Soil Data Warehouse – SSURGO Business Process

Current Process

The following is a brief description of the steps (major), process and validations for the current SSURGO process for the different business units.

Digitizing Unit

1. Gather Soil Business Data
    - Compilation Certification Letter.
    - Correlation Document/amendments (if no correlation then mapunit Legend, publication survey report).
    - Metadata (source data, compilation information).
    - Source data to digitize/existing digital data.
2. Digitizing Soil Map
    - Quad and full survey basis.
    - Get adjoining survey valid overlap/gaps – correct.
        a. SSURGO archive for adjoining surveys or new edited survey data.
        b. Coordinate with state on spatial data.
        c. Coordinate with other DU.
    - Create acre report of digital data, send to State and MO.
3. Quality control and assurance
    - Author error corrections (unlabeled polygons, common lines between polygons) - corrections made. (State office).
    - Check plots 100 percent – correct errors.
    - Send 10 percent check plots to MO – corrections made (repeat).
    - Create/write DLG.
    - Get tabular data (NASIS export file from state office).
    - Run AML validation (topology, header info, coordinate values, collars, data consistency, check tabular tables all there, certification status attribute) – coordinate with state on tabular data corrections (state office). Join spatial and tabular.
4. SSURGO Product Creation
    - Final certification letter.
    - Generate packages for archive – DLG NAD27 and/or 83, NAD83 UTM Coverage in quad basis (7.5 or 3.75), full survey coverage in geographic coordinates, Arc interchange exports of coverage (quad and full survey), shapefile full survey in geographic. All contain polygon, point, line, quad boundary, and survey boundary. Metadata and tabular data in separate directory and zipped with shapefile.
    - FTP to NCGC to a private directory at NCGC.

MLRA Office

1. Certification process
    - During 10 percent checks coordinate certification letter with State.

State Office

1. Gets acre report from DU, possible update tabular database.
2. Get special feature summary and update 37A.
3. Sends Metadata to DU.
4. Sends certified NASIS export to DU (or notifies DU that export is on NASIS server).
5. Some get the full survey cover and load into SDV for QA.
6. Signs SSURGO Certification Letter.
7. Some get data from FTP site and create field office datasets (spatial and tabular).

NCGC

1. Automatic notification that new data has arrived.
2. Run validation checks (topology, header info, coordinate values, collars, data consistency, check tabular tables all there, certification status attribute, metadata).
3. Errors are returned to DU (spatial or metadata errors) – correction made.
4. Post packages to SSURGO FTP site.
5. (Fred loads into SDE).
6. Survey boundary is loaded in SDE for a status map for only SSURGO I and II (ArcIMS).
7. Resource Data Gateway footprint created (shapefile).
8. Update SSURGO status activity map (monthly). Boundaries are not updated with survey digital boundary. (Status of soil survey map (year), Status of soil survey digitizing SSURGO (monthly).
9. Metadata (txt) converted into SGML and HTML placed on FGDC clearinghouse node (within 48 hours).
10. Available SSURGO list (HTML page) is updated.
11. Data stored off-site (DLT and CD-Rom).

New Business Process (spatial and tabular)

The following is a brief description of the proposed new steps, processes, and validations for spatial and tabular data delivery when the Soil Data Warehouse and Soil Data Mart are implemented. This results in a change in our current SSURGO business process.

Tabular Data

1. State Soil Scientist is responsible to exports tabular data for a soil survey to staging server. Authenticated user runs NASIS export to staging server (authentication is limited to state group and using state code from area symbol or overlap tables).
2. Authenticated user performs staging server activities.
    - Join tabular survey area symbol is joined to boundary map area symbol (SSURGO status map).
        a) The area symbol for the tabular data must match the area symbol for the spatial data.
        b) Every mapunit symbol in the spatial data except "BLANK" must match a mapunit record.
    - If area symbol is not found in boundary map then:
        a) Tabular data fails to commit to the warehouse.
        b) If tabular data is a national non-MLRA survey area symbol or MLRA soil survey area, then boundary map needs updating.
        c) NCGC in coordination with state updates boundary.
    - If tabular data is new to warehouse the Metadata template is updated with contact info based on area symbol, bounding area from survey area boundary, etc. If tabular data is already in warehouse the warehouse version of metadata is updated with this version of tabular data.
3. Validations

    Validate tabular data (some may be on transactional (NASIS) or staging server, warning or required-fail error).
    - **Required attributes to be populated**: (export fails if null).
        a. Area_type_name
        b. Area_symbol
        c. Area_name
        d. Area_acres
        e. Project_scale from the legend table
        f. Mapunit_symbol
        g. Mapunit_name
    - **Data Validations:**
        a. Area symbol may not contain only letters and digits, no other special characters allowed. Area symbol will convert to upper case characters.
        b. At least one valid entry in overlap table for at least one state and valid county (for that state) excluding Pangaea site CZ, FN, PB, UM (export fails). The state overlap is used to validate authorized users to commit survey to warehouse from staging server.

c. Only Pangaea official "Non-MLRA Soil Survey Area" or "MLRA Soil Survey Area" area type. (export fails)
d. Unique mapunit symbol within a legend. All musym in spatial will match a mapunit symbol in tabular excluding "BLANK." (export fails)
e. Not more than one hydric rating for a component. (export fails)
f. No null component months. (export fails)
g. A mapunit text record for Nontechnical Description may not have a null text category. (export fails)
h. Component_name is required if a row is open in component table. (export fails)
i. Correlation_date in the legend table is not required. Data values are stored in SDW and SDM.
j. Warning if mapunit_acres are null. Generate list of mapunits that have null acres.

- **Data Processed (SDW and SDM)**
  a. Soil_survey_area_status is not required. Value is stored in SDW, but value is nulled in SDM. (future SSURGO structure will drop this attribute)
  b. Mapunit_status is not required. Value is stored in SDW, but value is nulled in SDM. (future SSURGO structure will drop this attribute)
  c. Legend_suitability_for_use is not required. However, a warning if data value is other than "2 – current wherever mapped". Value is stored in SDW, but value is nulled in SDM. (future SSURGO structure will drop this attribute)
  d. Legend_certification_status is not required. However, a warning if data value is other than "3 certified". Value is stored in SDW, but value is nulled in SDM. (future SSURGO structure will drop this attribute)
  e. DMU_certification_status is not required. However, warn if data value is other than "3 certified". Value is stored in SDW, but value is nulled in SDM. (Future SSURGO structure will drop this attribute).

  Steps 1 and 2 will be done in NASIS before export through either a calculation or a report. The report would put the information together and the NASIS user would copy and paste it into legend_export_metadata.

  1. If value is "Not for Distribution", "Not Certified" or "Partly Certified" then create list of - map unit symbol, DMU certification status and DMU text where Kind = "certification statements".
  2. User is given option to include list (paste) into Legend_export_metadata text record in dataset on staging server (append to any existing text).

f. Export_certification_status is not required. However, a warning if data value is other than "3 fully certified". Values are stored in SDW, but only "2 partly certified" or "3 fully certified" are stored in SDM, all others are nulled.
    1. Warn if export certification status is "Partly Certified" and export_metadata is null. (export metadata text may be used to insert into metadata file)
g. Export_certification_date is populated electronically when the dataset is exported from NASIS transactional database. Further work needed on definition of export certification date. Current thinking is that this date should represent the actual date the data is exported from NASIS (not the date imported into warehouse). Other system version dates in the warehouse and data mart to be used as dates to identify the current data as well as the date range when a dataset was current.

Spatial Data

1. Gather Soil Business Data (Digitizing Unit)
    - Compilation Certification Letter.
    - Correlation Document/amendments (if no correlation document then use mapunit Legend or published survey report).
    - Metadata (source data, compilation information).
    - Source data to digitize/existing digital data.
2. Digitizing Soil Map
    - Full survey basis (no quads).
    - Get adjoining survey valid overlap/gaps – correct
        a. DU - Using survey boundary map to identify adjoining soil surveys.
        b. DU - Adjoining soil surveys extracted from the Warehouse/Soil Data Mart (UTM). Boundaries are corrected.
        c. Coordinate with state on spatial data overlap and gaps.
        d. Coordinate with other DU on surveys crossing DU regions.
3. DU - Quality control and assurance
    - Author error corrections (unlabeled polygons, common lines between polygons) - corrections made. (State office)
    - DU – Completes quality checks 100 percent.
    - DU – sends 10 percent check plots to MO.
4. DU - creates and sends acre summary and special features report to State/MO and request for any remaining business area paper work.
5. SO/MO - updates mapunit acres, validates that spatial mapunits exist in current legend, send preliminary NASIS data, amended correlation documents, 37A, etc.
6. DU - validates spatial and preliminary tabular data.
7. DU – adds area symbol, mukey to each feature class.
8. DU - send new spatial data in ESRI Coverage format (no DLG or shapefile - these formats are delivered dynamically from the Soil Data Mart) with metadata, and

any corrected adjoining spatial data to Staging Server. Metadata is sent in XML format. NCGC is notified that new metadata is on staging server.

9. Staging server validates spatial boundary with adjoining boundaries in warehouse.
10. SO selects and exports final tabular data to staging server.
11. Authenticated user logs onto Staging Server performs these task:
    - Join (spatial and tabular)
        a. Failed joined then SO and DU may need to correct. Resubmit data.
    - Spatial Validations
        a. Musym and mukey in spatial data must match musym and mukey in tabular data.
        b. Musym can not be null.
        c. Mukey can be null, only for musym "BLANK".
        d. Special feature featsym must match spatial and tabular (feature description)
        e. Check for unlabeled special feature
        f. Area symbol in spatial data must match area symbol in tabular data.
    - metadata reviewed/edit/update
        a. DU has edit capability to update metadata for spatial editing.
        b. SO updates metadata for tabular changes (from NASIS or use edit capability on the staging server).
        c. NCGC reviews and approves. This is verbal coordination of State Soil Scientist and NCGC before commit is done.
    - Quality review (reporting capability or optional by downloading and using SDV).

12. State Soil Scientist commits survey data to warehouse.
    - State Soil Scientist ensures that Certification Letter is signed (the two certification letters – Soil Survey Geographic Data Certification and FOTG certification will be replaced by one letter yet to be developed. The Regional Technology Specialists are working with management to review, revise GM and policy for FOTG certification issues.
    - On commit the system:
        a. Checks that the spatial data boundary still matches adjoining surveys. Soil survey area boundary is validated with existing warehouse boundaries to **3** (to 6 ?) meter resolution. (Commit fails)
        b. Spatial version tag is added to spatial data (polygon, line and point).

Data Mart

1. System sends metadata to FDGC clearinghouse (spatial/tabular only?) Need to refine rules on which type of data is published to clearinghouse.
2. Possible retrieve existing data from FTW ftp site for data not yet loaded into data mart. (Michelle Simmons FTW contact).

3. Downloading spatial data - add advisory that the data may have been compiled and developed on different base and that when viewed on different base the lines may not line up (for more detail see Metadata).