# Parent Material Simplification/Normalization Project

## Background

There is a vast amount of field-based knowledge about environmental properties and processes contained in the data structure of the national soil survey database. Creating pathways for sharing that knowledge with our customers is critical, so that individuals and society can make informed land use decisions.

Creating products for customers from the soil survey database is difficult for two primary reasons. One is the complexity of the data structure. The other is that soil survey work has been managed within administrative boundaries over the course of many decades. The result is that there are differences in how the data have been populated for different areas across the landscape. For both those reasons, both data analysis and visual products can be difficult to understand.

Two very useful soil properties contained in the component/map unit data structure are parent material *kind* (process) and *origin* (geology). These data represent tens of thousands of observations of parent material gained during decades of soil survey work in the field. Parent material *kind* represents a process. Parent material *origin* is the geologic origin of the parent material kind. Delivering these two fundamental properties as data for analysis or as map products, can inform many kinds of land use decisions. Examples of products and uses for these products are summarized below in the Outcomes section.

## Problem

There are several categories of problems with the parent material data in the soil survey database. These include naming conventions, data complexity, and differences in scale of application for the same data. These problems result in two challenges. One is the difficulty of using the data to elucidate relationships between the soils and these properties across the landscape. The other challenge is the visual confusion created by different, but conceptually duplicative, categories. Both challenge our ability to understand environmental processes across landscapes.

We would like to have the flexibility of scaling the complexity up or down, depending on the analysis chosen. But we want to be reasonably sure that the base level of complexity represents both real and useful differences. We also want to retain historical knowledge that captures real differences across the landscape.

## Outcomes

We want to produce useful products, that impart knowledge of soil properties and landscape processes, for internal and external customers. These products include:

* Thematic maps that represent varying levels of complexity, depending on user preference or delivery context.
* Raster files of parent material kind, origin, and the combination, for analysis in digital soils mapping and other modeling efforts.
* Underlying source data that minimize complexity where warranted
* Retention of historical knowledge where valid (though it is probably impossible to determine where subtle differences might or might not be valid).

## Methods and Procedures

There are three major (with current funding and staff resources) approaches:

1. Create new, or edit, source data in NASIS for delivery of ‘refined’ data to SSURGO. Extraction of data for analysis and display products can be done from SSURGO delivery systems.

* Advantages: Every customer has access to the refined data, and maps and analyses can be created from the official data.
* Disadvantages: Difficult (impossible?) to create automated processes that everyone can agree on for all surveys. Potential for loss of historic knowledge accessible by the public or even internal users.

1. Create records in NASIS that can be used internally, but don’t deliver to the public via SSURGO. Data products could be delivered but it wouldn’t be based on the official data

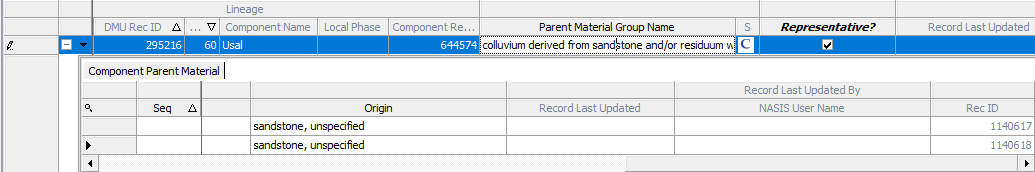
* Advantages: The data are maintained in the institutional database (i.e. use Sequence column in the parent material child table to note refined data?) with metadata about the origins of the records.
* Disadvantages: Products such as raster files and maps wouldn’t be based on the official data available to all customers. Official data, with contradictions and overlapping concepts, remain available for use to the public.

1. Extract data from NASIS or SSURGO and manipulate it in an outside data management environment.

* Advantages: No risk to historical data. No edits required to official data or to internal records.
* Disadvantages: Official data, with contradictions and overlapping concepts, remain available for use to the public. No institutional framework for retaining the record of decisions or the resulting data and products.

Method 1 is the preferred alternative, If the disadvantages listed can be resolved or mitigated

### Possible Method 1 Procedures



Want:

New single Parent Material child table with cleaned kind and origin

or

Parent Material Group Name with rv checked, calculated from old Parent Material choices.