Dylan,

This is a summary of some of the things that we talked about today and some comments on topics we may have missed.

Horizon Designations:

In the Lab Data Mart there are two columns for Horizon designations. One contained what was called Original horizon designations and the other is simply called horizon designations.

Last time I knew the Original horizon designations were static and represented designations in the database as of several years ago. They may or may not be original designations.

The other column contained the designations that come from the descriptions on the tags on the bags or were updated versions of the horizon designations in the original designation column.

Ideally both columns would be populated with the same information from the tags when the samples come in. Then over time the current one may change if it is updated at a later date. The reason you want the latest horizon designations in a single column is so that you can query and sort by it.

In NASIS there is a table called NCSS Lab Layer. In that table there is room for two designations. Originally the designations came from the Lab Data Mart. I then queried NASIS and tried to update the one column with the most recent horizon designation in NASIS. I queried NASIS for the Oldest pedons and updated the original designations with the oldest designation that I could find.

In theory the NCSS Lab Layer table contains the most up to date designations and the oldest designations in the database. There is much room for improvement. The reason for this is because NASIS and the laboratory database originally only had room for one designation. Therefore as folks decided to update the designations they had to modify the only place in the database that it existed.

There are queries and reports in NASIS intended to help people compare designations in different databases and from various pedons to get the latest horizon designations.

The reason that the whole process can not be automated is because the horizon designations sometimes are really layer or sample designations and the samples may come from more than one horizon or part of a horizon.

Another important note. In the laboratory database there is another column. I can not remember what it is called, but it contains non-horizon designations. That is all the junk that was originally in the horizon designation column but got moved into this other column. Things like dense layer, yellow seam, red seam and such got moved into that column. I can't remember if we put that column in NASIS but I don't think so. I think it is only in the layer table in the lab database.

As for Depths.

There are depths in the Lab Data Mart that hopefully have been updated to the most recent standards. I think that I did do that. However there is also the possibility that I might have missed some.

The toughest part was deciding what to do with missing lower depths. Sometimes the lowest depth was filled out with 150 or 202 or possibly with the depth of bedrock... There were various conventions used over the years so you will find many different things. Some people added 10 cm to the depth to rock, some added 25 cm...

I updated the depths in the NCSS Lab Layer table to reflect the most recent standards for measureing horizons. Those are the same depths that I hope to have put in the Lab Data Mart, but I can not remember if that was done.

I updated the depths in some of the pedons in NASIS that I owned to match the sample depths. Since I did not own all pedons it is possible that some pedons in NASIS never got updated.

There are new columns in the Pedon Horizon Sample table. These depth columns should have depths that match the sample depths exactly rather than the horizons. The intent is not for people to vary from the horizon depths except in the cases where horizons were combined or horizons were split for some reason or other. In general at least one of the two depths would correspond to the top or bottom depth of a horizon unless the samples were completely by depth. For example 0-5 5-10 10-15 15-20 20-25...

The intent for adding all the new columns in the Pedon Horizon Sample table is to allow the field to fully populate the table in NASIS and print out the report to accompany samples to the lab. This should reduce the chance of transcription errors and help the field make sure that samples do not get mixed. up.

Publishing data

There is a report in NASIS that is used to search for the most recent pedon, location and calculation. I think that I have a powerpoint presentation or two and a video or two that describes this process.

The report is run on NASIS and data mines NASIS for values which are used to update the Lab Data Mart. Rick has some checks on the classification so if the parts do not add up to the calculated whole then no updates are made.

The report can be used to update the Lat and Long as well as the classification.

The report is usually only used to update the database once a year, or once a quarter when the ACCESS database was created.

The goal of placing the metadata in NASIS is so the 6 tables will be completely documented. Those 6 tables plus a NCSS Site Location and NCSS Pedon Taxonomy table would be enough to place in an ACCESS or SQL Light database for distribution. They also would be great if a web service such as Soil Data Access were Pointed to them.

The NASIS documentation could be used to help document the data in the web service.

There are some columns for mehlic3 that I believe have been removed from the Chemical Properties Table. The corresponding metadata for those columns would need to be removed from NASIS. The scripts(views) in the Lab Data Mart need to be compared to the NASIS metadata to determine which need to be removed.

Rick would prefer to only upload the changes to Kansas city rather than whole tables.

We discussed the possibility of a unique identifier of the lab sample number and the preparation code to identify each row. We also discussed the possibility of just concatenating the entire rows and comparing them and replacing the rows that did not match.

If we did that the process that I described in the email would go as such.

1. Create a new table using the views and call it update.

1. Compare the new table to the published table and identify appends and replacements.
2. Compare the Published table to the update table to identify deletes.

If the replication process could identify which rows had been appended, replaced and deleted, that would be all that would be required.

It is suggested that we share the site table in the Lab DataMart as the first Web Service. This will allow folks to plot all of the laboratory sample points. One of these days it would be good to get a layer of tribal lands for which data is not to be shared. We could then intersect the points and remove any laboratory data that should not be shared. It is also possible that those agreements might change over time, so at some future date they might want the data shared.

If we share the site table in a web service it should be scanned for PII data. Hopefully there is none.

There are calculations in NASIS for updating horizon sample numbers, pedon laboratory sample numbers, user pedon ids, and user site ids. There are also calculations for updating any number of other tables like the NCSS Layer and NCSS Pedon tables.

I am sure I will think of more, but this is a start.

The Skeleton pedons:

There are skeleton sites and pedons for thousands of pedons in NASIS. A chunk of them are from Missouri. I think there might be another lab or tow that we made some for as well. There are about 8000 skeletons for pedons and sites that were in the repository.

*DEB: it seems like many of the skeletons are only site objects, but still manage to have pedon IDs. How can this be? See user site ID* ***79CA000003*** *as an example. You can only access this thing via site-only query in NASIS.*

Basically I used ODBC connections between a copy of NASIS that is in Lincoln and the laboratory database. I looked for differences in the Laboratory sample numbers.

If there were laboratory sample numbers in NASIS that were not in the Lab Data Mart I couldn’t really do much about that. If there were laboratory sample numbers in the lab data mart that were not in NASIS I was able to make these skeletons.

Over time it is hoped that the field would identify real pedons and replace the skeletons. Some people have just updated the skeletons and moved them to their ownership.

The goal is that there will be no more skeletons and that they will be either deleted, or moved to the ownership of the appropriate offices for maintenance.

The overall stepwise approach.

1. Create a web service for the site table.
2. Figure out how to update the Google Fusion Table/Map on the fly using the web service.
3. Create the easiest of the 6 tables in the Lincoln copy of the lab data mart.
4. Document a way of updating those tables that are created.
5. Get the remainder of the 6 tables created in the Lincoln copy of the lab data mart
6. Get the 6 tables replicated to Kansas city
7. Institute the refresh cycle for the 6 tables. Could be yearly, biannually, quarterly, monthly, weekly, or daily depending upon the automation tools involved.
8. Create a SQL Light and or ACCESS database of the 6 tables and get it posted on a regular basis whatever that interval will be.
9. Either include or point users to the metadata for the 6 tables.
10. Consider adding the 6 tables to NASIS.
11. Evaluate the possibility of editing in NASIS and updating the Lab Data Mart from NASIS
12. Identify the “authoritative source of the data so the work-flow would not accidently overwrite valid corrections.