

SSURGO Data Packaging and Use

December 2003

United States Department of Agriculture
Natural Resources Conservation Service

Table of Contents

INTRODUCTION	5
SSURGO EXPORT PACKAGING	5
CONTENTS OF DIRECTORY "SOIL_SSASYMBOL"	6
CONTENTS OF DIRECTORY "TABULAR"	7
CONTENTS OF DIRECTORY "SPATIAL"	7
<i>Non-Spatial Attributes Embedded in the Spatial Data</i>	9
DATA VERSIONING	10
LOCATING VERSION INFORMATION	10
MAKING SURE THE TABULAR DATA VERSION IS COMPATIBLE WITH THE SPATIAL DATA VERSION	11
USING SOIL TABULAR DATA	12
USING A SSURGO TEMPLATE DATABASE	13
USING SOIL SPATIAL DATA	13
PERSONAL GEODATABASE	13
THE CHALLENGE OF USING SOIL DATA IN A GIS	14
SOIL DATA VIEWER APPLICATION	16
CONTACTING SUPPORT	17

Introduction

Soil survey area data that is exported from the Soil Data Mart is provided in what is referred to as “SSURGO” format. This format defines which attributes are provided, how those attributes are grouped, and how those groups are related.

There are two categories of soil survey area data, tabular and spatial. Tabular data consists of records and fields. Tabular data is typically manipulated in a database or spreadsheet. Spatial data consists of the information needed to construct a graphic soil survey area map. Spatial data is typically manipulated in a Geographic Information System (GIS). SSURGO spatial data is provided in vector format.

In order for a soil survey area to be added to the Soil Data Mart, tabular data is required to exist, but spatial data is not required. For a survey area that is still in the process of being mapped, tabular data may be available before the corresponding spatial data is available. Even when both tabular and spatial data are available, tabular data can be exported without exporting the corresponding spatial data, and vice versa.

SSURGO Export Packaging

When soil data is exported from the Soil Data Mart, the end result is always a single zip file, regardless of what export options were selected. The format of an export file name is: *soil_ssasymbol.zip*, where *ssasymbol* is the symbol of the corresponding soil survey area. A soil survey area symbol uniquely identifies a soil survey area. An export from the Soil Data Mart always contains data for one and only one soil survey area.

There are two kinds of survey area symbols.

The most common kind of survey area symbol is a five character symbol where the first two characters are a U.S. state or territory postal code, and the last three characters are digits that represent the soil survey area within the corresponding state or territory.

The less common kind of survey area symbol is a six or seven character symbol where the first two characters are “SS” and the remaining digits and optional ending letter identify a Major Land Resource Area (MLRA). For additional information on MLRAs, please visit:

http://soils.usda.gov/soil_survey/geography/mlra/main.htm

A SSURGO export file can be unzipped using WinZip or an equivalent application. When an export file is unzipped, the following directory hierarchy is produced in the directory to which the export file was unzipped:

soil_*ssasymbol*
tabular
spatial

where *ssasymbol* is the symbol of the corresponding soil survey area.

Contents of Directory “soil_*ssasymbol*”

In addition to subdirectories “tabular” and “spatial”, this directory contains either three or four files:

1. soil_metadata_*ssasymbol*.txt
2. soil_metadata_*ssasymbol*.xml
3. README.txt
4. *something*.zip (optional)

The first two files contain the FGDC metadata for the corresponding survey area, in plain ASCII and XML format, respectively. FGDC is the acronym for Federal Geographic Data Committee. The FGDC metadata primarily pertains to spatial soil data. A smaller portion of the FGDC metadata pertains to the tabular soil data.

File “README.txt” contains a lot of the same information that is presented in this document. The top of this file documents the versions of the data included in the export, as well as any options that were specified if spatial data was included in the export.

File “*something*.zip”, if it exists, is a zipped Microsoft Access database, into which the tabular soil data can be imported. This file will only exist if the person who generated this export requested its inclusion. The embedded Microsoft Access database is referred to as a “SSURGO template database”. There is more than one SSURGO template database, so the non-extension portion of the file name can vary. SSURGO template databases are discussed in the section titled “Using Soil Tabular Data”.

If a SSURGO template database was not included in the export, one can always be downloaded from the following Soil Data Mart web page:

<http://soildatamart.nrcs.usda.gov/Templates.aspx>

Contents of Directory “tabular”

This directory contains a set of ASCII field and text delimited files. Fields are delimited by the pipe or vertical bar character. Text fields are double quote delimited because such a field can contain embedded field delimiters and end of lines. Within a text field, any embedded double quotes are doubled.

With the exception of “version.txt”, each of these files corresponds to a table in a SSURGO database. Which ASCII delimited file (“Import/Export File Name”) corresponds to which table (“Table Physical Name”) is documented in the “SSURGO Metadata – Tables” report.

File “version.txt” records the SSURGO version of the corresponding tabular data. The contents of this file is checked when tabular data is imported into a SSURGO template database.

Contents of Directory “spatial”

The exact contents of this directory depends on the spatial format option that was selected when the export was generated. Spatial data can be exported in any of the following formats:

1. ArcView Shape File
2. ArcInfo Coverage
3. ArcInfo Interchange

In addition to specifying a spatial format, a user exporting soil spatial data from the Soil Data Mart can specify a coordinate system and projection. When spatial data is included in an export, the corresponding spatial format and coordinate system are documented in the “README.txt” file, located in the root directory that was created by unzipping the export file. The export spatial format and coordinate system are also documented in the FGDC metadata, also located in the root directory that was created by unzipping the export file.

In discussing the spatial data, it helps to separate the logical from the physical. The possible logical spatial entities are:

1. Soil Survey Area Boundary Polygon(s) (Required)
2. Map Unit Boundary Polygons (Required)
3. Line Map Units (Optional)
4. Point Map Units (Optional)
5. Line Spot Features (Optional)

6. Point Spot Features (Optional)

7. Spot Feature Descriptions (Optional)

Required means that, if spatial data for the corresponding survey area exists, this data must be included. In other words, if spatial data exists, at a minimum, soil survey area boundary polygons and map unit boundary polygons will exist.

Each of the first six spatial entities represents what ArcInfo refers to as a “feature class”. This means that they corresponding to something that appears on a map. The last entity, Spot Feature Descriptions, is an ASCII field and text delimited file that contains narrative descriptions of any corresponding line or point spot features. In effect, this is “tabular data” that is considered to be part of the spatial data.

A single logical spatial entity is delivered as multiple files, sometimes with one or more subdirectories, depending on the spatial format that was selected when the export was generated. Some, but not necessarily all, of the files corresponding to a particular logical spatial entity have a name whose prefix denotes the corresponding spatial entity.

For spatial data in ArcView Shape File format, the following file name prefixes are used:

File Name Prefix	Spatial Entity
soilsa_a	soil survey area boundary polygons
soilmu_a	map unit boundary polygons
soilmu_l	line map units
soilmu_p	point map units
soilsf_l	line spot features
soilsf_p	point spot features
soilsf_t	spot feature descriptions

For spatial data in ArcInfo Coverage or ArcInfo Interchange format, the following file name prefixes are used:

File Name Prefix	Spatial Entity
ssa_a	soil survey area boundary polygons
smu_a	map unit boundary polygons
smu_l	line map units
smu_p	point map units
ssf_l	line spot features
ssf_p	point spot features
ssf_t	spot feature descriptions

We haven’t discussed the gory details about ALL spatial files and directories that may have been created. The reason for this is that the software you would use to manipulate the spatial data

most likely only needs to know the corresponding spatial format and root directory location of the spatial data. The GIS software will “take care of everything else” for you.

Non-Spatial Attributes Embedded in the Spatial Data

A handful of non-spatial (non-geometry) attributes are embedded into the various spatial entities. These attributes either logically identify a corresponding tabular entity, serve as the physical link to the corresponding tabular entity, or identify the corresponding spatial data version. See the section titled “Data Versioning” for more information about data versioning.

Attribute	Present in the Following Spatial Entities	Purpose
Survey Area Symbol (column name is always “areasymbol”)	All	Logically identifies the corresponding survey area.
Spatial Data Version (column name is either “spatialversion” or “spatialver” – for some spatial formats, column names are limited to no more than ten characters)	All	Logically identifies the corresponding spatial data version.
Legend Key (column name is always “lkey”)	Survey Area Boundary Polygons	Links survey area boundary spatial record to the corresponding legend table record.
Map Unit Symbol (column name is always “musym”)	Map Unit Boundary Polygons, Line Map Units, Point Map Units	Logically identifies the corresponding map unit.
Map Unit Key (column name is always “mukey”)	Map Unit Boundary Polygons, Line Map Units, Point Map Units	Links map unit boundary spatial record to the corresponding mapunit table record.
Spot Feature Symbol (column name is always “featsym”)	Line Spot Features, Point Spot Features, Spot Feature Descriptions	Logically identifies the corresponding spot feature.
Spot Feature Key (column name is always “featkey”)	Line Spot Features, Point Spot Features, Spot Feature Descriptions	Links spot feature spatial record to corresponding feature description record.

Data Versioning

With the advent of the Soil Data Mart, data for a soil survey area is now explicitly versioned. There are three different versions:

1. Survey Area Version
2. Tabular Data Version
3. Spatial Data Version

Since tabular soil data for a survey area can be updated without updating corresponding spatial soil data, and vice versa, tabular data and spatial data are versioned independently of one another. For a survey area, any new Tabular Data Version or Spatial Data Version results in a new Survey Area Version.

For any version, there are two version attributes:

1. Version Number (usually referred to as “version”)
2. Version Established Date and Time

Version numbers are serially incremented simple integers, starting at one.

Locating Version Information

Survey area version information can be found in the following locations:

1. In the README.txt file, located in the root directory that was created by unzipping the export file.
2. In the sacatalog table record for the corresponding survey area (sacatalog table or file sacatlog.txt), when tabular data is included in the export.
3. In soil report “Survey Area Data Summary” in the MS Access SSURGO template database, when tabular data is included in the export and that tabular data has been imported into a SSURGO template database.

Tabular data version information can be found in the following locations:

1. In the README.txt file, located in the root directory that was created by unzipping the export file, when tabular data is included in the export.

2. In the sacatalog table record for the corresponding survey area (sacatalog table or file sacatlog.txt), when tabular data is included in the export.
3. In the page footer of soil reports in the MS Access SSURGO template database, when tabular data is included in the export and that tabular data has been imported into a SSURGO template database.
4. In the body of soil report “Survey Area Data Summary” in the MS Access SSURGO template database, when tabular data is included in the export and that tabular data has been imported into a SSURGO template database.

Spatial data version information can be found in the following locations:

1. In the README.txt file, located in the root directory that was created by unzipping the export file, when spatial data is included in the export.
2. Embedded in spatial data for any spatial entity that represents a feature class, when spatial data is included in the export.

Making Sure the Tabular Data Version is compatible with the Spatial Data Version

Soil tabular and spatial data are compatible when:

1. For every survey area polygon feature, there is a corresponding legend table record.
2. For every map unit polygon, line or point feature, there is a corresponding mapunit table record.

Note that every mapunit table record IS NOT required to have a corresponding map unit feature. While this is rare, it is permitted. The case for which this exception is permitted is for a survey area in the process of being mapped, where the spatial data is not yet complete.

If the export file includes both tabular and spatial data, that tabular data is compatible with that spatial data. Where things can get out of sync is when tabular data is obtained independent of corresponding spatial data, or vice versa.

On your PC there is no easy way to verify if a set of tabular and spatial data are compatible. The most reliable way to verify compatibility is to visit the Soil Data Mart, access the Download page for the survey area in question, and look at the tabular and spatial version numbers displayed on the Download page. If your version numbers don't match the version numbers displayed on the Download page, you should request a new export that includes both tabular and spatial data.

Using Soil Tabular Data

The tabular data as delivered in a SSURGO export file, isn't very usable as is. The tabular data is distributed between approximately sixty ASCII delimited files. In order to effectively use this data, it needs to be imported into a database. Even after loading data into a database, you still need to understand what the tables and attributes represent, how tables are related and what data constraints are in place. This type of information is available on the following web page:

<http://soildatamart.nrcs.usda.gov/ssurgometadata.aspx>

This is also the location of this document.

At the current time we provide a Microsoft Access database into which soil tabular data can be imported. We refer to this database as a "SSURGO template database". In order to use this database, you have to have Microsoft Access installed on your PC.

In a SSURGO template database, the SSURGO database structure has already been created. Tabular soil data can be imported by running a macro that resides in the database. Once data has been imported, a variety of reports can be generated. If you are willing to learn a little bit about Microsoft Access, you can create your own queries against the data you have imported.

If a SSURGO template database was included with the exported data, it will be the only zipped file in the root directory that was created by unzipping the export. If a SSURGO template database was not included with the exported data, one can be downloaded from the following location:

<http://soildatamart.nrcs.usda.gov/templates.aspx>

A number of different template databases are typically available. A national template database serves as the default. For a national template database, the corresponding "state code" in the template database web page data grid is "US".

Some states have created a customized template database for their state. If a customized template database is available for the state in whose data you are interested, you should use that state's customized template database.

When selecting a template database, if possible, select one whose MS Access version is the same as the version of MS Access that you have installed on your PC. At the current time, national template databases are available for Access 97, Access 2000 and Access 2002. What MS Access versions are supported for a state customized template database varies from state to state. While you can convert a template database from one version of MS Access to another, this conversion is not always successful.

At the current time, we do not support the user of any database other than MS Access. If you are interested in loading soil tabular data into a database other than MS Access, and have questions, please contact the NASIS Hotline. See the section titled “Contacting Support” for the details.

Using a SSURGO Template Database

For information about importing tabular data into a template database, or information about the capabilities of the template database in general, do the following:

1. Open the MS Access SSURGO template database in the appropriate version of MS Access.
2. Click the Reports tab of the Database Window. The Database Window may be behind a form titled “SSURGO Import” or “Soil Reports”.
3. Either double-click the report titled "How to Understand and Use this Database", or select this report and then click "Preview".
4. After the Report Viewer window is displayed, either click the printer icon or select "Print" from the File menu. You can also browse this report using the Preview window.

Using Soil Spatial Data

The spatial data can't be used until it has been imported into a GIS. In order to import soil spatial data into a GIS, the GIS must be able to import one of the following spatial formats:

1. ArcView Shape File
2. ArcInfo Coverage
3. ArcInfo Interchange

Obviously the spatial data isn't of much use unless the corresponding tabular data is also available. Once you have successfully imported the soil tabular data into a database, and have also successfully imported soil spatial data into a GIS, it's pretty much uphill from there. See the subsection titled “The Challenge of Using Soil Data in a GIS” for more details.

Personal Geodatabase

If spatial data is available for the survey in which you are interested, and if you or someone you know has access to ArcGIS, a personal geodatabase can be created. To do this, import the tabular data into a SSURGO template database in the usual way, and then use ArcCatalog to

import each of the spatial data feature classes into that same SSURGO template database. The spatial data to be imported must be in either Shape File or Coverage format. The resulting personal geodatabase can be used with ArcGIS or ArcView 8 or later.

The personal geodatabase doesn't know what tables are related unless you explicitly create relationship classes to indicate how two tables are related. If you want to establish relationship classes to any of the tabular data tables, you must register ("register to the personal geodatabase") each of the tabular tables of concern. The relationships between all SSURGO tables are documented in the SSURGO data model diagram, which can be found at:

<http://soildatamart.nrcs.usda.gov/ssurgometadata.aspx>

If you create a personal geodatabase and want to be able to link to spot feature descriptions, the spot feature descriptions must be explicitly imported, but keep in mind that a survey area may or may not include spot features. Spot feature descriptions, when they exist, are bundled and versioned with the spatial data, but the spot feature *descriptions* are not a feature class. To import spot feature descriptions into a SSURGO template database, run the macro titled "Import Feature Descriptions". This macro will import spot feature descriptions into the table named "featdesc". The macro displays a dialog box that prompts for the fully qualified path of the file containing the spot feature descriptions. This file, if it exists, resides in the subdirectory named "spatial", and the name of the file containing the spot feature descriptions starts with "soilsf_t" or "ssf_t" for Shape File format or Coverage format, respectively.

One last thing to keep in mind is the size limit of an MS Access database, which is 1 gigabyte for Access 97, and 2 gigabytes for Access 2000 and later versions. The amount of space required for a given survey area's tabular and spatial data varies widely. You may or may not be able to create a personal geodatabase that includes data for more than one survey area, depending on the size of the survey areas involved. For a single very large survey area, you may be able to create a personal geodatabase, but performance may be so poor as to render it unusable.

The Challenge of Using Soil Data in a GIS

For compatible tabular and spatial data, the relationship between the tabular and spatial data has already been established before that data was ever exported from the Soil Data Mart. You are not responsible for establishing this relationship. (For information about tabular and spatial data compatibility, see the section titled "Making Sure the Tabular Data Version is compatible with the Spatial Data Version". For information about how the tabular and spatial data are linked, see the section titled "Non-Spatial Attributes Embedded in the Spatial Data".)

The challenge is due to the following:

1. Thematic maps are based on map units.

2. The vast majority of attributes for which you will want to create a thematic map are not attributes of a map unit but attributes of an entity that repeats for its corresponding map unit.

Map units are typically made up of one or more named soils, which are referred to as “map unit components”, or just “components”. The majority of attributes you will want to display in a thematic map are attributes of a component. This means that such an attribute has to be aggregated to the map unit level before a thematic map can be generated. If you want to create a thematic map for a soil horizon attribute, you have to deal with selecting either a specific horizon, like the surface horizon, or aggregating the attribute for a particular depth range, even before dealing with the issue of multiple components.

In addition to dealing with aggregating data to the map unit level, you have to decide how to aggregate that data. There are a variety of options:

1. Dominant Component – The interpretation or soil property of the component with the largest percent composition is used to class the map unit. Where two or more components have equal percent composition, the component with the most restrictive interpretation is used.
2. Dominant Condition – For the components in the map unit, the interpretation rating classes or soil property values are grouped into like classes or property values. The component percent compositions are summed for each group of rating class or soil property value. The rating class or soil property with the largest percent composition is used to class the map unit. Where two or more interpretation groups have equal percent composition, the group with the most restrictive interpretation is used.
3. Most Limiting – Out of all the selected components in the map unit, the component with the most limiting restriction for the interpretation is used.
4. Least Limiting – From all the selected components in a map unit, the component with the least limiting restriction for the interpretation is used.
5. Weighted Average – Weighted average is a method of processing numeric soil properties or productivity values (i.e., crop yields) for multiple components and arriving at a single value for the map unit. Weighted average means that each components value contributes to the final answer, based (weighted) on the component’s percent composition of the map unit. Components with larger percent composition contribute more to the final answer than minor components of small percent composition. For each selected component, the numeric property or productivity value is multiplied by the decimal percent of the component percent composition. These values are summed for all selected components and represent the weighted average value for the map unit.
6. Presence/Absence – All the components in a map unit are evaluated for the presence or absence of an interpretation or soil property. The map unit can be classed where the interpretation is present in all components, absent in all components, partially present in

all components, or unknown. This processing method is used for the hydric soil map unit rating.

7. All Components – All the components in a map unit are evaluated for the soil property or feature. This component processing method is more commonly used for physical or chemical soil properties, flooding, ponding, or water table.

Needless to say, performing one of these aggregations can be complicated. In many cases, some programming is required.

In NRCS, our GIS of choice is either ArcView or ArcGIS. Our on-site GIS expert tells me that attempting to do one of these aggregations using only the capabilities provided by the ESRI GIS user interface, is extremely difficult. He suggests that a better alternative is to perform the aggregation in the SSURGO template database, saving the results as a column in the mapunit table. Even doing one of these aggregations in a SSURGO template database is far from trivial.

The NRCS interim solution to this problem was to create an application that does the most commonly needed aggregations for us, “on the fly”. This application is known as “Soil Data Viewer”. Soil Data Viewer will be discussed further in the next section. But before that, we should probably mention the “muaggatt” table.

“muaggatt” is the physical name of the “map unit aggregated attribute” table. This table contains a sampling of interpretations and soil properties that have already been aggregated to the map unit level. The column for a particular interpretation aggregation won’t be populated unless the corresponding interpretation was included in the exported data. This table has a one to one relationship with the map unit table, on “mukey”. A description of the columns in this table is found in the SSURGO metadata report titled “SSURGO Metadata – Table Column Descriptions”.

Soil Data Viewer Application

Here is the description of this application from its home web page:

<http://www.itc.nrcs.usda.gov/soildataviewer/>

Soil Data Viewer is a tool built as an extension to ArcView that allows you to easily create soil-based thematic maps. Soil Data Viewer shields users and applications from the complexity of the soil database and incorporates rules for appropriate use of soil data. It provides an easy to use tool for geospatial analysis of soil information for resource assessment and management.

There are currently two version of Soil Data Viewer:

1. Soil Data Viewer 4.0 is installed as a component of Customer Service Toolkit 4.1. It is designed for NRCS field offices running CCE computers on Windows XP.
2. Soil Data Viewer 3.0 Stand-alone is designed for installation on non-CCE computers of other agencies, partners, or public.

Both of these versions can be downloaded from the Soil Data Viewer web site.

In order to use stand-alone Soil Data Viewer with its GIS components, you must already have the following software installed.

1. Microsoft Windows NT 4.0, Windows 95, Windows 98 or Windows 2000.
2. Microsoft Word 97 or Word 2000.
3. Microsoft Access 97, 2000 or 2002.
4. ESRI ArcView GIS 3.1 or greater (required for GIS data) installed in the default location.

A Windows XP version of Soil Data Viewer is currently under construction.

If you cannot use a version of the Soil Data Viewer application, your only option at the current time is to do your own aggregations from scratch.

Contacting Support

Questions should be directed to the National Soil Information System (NASIS) Hotline. The NASIS Hotline, which resides at the USDA NRCS National Soil Survey Center in Lincoln Nebraska, is staffed from 8:00 AM to 4:30 PM Central Time.

(402) 437-5378 – Steve Speidel
(402) 437-5379 – Tammy Cheever

e-mail: hotline@nssc.nrcs.usda.gov