The US Soil Information System: A Model for Data Collection, Curation and Distribution

This paper focuses on activities related to the collection, organization, curation and distribution of soils information in the United States along with a brief discussion of new initiatives that will allow the Natural Resources Conservation Service – Soil Science Division to provide useful data and information to our users in the future. The paper covers four major topic areas: 1) The types of soils data available in the United States; 2) The array of information products that address user needs for soils information; 3) The various distribution mechanisms for the data and information products; and 4) New initiatives to provide soils and ecological systems information.

Soils data currently available from USDA-NRCS

The primary data source for soils information is the Soil Survey Geographic (SSURGO) database. This database, maintained by the NRCS, contains hundreds of estimated properties for soil landscapes and components that cover over 90% of the United States at a 1:24000 spatial scale. The STATSGO database, provides a smaller set of estimated properties for the entire country at a 1:250,000 scale. The National Cooperative Soil Survey (NCSS) Soil Characterization database contains measured data on over 1000 soil properties, as well as many other calculated soil properties, obtained from over 63,000 sites throughout the US and the world. All of these datasets are developed based on consistent, well-documented standards and specifications.

Existing information products

The official US soils data (SSURGO) includes a small number of key calculated pieces of soils information that are commonly used to inform land management decisions including available water capacity BS and soil erodibility. Whole soil values for important properties such as total organic carbon can also be calculated. In addition to calculated values, the much larger group of soils information products are referred to as interpretations. These interpretations can be broadly divided into three categories: classes, suitabilities and limitations. What differentiates interpretations from calculations is that interpretations rely on sets of rules to combine soil properties and estimate their overall effect on soil behavior. The most widely used class interpretation, although it is often not thought of as an interpretation, is soil texture which is determined based on a set of rules about the percentages of sand, silt and clay in a given soil (the textural triangle). Other common class interpretations are prime farmland, land capability class, hydric soils and hydrologic soil group. The other two categories of interpretations are in a sense mirror images of each other. Limitation-based interpretations focus on the interactions between key soil properties that limit the ability of a soil to be used or managed for a certain purpose whereas suitabilities focus on the key soil properties that make a soil suitable for a given use. Because one of the major functions of interpretations it to provide land managers with information about what must be done if a given management action is to be successful, limitation-based interpretations are much more common. In total, the published SSURGO database contains just over 100 nationally applicable interpretations and a similar number of locally-specific interpretations.

Distribution mechanisms

In the past, soil survey data and information was distributed through county-based hard copy soil survey publications. However, with the exponential expansion of digital technology over the past 30 years, the distribution of soils information has changed significantly. The primary gateway to US soils data for most users of soils information is the Web Soil Survey which contains the official USDA soils data (refreshed annually and as needed). The Web Soil Survey (WSS) site is the most used USDA informational outreach site with over 152,000 unique users every month. On average these users create over 493,000 customized soil reports developed every month. The WSS allows users to identify and obtain data and information for "Area of Interest" of up to 40,000 ha including tabular reports, thematic maps and professional customized reports. The web soil survey also provides the capability for users to download both spatial and tabular data for one or more of over 3200 soil survey areas.

The SoilWeb app, a cooperative project between the USDA-NRCS and the University of California at Davis provides mobile phone users a simple and convenient way to access basic soils information. This application provides web access based on user defined location, or the user can enable their device GPS system to set their location. Due to programming constraints, this system uses a yearly download of the official soils data.

The USDA-NRCS also offers external web-service access to all of our official soils data through the Soil Data Access Facility. This web site and related resources provide web map services and web feature services. Data is obtained from the web site using custom queries written in standard SQL and can be easily integrated with other web applications as well as with other external applications such as simulation models.

The National Cooperative Soil Survey laboratory database contains laboratory data from over 60,000 pedons worldwide, although the vast majority are from the United States. All of this data is now available on-line through the NCSS soil laboratory data mart via both spatial (active map query using Google Fusion™) and tabular query mechanisms.

New initiatives

The final section of the paper will highlight several new initiatives to expand the existing suite of NRCS data and information to address changing user and societal needs. The first initiative is the collection and dissemination of data and information related to Ecological Site Descriptions or ESDs. An ecological site description is grounded in soils information but adds further detail and depth to that information by describing both the existing plant community, the different plant communities that may occur under different disturbance regimes, and the actions (natural and/or man-made) that cause the communities to transition from one state to another. The long-term goal of these efforts it to provide additional, soil-specific information to land managers that will allow them to better manage their land to meet their objectives without causing environmental degradation.

The second initiative is dynamic soil properties or DSPs. We in NRCS define DSPs as the soil properties that change over the course of days, months or years. With the increased public awareness of soil health and its wider use as a more holistic metric of soil condition, there is a need for better data to geospatially assess the present condition of key soil properties such as organic matter, compaction, and others over something larger than the individual site and to develop ways to measure and document changes over time. Collecting this sort of information over time using agreed-upon and replicable methods has the potential to greatly expand our understanding of soils and to provide more and better information to inform land management than we can at present. But, the challenges of collecting, storing, managing, interpreting and disseminating this data and associated information remain daunting.

The third initiative focuses on the increased use of raster and property data and related information products. In the US, current and legacy soil data and information is correlated, stored and disseminated primarily in polygon form. The USDA-NRCS has responded to the increased demand for raster-based products through the short term creation of the gridded SSURGO (gSSURGO) project, a raster representation of current polygon data. Efforts are currently underway to use environmental correlates to disaggregate existing map units to provide both a finer spatial scale and a better representation of uncertainty associated with soil components. Other groups in the agency are exploring a new paradigm of grid-based mapping to supplement existing soil-landscape polygon mapping activities. These efforts complement US participation in international efforts such as the GlobalSoilMap.net project sponsored by the International Union of Soil Scientists. Globalsoilmap.net has chosen to focus on a smaller number of key soil properties measured at specific depths in order to develop a truly global raster dataset that can be used both to inform local management efforts as well as to provide inputs to national, regional and global scale modeling efforts.

Conclusion

We recognize that these activities result from a sustained level of public investment that may be neither practical nor realistic for other countries to duplicate. However, we believe that the lessons learned through this evolving process, including the examples of the data collection, curation and management infrastructure the US has developed and the continually evolving suite of methods to distribute this information to an increasingly diverse customer base, may serve as a model for similar efforts throughout the world.