

**COMMITTEE 2**  
**DATA ACQUISITION FOR PROBLEM SOLVING**  
North Central Soil Survey Work Planning Conference  
Madison, Wisconsin  
June 24-27, 2002

The charges are:

**Charge 1 – What additional data is needed in NASIS**

Working and discussing with users of soil information for their data requirements are methods to determine what additional soil data elements are needed in the soil survey database. Sending out a questionnaire to users of soil information for their data requirements is another method to determine what additional soil data elements they need. This last method gets user's wish list and they may request items they do not know what the data element is or probably do not need or will ever use.

What data do NASIS users need that is not currently in NASIS?

A geographic linkage for each polygon of a NASIS map unit is needed. This is presently absent in NASIS 5.0 for the soil scientist to use. There is a real need to know what MLRA region or subdivision of a given MLRA that a polygon for a detailed soil map unit falls within. This is unknown today within the NASIS structure.

As for additional attributes, there is not really a need for further expansion of the already complex NASIS attribute structure. The current condition is that there is a place in a table for such information, but the tables are often empty.

- Data model that incorporates both spatial and attribute data.

Where something occurs on a landscape and in what part of the world is often as important as what occurs. This concept is important in data development and maintenance and in maintaining consistency and uniformity.

- Complete and accurate population

Provide accurate and complete population of basic properties in all data mapunits. There is a need for more measured data to substantiate the entries in NASIS. More time needs to be spent on collecting data to support what is in NASIS rather than adding more fields to populate without having the measured data to guide in populating these added data elements.

- Data quality

As data developers, we have an obligation to provide complete and accurate data that are free from bias and nuance. When the data are separated from the programs that operate on them, the data must stand on their own free from ambiguity. A null field can only imply a null meaning. Values of interpretation ratings must be aligned in the same direction. Representative flags must be set to pass along knowledge gained by the field mapper. No tool that processes the data can extract answers that are any better than the data. The data can only be presented in different views.

- Feature attributing

Provide capability to record variability in composition and properties for each polygon rather than at the map unit level. Often map units are mapped in large geographic areas. In geo-statistics, it is often

assumed that the farther the distance from the source, the greater the variability. In dealing with the soil pedon, we know that the distance can be rather short.

- Location

Populate landscape location of all components for each polygon in map units, especially for those that are contrasting and limiting. An example is a polygon for a map unit with a rock outcrop component. Need to be able to assign appropriate values to grid data used in models. Many models use properties such as permeability and runoff. If the areas of rock outcrop can not be located in the polygon, then the data generated by the model are essentially useless.

- Data elements that are proposed for addition to NASIS. They are Soil Moisture Status (Attachment A explains the proposal and is being considered – see the NASIS 5.1 release notes.), aggregate stability, and soils susceptibility to crusting. There is a need to improve our yield data and make it compatible with recent measured long-term yield. Our yields now in the soil survey database seem to be on the very conservative side.

One committee member's opinion is that we really do not need to add any new fields into NASIS. There is very little hard data to back up our NASIS entries. More time should be spent collecting data to support what we have rather than adding more fields and having no data to enter. It is a dangerous idea to try and predict how modelers will use our data and then design our database to match those predictions.

#### What data is needed for simulation models and soil quality assessments?

One member of the committee stated, "I believe the previous Committee 2 (2000 NCRSSC in Grand Rapids, MI) did a good job of identifying data needs. However, if the only reasons (we are collecting the additional use-dependent data) are for simulation models and soil quality assessments, then I would question if it is worth the time and effort. Creators of simulation models can generate their estimates with or without the soil data and these models are useful primarily when measured data is not available. If we have the required measured data we often will not need the models. The soil quality assessment effort is not totally accepted, rather poorly defined, lacks a scientific basis and is supported primarily by administrative efforts. To require another program (such as the soil survey program) to spend a lot of time and money to collect data for the soil quality assessment effort is questionable. That program should collect its own use or site specific data."

There seems to be some danger in trying to predict how modelers will use data in NASIS.

An AGNPS user requests Base Saturation as an item, as well as unstable aggregate ratio. One responder has a point, how important is and what is the sensitivity of base saturation in AGNPS? This same responder states "Unstable aggregate ratio is used in the RUSLE portion of AGNPS and is considered by a sister agency to be of low priority to incorporate at this time.

#### What should be the priorities for populating data elements?

A new strategy toward using NASIS should be considered. Identify those basic soil chemical and physical properties that are common to most of our USDA programs and other users, and then logically populate those properties within a consistent landscape region (sub division of MLRA or MLRA). If this is too difficult to attain, one must consider the usefulness of the existing system, for without reliable physical and chemical information, how can we hope to make reasonable interpretations for land use planning, etc. Such basic information includes: pedon based measurements and descriptions. Also included are estimated information - horizon depths, USDA texture classes with modifiers, rock fragment

content, depth of root zone for different plants (e.g. commodity crops), AWC within each horizon, dominant land use condition and, therefore, degree of anthropogenic influence (use-dependent data) for map units when mapped (use new National Land Cover 30 meter data or NRI), soil organic matter, land form, parent material, CEC, base saturation, estimate of mineralogy, most up-to-date Soil Classification (meets 1999 standard), and appropriate MLRA or subdivision of MLRA (Land Resource Unit or Common Resource Area) and general soil map unit (STATSGO or Digital General Soil Map of the US).

Two responders made this comment: “Much discussion has taken place over the past few years about data population in NASIS, but no one has sat down and calculated what the workload would be to meet the data population guidelines from the above reports and from other requests for data population. The two responders have begun to periodically check NASIS to see how quickly the number of DMU's (Data Mapunit) is increasing. The date is the date we checked NASIS and the number is the highest number DMU in NASIS on that day.

Here are some numbers to think about:

2-8-2002 409,100 DMU's

2-15-2002 409,662 DMU's an increase of 562 in 7 days

2-26-2002 410,573 DMU's an increase of 911 in 11 days

Could these numbers be misleading? DMU's can be added for various reasons. Some folks duplicate DMU's for archival purposes, others might be test or what-if things. Other folks do not know they are trying to maintain the same DMU seven times, i.e., they need training on the proper use of DMU's. None the less, it is difficult to tell which ones of this increase are “real” ones that need populating and/or maintenance.

Multiply the DMU number (410,573) by 750 (average number of data elements to be populated per DMU) and we already have a tremendous workload and the number of DMU's is increasing everyday. How does 307,929,750 data elements sound! This assumes that every data element needs populating for every soil. This may not necessarily be the case, but none the less the number is alarmingly large.

Some responders state - people will say that not all of those DMU's are or will be populated. Our answer to that is that someone still has to sift through the data to tell the good from the bad. The point is that workload should also be discussed when determining what should be populated in NASIS.” Another responder said, “Workload is taken into consideration before new data elements are added, or suggestions are made to populate something.”

- We need to be very conscious of what is set as RV's, as RV's are what many modeler's use. Example: Many soils that typically don't have gravel contain 0 to 15 percent fragments less than 2mm in the data to "cover the range" of the texture. Most of these soils should have a RV of less than one-percent coarse fragments, not the midpoint (7 percent). This situation infers that these data elements in the soil survey database need editing to reflect the coarse fragments for the soils in the soil survey.
- Some of our data elements, as currently entered, are subject to gross misinterpretation. Example: Component Slope Length. Many users view this as additive, i.e., a map unit has 5 components (major and minor) with slope length RV of 30 meters. Therefore, the slope length of the map unit is 5 times 30=150 meters. Slope length somewhat depends on the landscape, e.g., to say a map unit that contains potholes has a slope length of 100 feet is wrong for the soils in a pothole landscape, but is correct for the soils on the uniform slope. For map units with uniform slopes, slope lengths are more closely related to the map unit. Data elements need to be defined in enough detail, and these definitions must accompany the tables, so the user fully understands how to use the data element correctly.

One responder states - NCSS has collected soil data by soil survey area (mainly county or parish) for over 100 years. Much of this data is populated in NASIS, but entered by default from State Soil Survey Database (SSSD). Let's first review and certify this data. Let's not rush to populate NASIS by MLRA without doing some investigation.

## **Charge 2 – Who will collect, enter, and maintain the data?**

### Who will collect and enter the data?

Soil scientists in the field, who have made soil maps to National Cooperative Soil Survey (NCSS) standards and understand soil/landscape relationships, will collect the soil samples for the laboratory and field data as part of a systematic gathering of soil-landscape data. Data that has already been collected by NCSS, should be reviewed and then be included in the data set if the data meets the requirements of the NCSS standards for quality. Someone will need to define what quality data is? It is time and cost prohibitive to collect measurable data on all soil series. It is possible to come up with benchmark soils to collect complete soils data on, and then interpolate (not extrapolate) the data for soils in between. It would be easy to flag benchmark soil series data (measured data) and differentiate it from interpolated data. Each of the benchmark series should have complete characterization. Interpolated data can be flagged by saying something like - "The data for the Y series was developed (interpolated) by using data from the X and the Z series."

Data can be collected by public and private sectors. The Natural Resources Conservation Service (NRCS) is the lead federal agency for the federal soil survey program; therefore, at this time the “keeper” of the soils databases for the federal government. The standards to which the data needs to be collected, submitted, and stored is the responsibility of the NCSS. Data should be collected by or under the quality assurance of NCSS.

It seems logical that the responsibility should remain with the NCSS since it has existing standards that can be used for the collection of information. What seems to be missing is an easy way to bring existing and new information into soil survey database. There is a real need for good translation software to take full advantage of existing soil knowledge gathered in the University and private sector environment. Perhaps a 3<sup>rd</sup> party could develop such software?

Data should be made available at the discretion of the owner of the data. Our traditional soil survey with narrative, maps and interpretative tables should be made available via the Web and the new Soil Data Warehouse or Data Marts. The digital soil geographic data sets and laboratory measurements should also be made available as date stamped datasets on CD-ROM and via Web on a yearly or bi-yearly basis.

One responder stated that it is the responsibility of everyone that uses the data to have the ability to add to it. There will be standards set nationally for every type data that is collected. There is always the need for accurate data; therefore, it would be foolish not to accept data from other sources just because someone else collected it. NRCS should enter and maintain all the data. This should be the job of the designated soil scientist with guidance from the lead soil scientist responsible for data quality. The designated soil scientist should evaluate the data that are collected from the various sources and use it to create guidelines and populate the soil survey database.

One member of the committee stated, “Let the programs that need the soil data, actually collect and maintain their own data base. The private sector should not be required to contribute data unless it is being used to support a position before a government agency (for a regulatory purpose). At that time the soil data would become public information and could be stored by Government scientists for other uses.”

One responder states - Any soil scientist can collect data. The question becomes what data is legitimate. A distinction needs to be made between point data and aggregated data. Aggregated data entry needs a "gatekeeper" to ensure quality. Until some method is developed to incorporate others outside of NCSS into the correlation process, these data are suspect.

#### Data Collection

- Gather data in the field using devices such as a GPS unit, Palm Pilot, and a windows based computer with Pedon software.
- Dump the data into a GIS relational database for local use.
- Have NASIS set up so you can download data from a data recorder (Palm Pilot) and soil descriptions in Pedon directly into National Soils Information System (NASIS). This process is available for the new Windows Pedon Program.

#### What is the role of the private sector and other NCSS partners?

The collection of the data should be done as a partnership involving the private sector. The private sector can recommend and share data to fill data gaps for NCSS database such as soil quality, precision agriculture, water quality, carbon sequestration, soil chemical and physical properties, etc. Data should be collected by or under the quality assurance of NCSS. Someone in the NCSS must be ultimately responsible for the day-in-day-out quality of the data. At this time, it is someone in NRCS. The private sector should be able to add data with the appropriate computer permissions.

One responder states - We have partners in our state who are as much a part of the NCSS as NRCS. Their activities are considered equal to NRCS. They are full users of NASIS.

The Forest Service has done extensive Bulk Density work in some forests for logging road and skid trail interpretations and regulation. Perhaps this can be incorporated into the database.

#### What standards will be enforced and how?

The standards are developed by the NCSS (Soil Taxonomy, interpretation criteria, etc.). After they are proposed, reviewed, and consensus are reached by members of the NCSS, they are published in the National Soil Survey Handbook (NSSH) or other documents that are used as a standard for making soil surveys. If published in other documents, they are referenced in the NSSH. These standards are enforced by the people who delegated that authority. Appraisals are made by appropriate people to determine if the delegated people are following the standards in the NSSH. The computer software could enforce standards if the NCSS determines this procedure is appropriate.

One responder states - Data needs to be certified by the State Soil Scientist regardless of who collects or enters the data. Standards are in place. MLRA Office has responsibility for quality assurance for data collection and data entry.

#### Who will maintain the soil survey database (NRCS, experiment station, private sector, other)?

First, let us examine the need for a database – use for local, state, and federal programs, use in models for generating soil interpretations, and users wanting soil chemical and physical properties, soil moisture and temperature regimes, and landscape information. The answer in part depends on how “The” is interpreted. Some of our cooperators have different data needs so a lot depends on how flexible NASIS turns out to be. At present, there are fields of data in some cooperator’s soil property and interpretation databases that are not included in the published soil survey reports. If all those fields could be

incorporated into NASIS, there would be no need for maintaining different databases. However, until a complete incorporation is reality, there will be a need to maintain more than one database.

At this time and probably sometime into the future NRCS will be the custodian of the NCSS soil survey database. It is expected that NASIS will be the database. Also, it is envisioned that the maintenance and quality assurance of the soil database will be through NRCS – MLRA Soil Survey Offices and State Office Staffs using NCSS standards. Hopefully, there will be soil scientists to populate the soil survey database and State Soil Scientists whose job is to certify all the data in the soil survey database.

One responder states - Realistically the data needs to be maintained at a local level, either by project office or resource soil scientist. The State Soil Scientist needs to certify the data.

#### How will databases be made available to the public?

The data will be placed on a soil data warehouse and the public will be able to access the data through the internet. It will be available in paper copy and CD's. At the present time, the existing systems (all are not available to all the public) are: National Soils Information System (NASIS), State Soil Survey Database (3SD), Soil Interpretations Records (SIR), Map Unit Interpretations Records (National MUIR), Official Series Descriptions (OSD), Soil Classification File (SC File), National Soil Survey Laboratory Characterization Data (NSSL Data), and Cooperator's of the NCSS data bases. The 3SD, SIR, and MUIR are not supported at this time and will not be necessary after the soil data warehouse is operational. The information should be made available to the public through the internet in an easy to understand, easy to search, and easy to download format. Access to OSD's is a good example of this.

The Public owns the data. NRCS, U.S. Forest Service, Bureau of Land Management, the Agriculture Experiment Station, and Extension Service will continue to provide data. There has not been a real effort to incorporate Agricultural Research Service (USDA-ARS) data into the system, but USDA-ARS collects a lot of point data. With a little effort and the use of GPS locations one could probably capture a lot of their data and incorporate it into the soil map unit databases. USDA-ARS is more inclined to communicate on a soil series basis or even higher abstraction, such as, by soil association area. Data within NASIS are public information and shall not be copyrighted.

This last question to charge 2 is the most important part of the question. We should freeze the data in the data warehouse and make it available to the public. We too often delay the release of our information because we are never quite happy with it. We spend 95% of our time tweaking the last 5% of our projects. We need to freeze it, get it out there and then tweak that last 5%. We can download an updated version in the future. This philosophy needs to be incorporated into the design of the soil data warehouse. Various versions or vintages of the official data need to be stored in the soil data warehouse – each with a time stamp.

#### • Public availability

On-line browsers. The value of information is determined by several factors such as ease of access, timeliness or when it can be accessed, and how easy it is to understand it. Web-based GIS provides real-time access to online geospatial databases. Tools allow users to query, analyze, and visualize data through maps and extracted information.

One responder states - I think the data warehouse is the best solution. I have been receiving increasing requests for downloads of data. When the warehouse is available, this will reduce my workload. It is also the best method of providing the most up-to-date version of the data.

### **Charge 3 – How should we document use-dependent data?**

#### What are the effects of use-dependent and dynamic soil properties on the accuracy of interpretations?

This is where models come in. We can only hope to make good solid measurements and observations in the field for that time that we are set about creating soil surveys. Monitoring of soil properties is often in the domain of other programs such as National Resources Inventory (NRI) and ARS Research Stations. It is imperative that we record the land use condition and history (if known) and the geographic position (latitude/longitude) before we leave the soil profile. If this is done, this information can be used to help model other areas where we do not have measurements and observations.

Soil organic carbon (SOC) is one of the more important "use-dependent" properties. At this time, current soil surveys are built to handle only a few categories of land use condition that could estimate different values of SOC. There are many levels of scale to use-dependent properties (daily, seasonal, annual, el nino/la nina - multi decadal, and longer).

A better understanding of soil climate regime in temporal context with more frequent climate change cycles is presently missing. Again, time and place help to add logic to our soil datasets. There must be a test of this data against these parameters.

One member of the committee stated, "If the data is use dependent we should probably not be trying to collect soil data under each use or for each site. We only need to collect the range of soil values without stratifying by use or site."

A responder felt that there is a lack of documentation to really address this issue. The responder thinks that the data holes should be identified to the research community (universities, sister agencies, etc.). There could be a cooperative agreement to help with the research to fill in the holes. What really needs to be done, is to be able to run a set of interpretations against a pedon that is in NASIS. Unless there has been a misunderstanding, the capability to do this exists. A soil could be described and then enter some assumed data, like permeability, and run some interpretation against it. Then this would allow giving the user information on a specific soil at a specific site.

At this time, pedon level data can be interpreted in NASIS. Current interpretations are directed towards the component and map unit level tables. To interpret pedon data, the applicable interpretive rules, evaluations and properties would need to be written to access the pedon data.

Use-dependent properties can impact the accuracy of an interpretation, depending on the property and the interpretation ( see Attachment B: Use Dependent Data Study – Iowa). The use of these use-dependent properties can give us better interpretations.

#### What use-dependent and dynamic data is especially significant in the North Central Region?

- From an Agronomic standpoint, use dependent data of most concern in the North Central Region would be
  - Surface Soil Structure
  - Infiltration
  - Bulk Density
  - Ksat
  - aggregate stability
  
- Each component may have more than one set of Horizon Soil properties, depending on what "Data mapunit Interpretive Focus" and "Component Cover Kind" is being assessed. In the north central region, for instance, the above use dependent properties would be different for a cover type of "Cropland" than a cover type of "Rangeland" or "Forestland" for the same soil.

### Are changes in NASIS structure needed to accommodate and interpret these soil properties?

- + The current database contains many data elements that can be considered use dependent (e.g. Soil Organic Matter or Carbon). However, current NASIS map unit component data structure and dictionary does not provide adequate use dependent/dynamic soil property values, and the related interpretations, plant/soil interactions, and ecological functions.
- + There is not an organizational framework in place to capture use dependent information.
- + There are no use dependent data (measurements).
- + There are no descriptions of management practices, land use, land use histories, etc., that cause changes in soil properties with time (driving forces).
- + At this time two erosion prediction tools (WEPS 1.0 and RUSLE2) are being implemented. These models are dependent on the near surface properties and hence are highly use dependent (dynamic).

Some limitations of current system:

- + Capability to deliver dynamic soil properties at field level.

Example: Database entry for permeability for the near-surface differs significantly from measurements at the field level.

- + Does not have adequate capabilities to explore relationship between soil properties and management.

Example: The prediction of carbon for carbon sequestration legislation can only be done through a use dependent dynamic soil property database.

### What additional research or data is needed?

The availability of dynamic soil information will facilitate the application of new science in order to more accurately inventory the soil resource, make assessments, and develop better management alternatives.

This will allow NRCS to provide more complete and relevant technical assistance to our clients in order to better enhance and protect our natural resources.

Examples of new science and technology that require new soils data:

Farm-global scale carbon models

Plant growth and soil erosion models

Hydrology models

Precision agriculture

Development of ecological site descriptions and associated state and transition models

Bio-availability of heavy metals (including urban ecosystems)

### **Charge 4– How should estimated data be identified?**

There are different kinds of data:

1. Laboratory data, e.g., mechanical analyses, cation exchange capacity, pH, organic matter, clay mineralogy, bulk densities etc.
2. Calculated data (relational data), for example knowing mineralogy, clay content and organic matter content using the formula one can estimate the cation exchange capacity of the layer. Estimating the data by field method and then correlating to certain properties, for example, using field techniques to determine the Unified class and then correlate it to the liquid limit and plasticity index.
3. Copying the data from a similar soil, for example splitting a series into a typic and oxyaquic that are formed in the same parent material.

### Should estimated data be identified differently than data based on actual measurements?

Yes.

One member of the committee stated, "Most of the data is either estimated or a small measured sample of a much larger population. Often we do not know if the soil property range is normally distributed or where in the range the measured samples fit. I suggest we treat them all as estimates based on measured data. Perhaps we should be expressing the data as a mean and standard deviation rather than a range."

With some amount of doubt in mind, estimated data should be identified. The doubt comes in that there is a fear that there is a substantial amount of estimated data. There is a concern that the public may not accept the data when they realize how much of it is estimated. There is a need to be very professional when an explanation of how the estimates were made. If a good job of explaining what was used to base our assumption and estimate on, the public will accept it. As soil scientists know, most estimates are based on scientifically sound logic and assumptions. There is a need to explain this to the public and not just list a set of items and say that these are all estimates. This could be done easily in a hard copy. A list of properties would be put in the beginning of the Soil Survey, that were estimates and explain how each was arrived at. In NASIS and with the digital data this may be more difficult. A suggestion would be that an explanation of all the estimated data be put into each report. So, when someone generates a report, they get a list and explanation of all the estimated data used in that report. Another alternative is that an explanation could be in the SSURGO metadata file. A similar file could be included with most all data deliveries that is provided to the users.

All of our mapped soil geographic databases (digital maps and attributes) should be considered estimated. It was named "Estimated Soil Properties" on the Form SCS-SOI-5. Only pedon measurements or observations (pedon descriptions) should be considered measurements. A linkage (a geographic or taxonomic one) between real measurements and estimated soil map unit polygons and their component attributes will help assess the reliability of this information for the user. We presently do not use geography in our system; therefore, we cannot provide this information to the user. This can be done, but we will need to change our strategy in data quality review. This could be done at the state or MO level. It could also be done as part of a peer-review process by the University or Private Sector partner.

One responder stated - Most aggregated data is a combination of measured data, scientific guess, and tacit knowledge. Probably there is not very much aggregated data that is based totally on measured data. Is there any value in identifying estimated/measured data unless there is aggregated data that is totally based on measured data?

### How might this be done in the database?

Certainly, soil scientists should know the source of the data, how many profiles or pedons were averaged, etc., but there will not be complete data on all soils in our lifetime or probably two or three lifetimes at the present rate of data collection. However, even if there were a lot of data on many soils, there would still be a problem of how the data are presented. The NCSS can effectively use the concept of benchmark soils and use modeling methods to estimate properties of other soils that have similar characteristics or are grading between two benchmark soils. Therefore, can there be one database that contains the hard data that can be used to estimate properties of other map units? Perhaps most of the data used to make interpretations in the soil survey report will be made based on estimated data that has been modeled from the measured data. If presented in a professional manner, the general public will accept the best judgement of qualified soil scientists.

Some alternatives for indicating the type of data in the soil database are:

- + Under the remarks in the typical pedon description, a statement can be made that the properties are estimated i.e., available water capacity, CEC, Ksat, liquid limit, plasticity index are estimated.
- + In the NASIS database, it can be shown with an asterisk and then can be explained as footnote how the data was arrived.
- + By inserting a letter with the data like Ksat 2.0-6.0(e). Then define e as an estimated data.
- + If there is a way of entering certain data figures as an *italic*. Others regular or underlined, etc. Computer experts can come up with a way of showing it differently.
- + May be a general introduction, that certain values are calculated or estimated using certain formulas or copied from similar soils that have been determined by laboratory data.
- + Estimated data in the soil survey database should be defined in the NASIS Metadata overview.

NRCS in conjunction and cooperation with partners of the NCSS.

- + Centralized control of the NASIS database is essential. This will ensure the quality and accuracy of the data being used. Centralized control will allow any NCSS cooperator or private sector soil scientist to enter data after review.
- + Global population of the NASIS database is much more efficient, ensures greater accuracy of data entries and speeds up the time it will take to publish a soil survey update using NASIS.

Two respondents suggested tagging each data element with measured (M), calculated (C), estimated (E), copied (Cp), etc. from similar soil will be a nightmare for data entry person.

One responder stated that this charge is an inappropriate question to be asking. This charge would be more appropriately addressed by database designers. Members of this committee should be providing/stating requirements that the data should be addressing not providing data structure solutions.

Another responder suggested that if the decision is to do this, then this information could be put in the NASIS metadata and perhaps in prewritten material of soil survey reports.

#### Should the degree of confidence in the data be identified? How could this be done?

Yes, the soil survey database should provide for a reliability designation for the observation for the primary soil data elements. The reliability designation should consider scale (one delineation, county/parish, state, or nation), use being made of the data element, method used to collect the information to assign the numerical value for the soil data element, and other needed items .

A committee needs to be established by the soil survey leadership to resolve the issue of communicating a better understanding of the meaning of data supplied by the soil survey database. The need for describing reliability is recognized by both those making soil surveys and those using soil information. The soil survey database will accommodate the degree of confidence after one is developed and approved for use by the National Cooperative Soil Survey.

Acknowledgement: the chairperson wants to thank Samuel J. Indorante, Soil Survey Project Leader, and Dwayne R. Williams, Soil Scientist, MLRA Office, Carbondale, Illinois for their contribution during the preparation of this committee report.

## **RECOMMENDATIONS:**

The recommendations in this committee report supplements the recommendations that were in the June 18-22, 2000 published report for Committee 2 “Data Acquisition for Problem Solving” of the North Central Work Planning Conference held in Grand Rapids, Michigan.

The members attending the North Central Soil Survey Work Planning Conference agreed that this committee should be terminated unless there are new charges for this committee to address.

Recommendations for the Charges –

### **Charge 1 – What additional data is needed in NASIS**

#### What data do NASIS users need that is not in NASIS?

Currently, there are sufficient soil properties and interpretations in NASIS. Two responders stated that a geographic linkage for each polygon of a NASIS map unit is needed. Two responders stated that there is a procedure already established for NCSS cooperators to request additional NASIS data elements.

#### Data model that incorporates both spatial and attribute data.

Yes, the data model needs to incorporate both spatial and attribute data. Provide capability to record information (soil properties, landscape, climate, etc.) for each polygon rather than doing it only at the map unit level (provisional, approved, additional, and correlated). Populate landscape location of all components for each polygon in map units, especially for those that are contrasting and limiting. One responder questions if this is within the scope of the NCSS program.

#### What data is needed for simulation models and soil quality assessments?

Before a data element is added to the soil survey database for use by modelers, a determination of how important it is and what is the sensitivity of the data element to the model is needed? One responder stated, “If the only reasons (we are collecting the additional use-dependent data) are for simulation models and soil quality assessments, then I would question if it is worth the time and effort. Creators of simulation models can generate their estimates with or without the soil data and these models are useful primarily when measured data is not available. If we have the required measured data we often will not need the models.”

#### What should be the priorities for populating data elements.

Identify those basic soil chemical and physical properties that are common to most of our USDA programs and other users, and then logically populate those properties within a consistent landscape region (subdivision of MLRA, which could be parts of a county or counties, or MLRA).

### **Charge 2 – Who will collect, enter, and maintain the data?**

#### Who will collect and enter the data?

NCSS cooperators with access to NASIS and given permission, by the State Soil Scientist, to populate the database.

Other soil scientists whose work is acceptable by the NCSS can collect data. A distinction needs to be made between point data and aggregated data. Aggregated data entry needs a "gatekeeper" to ensure quality. Until a method is developed to incorporate others not familiar with the NCSS correlation process, their data need to be evaluated by a “gatekeeper” before it becomes part the NCSS database.

There is a real need for good translation software to take full advantage of existing soil knowledge gathered by the Universities and private sector. Perhaps a 3<sup>rd</sup> party could develop such software?

What is the role of the private sector and other NCSS partners?

The private sector soil scientists can be full partners in the NCSS, but their activities may be limited by legal, clientele, and/or professional statutes. Partners of the NCSS contribute to the funding, mapping and technical phases of soil survey, testing soil interpretations and performance data, and providing recommendations for improvement of the soil survey.

What standards will be enforced and how?

The standards that will be enforced will be those developed and adopted by the NCSS. These standards are enforced by the people delegated that authority. Appraisals are made by appropriate people to determine if the delegated people are following the standards in the NSSH. The computer software could enforce standards if the NCSS determines this procedure is appropriate.

Who will maintain “the” soil survey database (NRCS, agriculture experiment station, private sector, other)?

First, let us examine the need for a database – use for local, state, and federal programs, use in models for generating soil interpretations, and users wanting soil chemical and physical properties, soil moisture and temperature regimes, and landscape information. The answer in part depends on how “the” is interpreted. Some of our cooperators have different data needs so a lot depends on how flexible NASIS turns out to be. At present, there are fields of data in some cooperator’s soil property and interpretation databases that are not included in the published soil survey reports. If all those fields could be incorporated into NASIS, there would be no need for maintaining different databases. However, until a complete incorporation is reality, there will be a need to maintain more than one database.

How will databases be made available to the public?

The data will be placed on a soil data warehouse and the public will be able to access the data through the internet. It will be available in paper copy and electronic formats.

**Charge 3 – How should we document use-dependent data?**

Documentation of use-dependent data is dependent on the management method for soil survey. Three methods are available in NASIS to allow for the population of these data:

- A. Delineate separate map unit polygons based on use—thereby creating and populating data mapunits for each use.
- B. Use existing map unit polygon delineations, with the one data mapunit linked. Inside the data mapunit, identify those use-dependent components and populate its properties (component by use), e.g., eroded vs. non-eroded, irrigated vs. non-irrigated, drained vs. non-drained, cropland vs. grassland, etc. The “earth cover kind 1 and 2 data elements are available to facilitate the recording of use-dependent components.
- C. Use existing map unit polygon linked to more than one data mapunit. Each data mapunit would contain the specific use-dependent properties for the given soil component (data mapunit by use).

What are the effects of use-dependent and dynamic soil properties on the accuracy of interpretations?

The assumption would be that more specific soil property data by land use could result in more precise interpretation. This would depend on the sensitivity of the model to that soil data element.

Use-dependent soil data for the soil survey database will be presented at the level above point data. Simply cost alone prohibits the collection of point data as use-dependent data. Use-dependent data at a point changes depending on natural phenomena (fire), harvesting (woodland, cropland, etc.), management (type of conservation tillage), unusual climatic events (temperature, precipitation, etc.), etc. Today, models have algorithms to calculate changes in use-dependent data during the year for specific land uses. Use-dependent data for the soil survey database will be presented above the level of point data and then only if there is a statistical difference between land uses and management that is important to the users.

What use-dependent and dynamic data is especially significant in the North Central Region?

They are surface soil structure, infiltration, bulk density, Ksat, organic matter, available water capacity, and aggregate stability.

Are changes in NASIS structure needed to accommodate and interpret these properties?

The current NASIS map unit component data structure and dictionary is capable to provide adequate use dependent/dynamic soil property values, and the related interpretations, plant/soil interactions, and ecological functions. There is a need for an organizational (temporal/geographic) framework in place to capture use-dependent information. There needs to be descriptions of management practices, land use, land use histories, etc., that cause changes in soil properties with time (driving forces).

What additional research or data is needed?

Farm-global scale carbon models, plant growth and soil erosion models, hydrology models, precision agriculture, development of ecological site descriptions and associated state and transition models, and bio-availability of heavy metals (including urban ecosystems).

**Charge 4 – How should estimated data be identified?**

Should estimated data be identified differently than data based on actual measurements?

Most committee members agree that aggregated data developed for soil map units using one or both - “measured and/or estimated data” - need to be identified as estimated data. There is a need to explain point data to the users. Point data can be either measured or estimated. This could be done easily in a hard copy. A list of properties would be put in the beginning of a report, that were estimates and explain how each was arrived at. In NASIS an explanation of all the estimated data could be put into each report. So, when someone generates a report, they get a list and explanation of all the estimated data used in that report. Another alternative is that an explanation could be in the SSURGO metadata file. A similar file could be included with most all data deliveries that is provided to the users.

Aggregated data developed using measured, estimated, and calculated (calculated refers to data developed from algorithms or other methods) data need to be identified differently than either estimated or measured data unless it is considered estimated.

How might this be done in the database?

Two responders stated that the method of indicating estimated and measured soil properties in the soil database needs to be developed by the people doing the programming. Soil scientist’s responsibility is to decide if estimated and measured soil properties will be indicated in the soil database.

Some alternatives for indicating the type of data in the soil database are:

- + Under the remarks in the typical pedon description, a statement can be made that the properties are estimated i.e., available water capacity, CEC, Ksat, liquid limit, plasticity index are estimated.
- + In the NASIS database, it can be shown with an asterisk and then can be explained as a footnote how the data was arrived.
- + By inserting a letter with the data like K sat 2.0-6.0(e). Then define e as an estimated data.
- + If there is a way of entering certain data figures as an *italic*. Others regular or underlined.
- + May be a general introduction, that certain values are calculated or estimated using certain formulas or copied from similar soils that have been determined by laboratory data.

Estimated data in the soil survey database should be defined in the NASIS Metadata overview.

Should the degree of confidence in the data be identified? How could this be done?

Yes. A committee needs to be established by the soil survey leadership to resolve the issue of communicating a better understanding of the meaning of data supplied by the soil survey database. The need for describing reliability is recognized by both those making soil surveys and those using soil information. The soil survey database will accommodate the degree of confidence after one is developed and approved for use by the National Cooperative Soil Survey.

## ATTACHMENT A

### Soil Moisture Status-- Proposal for a new data element in NASIS

Currently, many loamy, clayey and silty soils in the "Western Udic" region are considered freely drained and lack saturation in the series control section (within 60 inches of the surface). It has been well documented however, that many of these same soils also have redox depletions and concentrations above 60 inches. The past explanation for this was that these redox features were relic.

Piezometer data from many scattered and varied sources, however, shows that many of these soils have saturation in some years. If one makes the correlation that these soils are wet during high rainfall years, it means they are saturated about 25 percent of the time at Brookings South Dakota and about 20 percent of the time at Highmore South Dakota. Since NRCS defines saturation in a profile as that which occurs in normal years (mean values) and it lasts for about one month (20 consecutive or 30 cumulative days), this actual lower frequency saturation goes unaccounted for and uninterpreted in our reports and database.

There is a concern about transitory soil saturation for some uses (building suitability, drainfield design, and other urban uses) even if it only occurs during wet years or heavy rainfall events. (low frequency events).

In order to alert users to the potential of saturation in the lower part of the soil profile during years of high rainfall, we propose the addition of a "saturation frequency" data element to the NASIS component month table. We could then use the term "occasional" saturation just as we use the term occasional for flooding frequency. It would be defined as a 5 to 50 percent chance (5 to 50 times in 100 years). We would populate the soil moisture status and "saturation frequency" in NASIS based on data, observation, morphology, tacit knowledge, etc.

In the map unit description we could use a line such as: *Occasional saturation occurs as shallow as 40 inches in the months of April through June during above normal rainfall patterns.*

There are several advantages of doing this: They are:

- + It acknowledges that the redox features are indicative of wetness, albeit only in wet years.
- + It increases the credibility of our work as others in the field of soil science treat these redox features as "real" and not relic. We also have had a hard time defending the term relic as we did not have piezometer data to show it lacked saturation.
- + It alerts users to the potential to have problems with wetness in the lower part of the profile. They can do their own risk assessment based on the fact it is occasional wetness.
- + It resolves problems between states that in the past have stated that the same series was either saturated most years at specified depths or it was not wet at all (by definition of a water table). Being saturated on an occasional basis is likely much closer to being correct in the transitional areas where udic moisture grades to ustic moisture.

## **Attachment B**

### **USE DEPENDENT DATA STUDY - IOWA**

In May 2001, Bob Grossman and Berman Hudson assisted the Iowa soil survey program in sampling a Fayette silt loam (fi-si, m, superactive, mesic Typic Haplustalf) 2-5 % slopes. Soils were described in conventional cropland, reduced tillage cropland, bluegrass pasture, and forest. The samples were collected in Iowa County, Iowa, near the Amana Colonies. Iowa County is currently being updated.

Results show definite differences in aggregate stability, soil pH, CEC, carbon, and bulk density. It is our intention to be able to publish this data with the Iowa County report as a subset of the standard tables. We are hoping to use NASIS to develop a supplementary set of properties and interpretations for this one map unit to demonstrate to the soil survey user that management impacts soil properties and interpretations and that one size does not fit all. We would also develop a pre written material to explain the use dependent data to the user and the significance of the information. We believe erosion factors, hydrologic groups, runoff curves, fertility and productivity, and various other properties would be affected.

The Fayette series is extensive. It is mapped in Iowa, Illinois, Minnesota, and Wisconsin. By selecting only one map unit, we believe the project to be manageable in order to complete a subset update on schedule. The data should be applicable to other areas and to similar soils. We plan to collect use dependent data on other series in conjunction with other soil survey updates and would be very comfortable using use dependent data from other states.

If anybody would like a comprehensive set of the laboratory analyses and Bob Grossman's in field measurements, contact [mike.sucik@ia.usda.gov](mailto:mike.sucik@ia.usda.gov)

Table 1. Comparison of surface horizon properties from different management/land uses on the Fayette soil in Iowa County, Iowa.

Management	Sand	Silt	Clay	Total			CEC/ clay	Db	Water content	CEC	pH	Water dispersible clay		Aggregate stability	Morphology index
				N	S	C						1/3	2/3		
									----	1/3 bar----	1:1				
Conv. tillage	3.2	80.2	16.6	0.157	0.02	1.45	0.86	1.47	26.0	14.2	6.9	5.9	5.9	7	58
Pasture	3.8	77.6	18.6	0.266	0.04	2.17	0.93	1.39	24.6	17.3	6.0	5.6	5.6	16	70
Reduced tillage	3.8	76.7	19.5	0.282	0.04	2.23	0.91	1.38	27.8	17.8	6.6	6.3	6.3	11	83
Woods	1.3	80.3	18.4	0.302	0.05	3.74	1.19	1.23	25.9	21.9	7.4	6.0	6.0	85	88

Observation of the data in Table 1:

1. The data is not replicated, only one site was sampled for each management system or land use. Therefore no statistical analyses could be used to indicate significant differences. It is recommended that the sites be replicated. To have three replications, this would require having three sites that are in woods, three sites that are in pasture, three sites that are in conventional tillage, and three sites that are in no-tillage.
2. The total carbon content was the highest for the woods site and lowest for the conventional tillage site. The pasture and reduced tillage sites had similar total carbon contents. In general, the total carbon content appears to decrease with the increased amount of soil disturbance.
3. The increase in CEC appears to be due to the increase in total carbon.
4. The bulk density appears to decrease with increasing amounts of total carbon. The woods site has the lowest bulk density, while the conventional tillage site has the highest bulk density.
5. Total N and S increase with increasing total carbon.
6. There appears to be no trend in the water content, pH, or water dispersible clay.
7. For aggregate stability, the least disturbed soil condition with highest total carbon amounts has the greatest aggregate stability. The soil condition with the greatest soil disturbance with the lowest total carbon contents has the lowest aggregate stability. There is a problem with this method in the estimation of the sand content, therefore the data may be somewhat inaccurate.
8. The morphological index appears also to follow the same trend as the total carbon content.

Summary: The biggest differences in surface horizon properties (total carbon, bulk density, CEC, aggregate stability, morphological index) were between the conventional tillage site and the woods site. The woods site had a higher total carbon content, a greater CEC, a higher percentage of stable aggregates, a higher morphological index, and a lower bulk density than the conventional tillage site. The pasture and reduced tillage sites have a similar soil condition, which is better than the conventional tillage site, but not as good as the woods site.

Table 2. Saturated hydraulic conductivity (Ksat) and infiltration rate measurements on the Fayette soil in Iowa County, Iowa. The values represent the average of three measurements measured one meter apart.

Management	Ksat	Infiltration sprinkler‡	Infiltration FH 5 min §	Infiltration FH long-term
	in/hr	in/hr	in/hr	in/hr
Conv. tillage	0.21 (0.13) †	3.4	3.87 (1.76)	0.51 (0.78)
Pasture	0.11 (0.03)	0.87 (1.41)	4.90 (1.12)	0.094 (.006)
Reduced tillage	0.05 (0.02)	2.63 (0.75)		
Woods	0.91 (0.50)	7.83 (3.57)		

† Number in parentheses is the standard deviation.

‡ Infiltration rate determined with the Cornell sprinkler infiltrameter.

§ The falling head (FH) method was wetted the previous day with 10 cm of water, the next day 5 cm of water was placed in the cylinder for 5 minutes and infiltration rate measured, and then a second 5 cm of water was added and the long-term (1-2 hr) infiltration rate was measured.

1. There were three measurements for each test, so statistical analyses were conducted to determine if there were significant differences between the management system or land uses. However, the three measurements are not true replicates, they are subsamples of one replicate or site. To have replication, three sites of the same management system or land use would need to be selected and measurements conducted. In other words, we would need to have two more sites for each of the management systems or land uses listed in Table 2 to have three replications (on the Fayette soil). Replication is important for getting estimates of the spatial variability.
2. For Ksat, there were no significant differences in Ksat between the conventional tillage, pasture, and reduce tillage sites. The woods site had a significantly higher Ksat than the Ksat on any of the other three sites.
3. For infiltration using the sprinkler infiltrameter, there were no significant differences in infiltration between the pasture and reduced tillage site. The conventional tillage site could not be analyzed because only one measurement was taken. The woods site had a significantly higher infiltration rate than either the pasture or the reduced tillage sites.
4. For infiltration using the falling head-5 minute or the falling head-value methods, there was no significant difference in the infiltration rate between the conventional tillage and pasture sites. Infiltration, using these methods, was not measured on the reduced tillage and woods sites.

Summary: The woods site had a significantly higher Ksat and infiltration rate than the Ksat and infiltration rate on the other three sites. There were no significant differences in the water transmission properties between any of the other three sites (between conv. tillage, pasture, or reduced tillage sites). The high spatial variability at each site could be masking any possible significant differences (refer to the standard deviations in Table 2)

