used as input to digital soil mapping.

In Indonesia, soil resource inventories have been conducted since 1905 by Indonesian Centre for Agricultural Land Resource Research and Development (ICALRD) and its colonial and post-independence predecessors for various purposes (e.g. agricultural planning, erosion hazard assessment, and soil fertility monitoring). This has resulted in numerous soil survey reports and soil maps. Soil survey reports commonly store soil profile descriptions, i.e., soil morphology, and selected basic physico-chemical properties.

Various databases have been developed to store soil data in Indonesia. Since the 1990s, ICALRD has created and stored its soil profile data in a Site and Horizon Database (Pawitan et al., 1990) and SHDE4 (Wood-Sichra, 1995a). The soil analysis data are stored in the Soil Sample Analysis Database (Muslihat et al., 1990) and SSA3 (Wood-Sichra, 1995b). Meanwhile, land unit information is stored in a land unit database (Wijayanto and Wood-Sichra, 1990). Lindert (2000) developed another database that stores information on selected soil physical and chemical properties from the first and second layers of soil profiles collected from 1920 to 1990 for use in assessing soil changes with time. Soil survey activities are still ongoing and more data are collected and stored in soil survey reports and maps as well as in existing databases.

Soil survey activities in Indonesia are not well known internationally. For example, Hartemink (2008) based on Zinck (1990) reported that only 40% of Indonesia is covered by small scale soil surveys (1:500 000–1:100 000), 10% by medium scale (1:100,000–1:50,000)

* Corresponding author. Tel.: +62 251 8323012.
E-mail address: yies2001@yahoo.com (Y. Sulaeman).

0016-7061/$ – see front matter. Crown Copyright © 2012 Published by Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.geoderma.2012.08.005

and 0% by large scale (≤1:25 000). Since Zinck's report, the entire Indonesian land has been covered by a reconnaissance soil map at the scale of 1:1,000,000 (Puslittanak, 2000). This map has about 180 soil mapping units with 44 great groups from 8 orders of US Soil Taxonomy (Soil Survey Staf, 1975), i.e. Inceptisols (7), Alfisols (2), Ultisols (9), Spodosols (2), Oxisols (2), Histosols (4), Entisols (10), Mollisols (5), and Vertisols (3). As of 2010, 72% of Indonesian soil is covered by a 1:250 000 map and 21% by semi detailed maps (≤1:100 000) (Shofiyati et al., 2010). In addition, a land system map at the scale of 1:250,000 is available in the whole country (RePPProt, 1985). However, due to spatial data infrastructure issues, soil data in Indonesia are not widely known and accessible.

The existing soil databases are not well managed and become uncoordinated with various changes in ICALRD's structure over time (ICALRD has experienced 14 name changes since it was first established). The current soil databases use an obsolete data storage format so that there is an effort to migrate the data to a current computer operating system and newer database format. This becomes tedious as the old database can only be accessed by operators specifically trained on the software. As this effort is ongoing, soil survey reports and hardcopy maps become the final choice for soil research and assessment.

Legacy data rescue and renewal as suggested by Rossiter (2008) is an important step before any digital mapping can be considered. Lagacherie and McBratney (2007) suggested that the available soil legacy data must be well organized before applying any methodology for digital soil mapping. Therefore, to be a success, an appropriate database system must be developed to store the harmonized legacy data.

This paper addresses the issues of the available legacy data that may hamper developing countries becoming involved in DSM. We will review the steps needed to prepare legacy data for application in global soil mapping (GlobalSoilMap.net, Sanchez et al., 2009). We then address the challenges and the choices that are being made in Indonesia. The specific goals are to secure, harmonize, and summarize current soil observation data. Based on the data condition and database construction experiences, we discuss soil legacy data management and harmonization and its potential use for soil monitoring.

2. Preparing legacy data for digital soil mapping

The application of digital soil mapping (DSM) and subsequently fulfilling the GlobalSoilMap.net criteria requires representative and spatially distributed soil data. In most developing countries, we are only equipped with legacy data collected from different surveys at different time for different purposes. Here we present how Indonesia prepares such data for DSM requirement in several steps i.e.:

(i) Legacy data identification and collection,
(ii) Data selection,
(iii) Database development and population,
(iv) Data harmonization and display, and
(v) Dataset integration.

2.1. Legacy data identification and collection

Data identification and collection is a crucial step in DSM based on legacy data. In this step, the challenges are to answer the following questions: (i) what type of legacy data are available, (ii) how much legacy data are available, (iii) where they could be found, and (v) how to get them. These challenges can be addressed if there are properly recorded metadata. We went through documents that contained metadata about soil mapping activities. There is only a catalogue developed by Puslittanak (1996) that shows areas that had been surveyed for every province in Indonesia prior to 1996. This catalogue provides information on the area surveyed, mapping scale, available data, and publisher. The legacy data include soil observations (in the form of soil profile, soil minipit, or soil laboratory test data) and polygon data (soil maps with legends). However this catalogue does not show where we can find the report. Hence, we must manually check whether the survey reports are available in the library. In some cases, this becomes a detective work, finding the right person who happened to own it.

Table 1 lists potential reports in which soil maps and soil observations may be available. These reports are currently stored in documentation section and library of ICALRD. This data identification is only based on the limited data managed by ICALRD only. In fact, other institutes and agencies also conducted soil survey and mapping. Currently, there is no protocol yet to gather and store this dataset in single national database. In the future, a protocol for data sharing should be created.

2.2. Data selection

The previous step identified the legacy data and collected reports and maps. The next challenge is how to select useful data for DSM purposes. As Minasny and McBratney (2010) outlined, these data can be in the form of soil maps with legends, and soil observations. In this context, useful data include maps that have a good reference system and soil observations that are ready to use, meaning the data were thoroughly tabulated with clear location (geographical coordinates).

Based the information in Table 1, we examined the reports and grouped them as: (i) reports having soil observation data and soil map, (ii) reports having soil map only, (iii) reports having soil observations only, (iv) reports having no soil observation and no map. Also, we found map sheets with no report but we decided not to use them.

Further selection is based on polygon data and point data. For soil point data, we collected reports that have the complete soil profile description with soil physical and chemical analysis. In addition, the location of soil observations must be known either registered in geographic (LAT/LON) reference system using the WGS 1984 datum or in a UTM system.

Regarding polygon data, we found two groups of maps i.e. maps having clear reference system, and maps having no reference system. For the latter, georeferencing is a problem as we need to manually identify some land marks and perform GIS corrections with the help of topographic map. However this becomes tedious and not cost effective. The maps (polygon data) need to be processed to digital formats (scanning, and registering). Considering the limited equipment, time and operator, this cannot be achieved efficiently. Thus, for the moment polygon data are not being used for DSM. However map digitizing is still ongoing and these maps will provide rich data sources in the future.

2.3. Database development and population

After the data have been selected, the next challenge is how to create a useful database and then populate it. The requirements are the database is user friendly, no special training is required as budget is limited that and the data can easily be exported or used by any GIS and statistical software. The data is mainly for modelling soil–landscape relationship or for creating pedotransfer functions.

We first developed a simple database designed to accommodate the data needed for digital soil mapping. Due to time constraints and a limited budget, we exercise a stepwise data input approach. It means that we only input data as required. We designed two tables for storing different data; the first is to store the data on when and where soil observations were taken. For each soil observation, we constructed a table which contains general information as presented in Table 2.

The second table is to store the selected soil properties per horizon. As presented in Table 3, these are observation code, upper horizon