GlobalSoilMap.net - from planning, development and proof of concept to full-scale production mapping

R.A. MacMillan[A], A. E. Hartemink[A] and A. B. McBratney[B]

[A]ISRIC - World Soil Information, PO Box 59, 6700 WB Wageningen, The Netherlands, Email Bob.MacMillan@wur.nl, alfred.hartemink@wur.nl
[B]Faculty of Agriculture, Food and Natural Resources, The University of Sydney, Sydney, Australia, Email alex.mcbratney@sydney.edu.au

Abstract
The *GlobalSoilMap.net* project aims to produce predictions of nine key soil properties at continuous depth intervals at a spatial resolution of 90 m for the entire world. These maps of soil properties will be produced by a participants working under the coordination of regional node leaders with responsibility for organizing and delivering results for eight defined geographic regions of the world. This paper identifies and discusses the technical impediments to moving towards commencement of operational production mapping. These are: i) agreement on specifications for all products, ii) location, digital capture and harmonization of legacy soil data, iii) assembly of covariate databases, iv) documentation of prediction methods, v) specification of data model(s) to use to capture, store and disseminate maps and data, vi) selection of cyber-infrastructure to support map production and dissemination vii) end user surveys assessment and verification, and vii) identification of methods for assessing the uncertainty and accuracy of predictions. Actions undertaken to date to address these challenges are presented and progress is evaluated. There are no significant technical reasons for not moving towards planning and implementing operational production mapping.

Introduction
The *GlobalSoilMap.net* project has an ambitious goal of producing predictions of continuous horizontal and vertical variation for ten key soil properties at a spatial resolution of 90 m for the entire world (Sanchez *et al.* 2009). Production of these grid maps will be achieved through the combined efforts of diverse participants whose contributions are being coordinated by regional node leaders for seven continent-sized geographic regions of the world. The wide diversity of contributors, and of soils and landscapes for which predictions will be made, results in a need to identify and address the main challenges to moving forward towards operational production mapping of soil properties.

Eight challenges to moving forward towards operational production mapping
Eight technical challenges to moving towards rapid commencement of operational production mapping have been identified and actions to address these have been initiated. There may well be other technical or organizational issues that conspire to delay initiation of operational mapping, but it is felt that the challenges identified here represent key impediments that must be addressed in order to move forward towards operational production mapping.

Specification of output products
It is not possible to begin production mapping until the specifics of what is to be produced have been agreed to by all project participants. Specifications spell out explicitly which soil properties will be predicted, at what spatial resolution and depth intervals they will be predicted and in what spatial framework (coordinate system, projection, datum) they will be stored and distributed.

An initial set of project specifications was discussed and agreed upon at a *GobalSoilMap.net* consortium meeting held in Seoul, South Korea in October, 2009. These specifications call for prediction of a mean value for each grid cell of 90 x 90 m horizontal dimensions (3 arc-seconds) for ten critical soil properties (organic carbon, sand, silt, clay, coarse fragments, pH, depth to bedrock, effective soil depth, bulk density and available water holding capacity) at six specified depth increments (0-5, 5-15, 15-30, 30-60, 60-100, 100-200 cm). A spline function will be fitted to the soil property values for each depth increment to permit depiction of continuous variation in soil properties with depth (Malone *et al.* 2009). An estimate of the uncertainty associated with each prediction at each depth will accompany each property value. The predictions from diverse sources will be delivered for assembly and redistribution in geographic coordinates (lat/long) using WGS80 datum.

1 – 6 August 2010, Brisbane, Australia. Published on DVD.

Location, capture and harmonization of legacy soil profile and map data
The *GlobalSoilMap.net* project has a vision to capitalize on the major investments that have been made in collecting soils information and producing soils maps, locally and regionally, over the last 50 to 80 years. This archive of legacy soils data has been under-used and under-appreciated, because it has not been collated and harmonized into a uniform and easy to access or interpret whole. Using pre-existing sources of soil information for predicting a specific set of individual soil properties at a consistent resolution aims at coaxing a minimum level of consistency and uniformity from this diversity of legacy data. This project will simplify the complex presentation of soils information, often associated with conventional soil maps, into a model of continuous variation in the values of single soil properties both horizontally and with depth.

Before doing so, it must first locate and obtain the majority of existing data and then find ways to harmonize or standardize data of different age, quality, information content and density so that outputs of consistent content and appearance can be generated. A first objective of the *GlobalSoilMap.net* project is therefore to identify, locate and obtain, or rescue, as wide a selection of existing information about soils as is feasible and practical. This legacy data is viewed as having two main forms, namely legacy profile descriptions and accompanying analytical data, which mostly describe point locations, and legacy map data, that describe the horizontal variation of soil classes in space.

In the context of capturing legacy point data, procedures and data entry protocols have been developed to facilitate the capture and storage of a consistent subset of soil profile attributes from a wide diversity of sources of soil information. Metadata are recorded for each soil profile to enable identification of the source of each piece of soil information and association of it with a defined analytical method or data dictionary. This capability will be used to identify whether the values for specific soil properties reported by data from different sources are comparable and equivalent or whether there are systematic differences that need to be identified and resolved.

In the context of capturing legacy map data initial efforts are focussing on simply identifying, obtaining and scanning those existing legacy maps that have not yet been captured, even as digital images. In the longer term, the project will investigate how, or if, these maps can be used to contribute to procedures for predicting the spatial variation in the ten selected soil properties. Ultimately, it will be necessary to devise procedures for standardizing and harmonizing the content of existing legacy soil maps, regardless of whether they are already topologically structured or exist only as simple scans. One option is to use existing maps to inform the creation of uniform, country-wide or continent-wide, harmonised soil-landscape maps that are then used as one of the main covariates in prediction of individual soil properties on a continuous basis.

Assembly and pre-processing of global and regional databases of environmental covariates
Most methods for predicting the spatial variation of soil properties analyse relationships between evidence, in the form of point soil profile data or soil maps, and explanatory variables that represent environmental conditions believed or expected to influence the spatial distribution of soils and soil properties. Environmental covariates are selected to represent the ***scorpan*** factors as outlined by McBratney *et al.* (2003). Many environmental covariates represent the influence of the land surface on variation in soil properties. These are computed as derivatives of digital elevation models (DEMs) and are the key, but not the sole, predictors used in many soil prediction models. At the time of conception of the *GlobalSoilMap.net* project, the viable digital elevation model available for most of the world was the 3 arc-second (90 m) SRTM DEM. For most portions of the world, this 90 m SRTM DEM will represent the finest resolution DEM that is consistently available across entire continents of interest. So this SRTM DEM, and derivatives computed from it, will provide a significant contribution to the covariates available to predict individual soil properties. The cost benefit of trying to obtain and use derivatives computed from finer resolution 1 arc-second (30 m) SRTM DEM data is also being evaluated.

While many key covariates will be extracted from the best available DEMs, other sources of digital data will also be necessary to capture and represent the influence of other ***scorpan*** environmental variables, such as climate, organisms (vegetation), parent material, age and spatial context. A task has been defined to identify all major digital databases of environmental covariates that are currently produced and available in digital format at global to continental extents. We wish to avoid having different partners, working in different nodes, needing to identify and obtain these same data sets separately. The job is being be done once and the resulting global to continental scale databases can be accessed and used by all project participants.

In addition to simply collecting and collating existing global scale data sets of environmental covariates, these data sets may be used to devise and apply a global stratification into ecological regions. This definition of ecological domains or pedological provinces will be coordinated with other, existing efforts to define an agreed-upon global framework of ecological strata.

Identification, evaluation and documentation of suitable prediction methods
One of the main questions posed to (and by) project participants has been "what prediction methods will be used to produce these maps?" This is a key issue for which answers need to be provided before operational mapping can begin. A general conceptual design has been proposed which envisages using different methods in different areas depending upon the type and amount of soil evidence data available (point and map), the kind of landscapes and the type and strength of soil-landscape relationships, and the availability of suitable environmental covariate data sets (Minasny and McBratney 2010). Four "proof of concept" areas have been identified for areas in the USA and Canada, Australia, Europe and Africa. These areas range in size from about 50 km by 100 km for the USA/Canada pilot area up to 850 km by 250 km for the entire country of Malawi. Data for these sites have been obtained and are being used to produce digital maps of soil properties using different methods. These specific examples illustrate general concepts of mapping using each main combination of available data and landscape attributes. This exercise will help users in different areas identify and select the prediction method or methods that are best suited to the conditions with which they have to deal. The resulting documentation and accompanying data sets can be used as tutorials and training manuals to help participants learn how to apply selected methods in their own areas

Specification of conceptual and physical data model(s) for GlobalSoilMap.net data
The *GlobalSoilMap.net* project offers a unique opportunity to redefine the structure and content of global soil databases using the most recent database concepts and modelling tools. This opportunity has been seized upon to develop a proposal to define a new global standard for the storage and exchange of soils information. This global standard will encompass the full range of types of soil information and will not just be specific to the grid maps of soil properties being prepared by the *GlobalSoilMap.net* project. It will set standards for point profile data, area soil samples, conventional polygonal soil maps and continuous raster soil property maps. A task group has been proposed to work under the auspices of the International Union of Soil Sciences (IUSS) to debate and design a new global SoilML based upon UML and XML standards.

Selection and implementation of appropriate cyber-infrastructure support
Storage and on-line delivery of the large global data sets of soil properties that the project will produce represents another challenge to be addressed. The project is building a cyber-infrastructure that can support collaborative efforts to acquire and process data sets of legacy soil data (points and maps) and environmental covariates of global to continental extent. This cyber-infrastructure will provide immediate support for on-line collaborative sharing of data, inter-active sharing of prediction methods and tools, and inter-active discussion forums and communication amongst project participants. In the longer term, a task group has been set up to draw on the experience of *GlobalSoilMap.net* partner agencies (e.g. USDA-NRCS, JRC, CIESIN, CSIRO, ISRIC) in storing and delivering large volumes of soils data on-line. This task group will investigate and recommend options for facilitating the storage, discovery, visualization, analysis and downloading of the data produced by the project.

Assessment and verification of ability to meet needs of major end users
This task seeks to validate the underlying assumption that "if we build it they will come". Essentially, the plan is to conduct a complete and systematic identification of all of the potential users of the soil property maps that will be prepared by the project. We then propose to identify the specific needs of each major user for soils data and assess the degree to which these needs can and will be met by the products that the project proposes to deliver. A task group has been formed with a mandate to systematically identify all major potential end users, to identify the specific needs of these end users for soils data and to verify the extent to which soil information produced by the project will meet those specific needs. The results of these investigations should be available for reporting by the time of the meeting at which this paper is presented.

Identification and implementation of methods for assessing uncertainty and accuracy of predictions
One of the exciting and important new aspects of the project is its commitment to providing an estimate of the uncertainty attached to each soil property prediction at each depth at each grid cell location. Awareness of the high level of uncertainty for particular areas may stimulate the collection of new data. A method has

been proposed for estimating the uncertainty of predictions of soil properties by depth (Malone *et al.* 2009) that makes use of analysis of all geo-referenced soil property values reported for a given area. On a second front, a task group has been set up to investigate and identify viable options for collecting independent field samples to support computation of estimates of predictive accuracy for the soil property maps at different levels of aggregation.

Planning of operational production mapping

Production mapping will not start unless the individual nodes agree to set specific targets for both initiating and completing production mapping for specific areas of large extent. Node leaders are being encouraged to consider how they can obtain the resources (human and financial) to support full scale operational production mapping and also need to set targets for extents to be mapped within specific time frames.

Conclusions

The *GlobalSoilMap.net* consortium aims to move forward from initial planning and proof of concept activities towards full scale operational production mapping. Several challenges have been identified and actions are being undertaken by specific task groups to address them. Sufficient progress has been made to encourage project participants to begin thinking about how they can move forward from planning and proof of concept to operational mapping. Initial examples from some participating nodes have demonstrated that it is indeed possible to produce predictions of soil property values by depth at 90 m for entire countries or states. Node leaders are being encouraged to set targets and timelines for operational mapping of their nodes and are actively seeking the funding and resources that will be required to succeed in these challenges..

References

Malone BP, McBratney AB, Minasny B, Laslett GM (2009) Mapping continuous depth functions of soil carbon storage and available water capacity. *Geoderma* **154**, 138-152

McBratney AB, Mendoca Santos ML, Minasny B (2003) On digital soil mapping. *Geoderma* **117**, 3-52.

Minasny B, McBratney AB (2010) Methodologies for global soil mapping: Digital Soil Mapping: Bridging Research, Production, and Environmental Application.In ‘Proceedings of the 3rd Global Workshop on Digital Soil Mapping (DSM 2008 USA)’ Progress in Soil Science, Vol. 2 (Eds JL Boettinger, D Howell, AC Moore, A Hartemink, S Kienast-Brown). (Springer)

Sanchez PA, Ahamed S, Carré F, Hartemink AE, Hempel J, Huising J, Lagacherie P, McBratney AB, McKenzie NJ, Mendonça-Santos ML, Minasny B, Montanarella L, Okoth P, Palm CA, Sachs JD, Shepherd KD, Vågen TG, Vanlauwe B, Walsh MG, Winowiecki LA, Zhang GL (2009) Digital soil map of the world. *Science* **325**, 680-681.