Foundational Soil Survey Data-A Case for Linking Basic Soil Data and Interpretations to the Official Series Description

Samuel J. Indorante

The purpose of this article is to present some ideas and concepts that will enhance the consistency, manageability, utility, and quality of the soils interpretations within NASIS for the current updating and digitization phase of the National Cooperative Soil Survey. Following is a case for creating and linking basic soil data and interpretations to the Official Series Description within the National Soil Information System (NASIS).

Soil Survey Interpretations

The Old Way of Doing Business (early 1970s-1998)

Early in our careers we were taught individual soil series concepts and that with each soil series concept (Official Series Description, or OSD) there were corresponding basic data and interpretations (Soil Interpretation Record, or SIR). Together, the OSD and SIR defined the central concept of the series, listed ranges of key properties of the series, and provided basic interpretations for the series as a starting point for interpretations for soil map units or phases. We were also taught that no OSD was complete without an SIR because the OSD alone did not contain enough information to document all ranges of the series. Current policy does not require SIR information, but some Major Land Resource Area (MLRA) offices have attempted to capture some of the data lost with the dropping of the SIR. For example, see ***http://ww2. ftw.nrcs.usda.gov/osd/dat/C/CECIL.html for*** an example of an OSD with SIR information, or see any OSD under the responsibility of MO14 Office in Raleigh, NC (John Kelley, 2007, personal communication).

Presently, more than 19,000 series have been recognized in the United States (USDA-NRCS 1999a, 1999b). For most of these 19,000 OSDs there is no attached SIR or foundational data. Keep this important number 19,000 in mind – it will be discussed again later.

Early in our careers we were also taught that soils are landscapes as well as profiles (USDA-SCS, 1951, p. 5-8; USDA-SCS, 1993, p. 9-11). It is this understanding of soil-landscape relationships that allows us to identify and delineate soil-landscape units and soil map units. We then use Soil Taxonomy (USDA-NRCS, 1999b) to identify and label soils (series or phases of series) that occur within the soil—landscape units and soil map units. The final step is to provide interpretations for the soil—landscape units and soil map units by selecting and editing the basic SIR data and interpretations that correspond to the local conditions for the soil(s) or component(s) for the map unit. I have simplified a complex sequence of events, but the details can be easily found in Soil Taxonomy (USDA-NRCS, 1999b) and the Soil Survey Manual (USDA-SCS, 1993, p. 9-11).

S.J. Indorante, Soil Scientist, USDA-NRCS, 148 E. Pleasant Hill Road, Suite 105, Carbondale, IL 62903 (email: **sam.indorante@il.usda.gov**).
Published in Soil Surv. Horiz. 48:43-45 (2007).

The Soil Survey Handbook (USDA-NRCS, 2005) outlines the development of soil interpretations ratings and also provides a “ratings guide" for the various interpretations. It should be noted that the interpretations are developed from a mixture of field data, laboratory physical and chemical data, and estimates of data by interpolation and extrapolation. To my knowledge there was no distinction between any of the data types used to develop the SIR interpretations. In a recent article Livingston (2006) presented the challenges and offered some solutions to data quality and quantity within NASIS.

NASIS: The New Way of Doing Business (~1998—present)

With the advent of NASIS in 1998 the SIR was dropped along with the link between the OSD and SIR. Interpretations were now made within NASIS. Calhoun (2001) summarized the three major concepts for interpreting soils within NASIS:

- We will interpret for the entire range of any component within the data map unit (DMU).
- We interpret for what is actually there.
- We are not confined to the series concept, and there is no need to be concerned about the limits and boundaries of a soil series.

The Impact of the New Way of Doing Business

The number of soil series currently recognized in the United States is approximately 19,000. If there are 19,000 OSDs, there are/were at least 19,000 corresponding SIRs (one or more, including phases, for each soil series). Each SIR had between 30 and 50 data records (e.g., data cells) or elements to populate (assuming four soil layers). Some records had more than one number or entry. Suppose the average number of data records to populate is 40. Multiplying 19,000 by 40 results in 760,000 data records to populate. Not only did we have to populate the data records, but they had to be reviewed, edited, updated, and revised periodically. Each state was responsible for statewide OSDs and SIRs; the workload seemed manageable because it was divided among the 50 states. Each state was roughly responsible for 15,200 data records. Keep in mind that these 19,000 plus SIRs were developed over a period of more than 20 years.

With the advent of NASIS we have to redo the math. By my estimate a DMU with four layers in NASIS can have somewhere between 500 and 1000 data records to populate. This includes legend, map unit, data map unit, component, and horizon information. Multiplying 19,000 OSDs by an average of 750 data records results in about 14,250,000 data records to populate!

But, wait a minute, we are not finished yet. The 19,000 is the minimum number of data map units to populate. The number will likely be much greater, because there will be one or more DMUs for each series. As an exercise multiply 14,250,000 by an average of 1.3 DMUs for each series. This big number gets much bigger very fast (18,525,000 data records). Divide 18,525,000 data records by 50 states. Each state is now responsible for 370,500 data records on average. The data workload for a state is 25 times larger! In reality the states with large legends would get hit much harder. For example, Illinois has approximately 5200 map units on its state legend. Multiply 5200 by 750 and the number of data records to populate in Illinois with its present legend would be about 3,900,000. Even if the Illinois state legend was correlated to one-half its present size, Illinois would still have 1,950,000 data records to populate.

It is difficult to foresee all of the implications of linking data map units and copying and changing a few data records in a data map unit to create another data map unit. It is obvious that as the number of DMUs increases, it will become more and more difficult to control the quality of the data (e.g., ranges in soil properties) and interpretations attached to the OSD.

Another consideration is that there are already many DMUs that were created at one time or another for a specific purpose, but are no longer used. The number of DMUs in the U.S. is now around 500,000, and we have only been using NASIS for about 8 yr. One thing I am certain of – the number of DMUs and the complexity of maintaining them will continue to increase.

These estimates emphasize the need for a structured, focused, and coordinated population of the NASIS database. NASIS has many capabilities (USDA-NRCS, 1999a), but to meet the current customer demands for soil survey information, we need to focus on building and supplying sound foundational soils information to our soil survey users.

Now that the very rough math has been completed, what questions do we need to ask next?

- Who will populate and manage the more than 18 million data records, and what will it take in terms of time and resources?
- How can we check the quality of such a large number of data records?
- Where will the numbers for populating the data fields come from (Livingston, 2006)?
- Is there a possibility that soil series concepts (The Official Series Descriptions and Range in Characteristics) will become weakened and obscured due to data quality and data quantity concerns?
- What are the implications of spending the majority of a soil scientist's time populating and managing the database instead of studying soils and landscapes in the field?
- NASIS is a “Transactional" database and not a data storage and summary tool. This means that it will likely change with time. Is it possible to maintain the quality and consistency of our soil data if we have multiple users editing and changing a shared data map unit?

Livingston (2006) mentioned some of these issues. I will not attempt to answer the above questions, but here are some suggestions for how to make the database more manageable, serviceable, and useable.

Something to Think About

The traditional SIR provided basic soil interpretations for taxonomic units and phases of taxonomic units, and NASIS DMUs provide interpretations for soil mapping units. By definition, SIRS are relatively static compared with DMUs. As pointed out by Calhoun (2001), this flexibility in NASIS DMUs brings up the possibility of totally losing the class-limiting information located in the SIR.[1] So the question should be: How do we maintain the Class limiting information located in the SIR and still provide flexibility to provide interpretations for soil mapping units?

Part of the answer is that we need to go back to using some form of the SIR concept (such as practiced in MO 14 as mentioned above). If each OSD had a corresponding NASIS-SIR, then everyone who builds a data map unit will begin with the same building blocks. The NASIS-SIR could be numbered or indexed in such a way that when someone uses (copies or links to) a DMU they know whether or not they are using the NASIS-SIR or a copied and edited version of the NASIS-SIR. The text notes in NASIS can supply a user with additional information on editing, but an indexing system is needed for consistency and ease of use. The advantage is that we can concentrate on strengthening our OSD concepts (limits and boundaries) and our interpretations, while still allowing flexibility in customizing data for specific DMUs.

To one degree or another, SIR data have been transferred to NASIS, but little or no effort has been put into cleaning up, editing, updating, and connecting the old SIR to a new NASIS-SIR. We can begin with the old SIR concept and utilize the flexibility of NASIS to develop a NASIS-SIR that includes foundational data. The NASIS-SIR will likely include more data than the traditional SIR, but all data elements would not have to be populated. Developing NASIS-SIRs and gaining greater control over the quantity and quality of data (Livingston, 2006) would be well worth the effort in the long run.

A Proposal for a NASIS-SIR or an OSD Data Object

At the State Soil Scientist/Soil Data Quality Specialist meeting in Kansas during April of 2001, Rick Fielder (Soil Data Quality Specialist, Little Rock MLRA Office, Retired) gave a presentation entitled "OSD Data Object-NASIS Data Quality/Population Tool." He proposed the development and use of an OSD Data Object.

Fielder's proposal would involve creating an OSD Data Object that includes all soil properties considered to be standards for the series. This would include series properties (ranges) that were formerly included in the SIR and some additional properties within NASIS. Minimum data population standards for an OSD Data Object can be used to ensure that critical soil and landscape properties are populated. It is likely that the minimum data set for the OSD Data Object involves populating between 250 and 500 data records. The immediate benefits of the OSD Data Object are:

- We are building on the SIR information that has been developed over a period of 25 to 30 years.
- We are focusing our energy on first populating critical soil and landscape properties that will be included in all OSD Data Objects.
- This allows soil data users to use the data as is or to copy the information from the OSD Data Object and edit or add to the OSD Data Object to customize the information for specific uses. Note, once the OSD Data Object is copied and edited, it is no longer an OSD Data Object, but a derivation of it and thus needs to be renamed or indexed as such.

---

[1] As NASIS is set up now there is also a distinct possibility of having different interpretations for the same phase of a soil series (taxonomic unit). While it is acceptable to have different phases of a soil series with the same interpretation, it is considered unacceptable to have the same phase of a soil series with differing interpretations.

A key point in Fielder's proposal is that an OSD Data Object would provide the capability to compare a DMU component with the OSD concept to confirm if it is within the defined ranges. This kind of comparison can be done quickly and accurately with a digital database and will help us develop and strengthen series concepts and their associated interpretations.

Key Functions of the OSD Data Object

Fielder's (2001) proposal could provide a number of useful functions:

- Identify soil properties that are considered a standard for the series.
- Capability to compare component data to the OSD record to confirm component is within defined ranges. This could be accomplished with a data comparison capability within NASIS, or by hard copy with component record and OSD record reports.
- Provide OSD record report to accompany an OSD for review and comment.
- Facilitate quality assurance with a nationally recognized standard.
- Provide base DMU component for MLRA project updates.

Official Series Description data and interpretations linked to an OSD would help meet the demand for foundational soil survey information that is well defined, well documented, up to date, and easily accessible. One of the most popular soil survey products is the Official Series Description Database. It is easily accessible and easy to use. But many users would like basic or foundational information (soil data and interpretations) linked to the OSD. It is easy to envision a link from the OSD directly to the soil data and interpretations (Fig. 1). It must be made clear that the data and interpretation apply to the OSD and are just a starting point for developing ranges of data and interpretations for map units in a soil survey. The soil data user would be directed to the NRCS Soil Web Page for further information on NASIS, SSURGO, or Web Soil Survey.

Summary

We are well into the digital age of soil survey, and this means we have two major sources of data to collect and digitize. We have both spatial data and associated attribute data. Geographic Information Systems (GIS), Global Positioning Systems (GPS), and Personal Digital Assistants (PDA) now allow us to build and quantify the spatial portion of our soils database in both two and three dimensions. Collecting data systematically across the soil-landscape over time will strengthen our understanding of soil—landscape relationships and our series (OSD) concepts. It follows that as we strengthen our series (OSD) concepts we can focus on populating and building OSD Data Objects that will provide the foundation for building associated DMUs for MLRA project updates. By building DMUs from this foundation we can lessen the workload to a manageable level and help assure the quality of our series concepts. An OSD data object would ensure that everyone who builds a data map unit will be starting with the same building blocks. The OSD Data Object could be numbered or indexed in such a way that when someone uses (copies or links to) an OSD Data Object they know whether or not they are using the OSD Data Object or a copied and edited version of the OSD Data Object. Data Map Units can be systematically built from the OSD Data Objects, and modelers and other users of soil survey information can copy the indexed OSD Data Object and customize them for their specific needs.

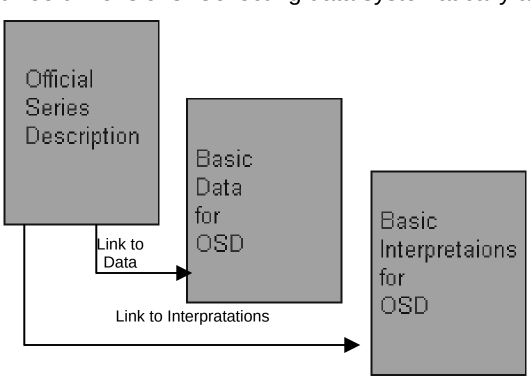

Fig. 1. Model for accessing foundational soil series information.

Yes, the workload is still substantial, but at least we can systematically build the database so it will be easier and more efficient to populate and update. It would also be easier to maintain the quality, consistency, and integrity of the data. The next logical step would be to link the OSD Data Objects to the OSDs on the web (Fig. 1). The NCSS has been praised for making OSDs easily available through the internet. To link the OSD Data Object to the corresponding OSD would make our soils data and interpretations available to anyone with a web browser. This is a product that would satisfy the vast majority of users of soil survey information.

The last step in this process would be to link the spatial data (e.g., SSURGO) with the attribute data (NASIS). This step would produce a synergistic effect. We would then be taking the best that each database has to offer to produce a digital soil survey. We could utilize the classifying power of a spatial database and the interpretive power of NASIS.

Let each database (SSURGO and NASIS) do what it does best, and then put them together to provide our customers with a digital product that is easy to use and understand. Also, let field soil scientists do what they do best—study soil-landscapes, develop and strengthen series concepts, and develop and strengthen the associated series interpretation records (OSD Data Objects). Having foundational SIRs in NASIS (e.g., OSD Data Objects) will increase the manageability, utility, and quality of our national soils database.

Acknowledgments

Thanks to Rick Livingston and John Kelley for reviewing the article and for their constructive comments.

References

Calhoun, T.E. 2001. Are classification and interpretation mutually exclusive processes? Natl. Coop. Soil Surv. Newsl. May 2001. Issue 15.

Fielder, R. 2001. OSD Data Object, NASIS Data Quality/Population Tool. Presentation at the April 2001 State Soil Scientist/Soil Data Quality Specialist Meeting, Kansas.

Livingston, R.L. 2006. "Truthiness" and the National Soil Information System: How are they related? Soil Survey Horiz. 47:43-45.

USDA-NRCS. 2005. National soil survey handbook, title 430-VI [Online]. Available at http://soils.usda.gov/technical/handbook/ (verified 16 Mar. 2007).

USDA-NRCS. 1999a. National Soil Survey Information System v 4.1. Soil Surv. Staff.

USDA-NRCS. 1999b. Soil taxonomy: A basic system of soil classification for making and interpreting soil surveys. Soil Surv. Staff. USDA Handb. 436. 2nd ed

USDA-SCS. 1951. Soil survey manual. Soil Surv. Div. Staff. U.S. Dep. Agric. Handb. 18.

USDA-SCS. 1993. Soil survey manual. Soil Surv. Div. Staff. U.S. Dep. Agric. Handb. 18.2nd ed.