

Version

2.2

FARM SERVICE AGENCY



ArcGIS Data Loader

User Guide

GIS TRAINING MATERIAL

ArcGIS Data Loader User Guide

Farm Service Agency, USDA

Please forward comments and suggestions about this document to:
Jim Heald, FSA HQ, Washington DC, jim_heald@wdc.usda.gov
Phone (202) 720-0787

Using the ArcGIS Data Loader

Note: This tool is installed when the ArcGIS CLU Maintenance Tool is installed.

Prior to running the ArcGIS Data Loader, it is highly recommended to create a spatial index on the CLU shapefile. A spatial index will greatly improve performance.

To create a spatial index index within ArcView 3.X:

- 1 Load the CLU shapefile into the ArcView CLU Maintenance Tool
- 2 Make the CLU shapefile the active theme in the View
- 3 Click the Open Theme Table button
- 4 Make the Shape field active by clicking it
- 5 Choose Create Index from the Field menu (if Remove Index is present in this Menu, an index has already been created and no action is required)
- 6 Exit the the ArcView CLU Maintenance Tool

Note: It is also highly recommended to run the *ArcGIS.3.x Quality Control tools* on your data prior to data load. This will catch any outstanding errors within your CLU shapefile that will stop your data from being imported into ArcSDE.

To run the ArcGIS Data Loader click on “**Start**”; click “**Programs**”; click “**USDA Applications**”; click “**Common Land Unit**”; and then click “**ArcGIS Data Loader**”.



The **ArcGIS Data Loader** is a Tool that converts the CLU data (CLU shapefile, CRP table, and Wetland shapefile) to ArcSDE database format. It is an executable file that must be run every time you want to add CLU data into ArcSDE. It must be run at least **once per Service Center** prior to working with CLU data in the ArcGIS CLU Maintenance Tool.

The ArcGIS Data Loader will automatically look for existing counties based on CLU shapefiles, CRP tables and Wetland shapefiles in the following directory location: F:\geodata\common_land_unit\fsa_clu. The ArcGIS Data Loader will confirm the existence of these files and begin the process to convert them into ArcSDE viewable database files.

ArcGIS Data Load Process

For the ArcGIS Data Loader to work correctly the CLU shapefile must be stored in the proper directory location (F:\geodata\common_land_unit\fsa_clu) and must be named correctly, that is, the names of the files must abide by USDA FSA naming conventions for geospatial data. If the CLU shapefile has an associated CRP table and/or Wetland shapefile, those files will also need to be stored within the same directory and named appropriately:

- clu_a_<stnnn>.shp -For CLU shapefile data
- crp_t_<stnnn>.dbf -For CRP tables
- wet_p_<stnnn>.shp -For Wetland shapefile data

where <st> is the 2 digit State abbreviation and <nnn> is the 3 digit FSA County code.

The following is a list of QC checks performed on the CLU, Wetland and CRP data prior to data load that would cause the load to fail.

- Attribute Structure (All data types)
- Invalid or Multiple CLUIDs (CLU data only)
- Multi-Part Feature (CLU data only)
- Overlapping Features (CLU data only)
- Sliver Features (CLU data only)

The following is a list of QC checks performed on the CLU, Wetland and CRP data that would allow the data to load and would be reported.

- Invalid acreage values (CLU data only)
- Multiple Tracts within Farms (CLU data only)
- Unique CLUs within Tracts (CLU data only)
- Duplicate and orphan CRP records (CRP data only)
- Invalid date values (CRP and Wetland data)

The ArcGIS Data Loader recalculates all acreages within the new CLU layer created in the ArcSDE database. These values will be compared to the acreage values in the CLU shapefile and report any discrepancies.

The ArcGIS Data Loader also deletes all duplicate and orphan CRP records within the ArcSDE database.

All Wetland Layer and CRP Table invalid dates will be discarded within the ArcSDE database. Valid dates are comprised of a Month, Day and Year, not a subset of these elements.

After confirming the existence of these files, the ArcGIS Data Loader will run numerous QC checks on the data to ensure the data contains no major topological or attribution errors. Upon completion of these checks, the ArcGIS Data Loader will place the CLU data into a new CLU database structure for use within ArcGIS CLU Maintenance Tool and a report will be generated identifying which counties have successfully loaded or which ones failed. If it finds errors within the data, the report will identify what caused the data load to fail. See 'Fixing Special Errors' at the back of this document to see instructions to fix 2 common errors that occur with the data loader.

To fix the errors, run all the QC tools found within the ArcView 3.x CLU Maintenance Tool. If the data is successfully loaded within ArcSDE, the original CLU shapefile data can then be archived, as this CLU data will no longer be used with the ArcGIS CLU Maintenance Tool.

If there are no CRP tables or Wetland shapefiles associated with a CLU shapefile prior to running the ArcGIS Data Loader (i.e. either do not exist or are not stored within the same directory as the CLU shapefile), the ArcGIS Data Loader will automatically create and name an empty feature class and/or table within the ArcSDE database.

If, during the data load process, the ArcGIS Data Loader encounters any problems such as incorrectly stored or named files, the data load process will be terminated and a message will appear informing the user why the process was halted.

Upon completion of the data load process a report will be issued for the user to examine.

Operating the ArcGIS Data Loader

Steps:

1. Make sure your CLU, CRP and Wetlands data are stored at the following location (F:\geodata\common_land_unit\fsa_clu) and properly named.
2. To run the ArcGIS Data Loader click on "**Start**"; click "**Programs**"; click "**USDA Applications**"; click "**Common Land Unit**"; and then click "**ArcGIS Data Loader**".
3. A login screen will appear: Type in the **Password** for the SDE account and press "**OK**."



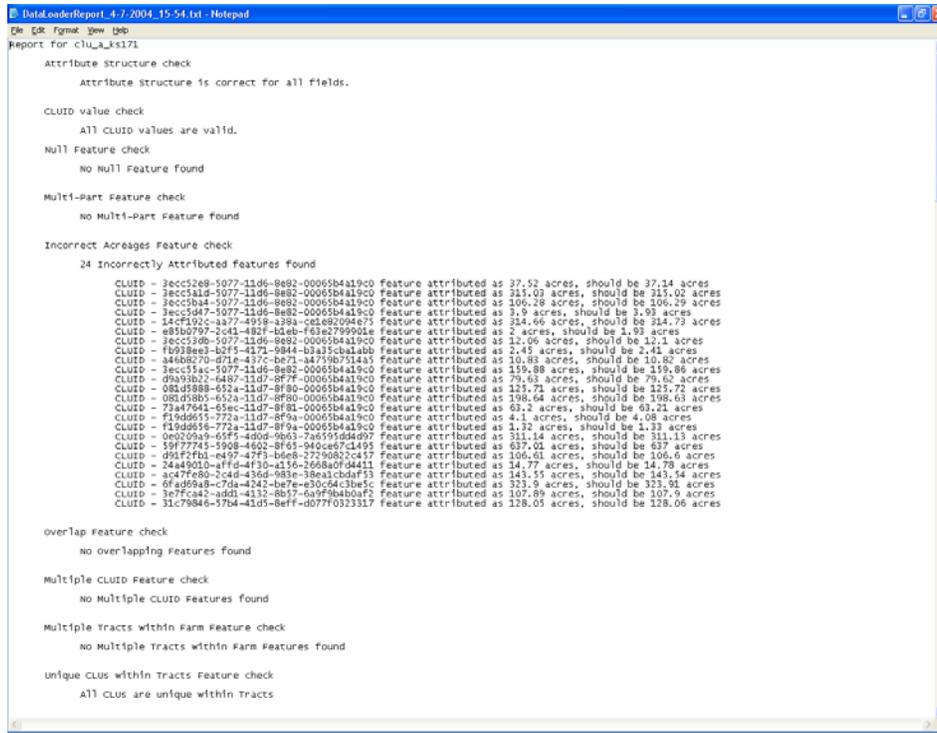
4. After the Password has been verified, the data will be processed and several messages will appear in the status bar.

After processing the data the following message will appear:



It will identify how many counties either succeeded or failed during the load process. Press the **“Yes”** button to view the report, press the **“No”** button to exit the Tool. If failures are encountered one of three situations has occurred. The county has already been loaded into the CLU database or errors were encountered during the Quality Assurance checks performed by the data loader. The third possibility is the UTM zone the county data resides in has not been entered in the UTM cross reference table. If this occurs please contact your state GIS coordinator or your state IT staff.

The Report will look similar to this:



If you desire to review the ArcGIS Data Loader’s report at a later date, the reports are stored in the following location:

C:\Program Files\USDA\FSA\ArcGIS\Maintenance\logs

Note: For offices that administer more than one county’s dataset, the ArcGIS Data Loader will process multiple county CLU, CRP and Wetland data if they are located in the same directory structure together. There is no need to run this process for each individual county because the ArcGIS Data Loader will process them all at once. CLUs without associated CRP or Wetland data will have those files created for them in the database(s) as empty tables.

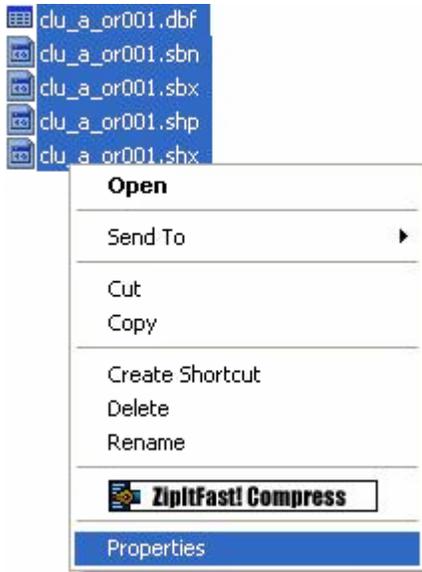
You are now ready to view and edit the data using the ArcGIS CLU Maintenance Tool.

Note: Make sure the clu shapefiles, wetland shapefiles, and crp tables are set to Read-Only after completing the data load process.

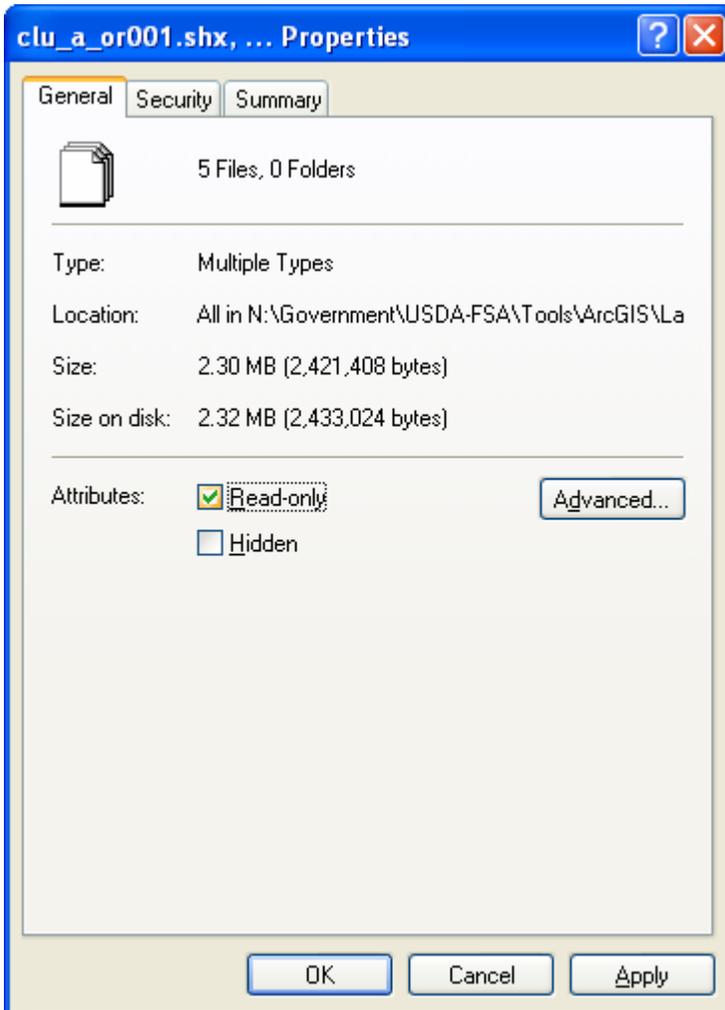
In windows explorer or (My Computer), navigate to the directory where your CLU, Wetland, and CRP files are stored:

Example: “F:\geodata\common_land_unit\fsa_clu”

Select all CLU/Wetland/CRP files related to counties that have been loaded by the ArcGIS Data Loader.



Right click on one of the selected files and choose “Properties”



Check the box next to “Read-only” and click on “OK”.

Understanding the Data Loader Report

In the process of importing the shape file into SDE, what does the data loader check for and under what conditions does the data loader abort the import of the shape file?

1. **The CLU Attribute Structure Check:** This procedure looks at the CLU shape file to determine if the proper fields exist in the shape file. Only valid fields will be brought into SDE. If there any extra fields beyond what is given in the following table, they will be skipped and not loaded into SDE. If there are any fields missing from the authoritative shape file (see below), the tool continues with validation checks, the missing fields are listed on the report in this section, but the file is not loaded into SDE. Field type and width are also checked.

CLU Shape File Fields

Alias	Type	Precision	Width	Indexed?
SHAPE	FIELD_SHAPEPOLY	0	8	TRUE
COMMENTS	FIELD_CHAR	0	80	FALSE
STATECD	FIELD_CHAR	0	2	FALSE
COUNTYCD	FIELD_CHAR	0	3	FALSE
FARMNBR	FIELD_DECIMAL	0	7	FALSE
TRACTNBR	FIELD_DECIMAL	0	7	FALSE
CLUNBR	FIELD_DECIMAL	0	7	FALSE
CALCACRES	FIELD_DECIMAL	2	8	FALSE
HELTYPCD	FIELD_CHAR	0	1	FALSE
CLUCLSCD	FIELD_DECIMAL	0	2	FALSE
FSA_ACRES	FIELD_DECIMAL	2	8	FALSE
CLUID	FIELD_CHAR	0	36	FALSE

2. **The CLU Value Check:** The CLU value is a Microsoft generated number when the CLU polygon is created which is called a GUID. More specifically a GUID is short for *Globally Unique Identifier*, a unique 128-bit number that is produced by the Windows OS or by some Windows applications to identify a particular component (the CLU polygon). Windows also identifies user accounts by a username (computer/domain and username) and assigns it a GUID. Some database administrators even will use GUIDs as a primary key value in databases. The GUID for FSA is either a 32 character string or a 36 character string.

The check will evaluate all CLUID values to see if the value is either 32 or 36 characters long. If the data loader finds anything else, the data loader will continue

to perform the other checks but will not import the shape file into SDE. The report will identify what CLUIDs do fit the length convention. The GIS user should go back to the 3.3 QC tools and perform the CLUID check which will take care of this issue.



The Check for Duplicate/Invalid CLUID's Tool on the QC CLU Maintenance Tool Bar in ArcView 3.3

3. **The CLU Null Feature Check:** This checks to see if there are null polygons within the CLU shape file. A null polygon is a feature for which there is no associated record in the table. A null polygon can also happen when a record has no associated feature or polygon.

If a null polygon is found the tool continues with validation checks, the number of the CLU polygon with this error will be listed on the report under "Null Feature Check", but the file will not be converted. The GIS user needs to use the "Clean Tool" in the ArcView 3.3 environment. This will delete the record that has no associate feature or the feature that has no associated record.



The Clean Tool on the QC CLU Certification Tool Bar in ArcView 3.3

4. **The CLU Multi-Part Feature (Polygon) Check:** A multi part polygon is a polygon (feature) that is composed of more than one physical part but only references one set of attributes in the database. If a multi-part polygon is found the tool continues with validation checks, the number of CLU with this error are listed on the report under "Multi-part Feature Check", but the file will not be loaded into SDE. Go back to the ArcView 3.3 QC tool that checks for multi-part polygons. If only a few exists, use the CLUID numbers that are given in the report to resolve the problem(s). The multi-part correction efforts require manual intervention. The ArcView tool takes you to where the problem exists but the user must fix the problem.



The Show Multi-Part Polygons Tool on the QC CLU Certification Tool Bar in ArcView 3.3

5. **CLU Incorrect Acreages Feature Checks:** A problem arises when the actual Calc_Acres figure does not accurately reflect the system generated acres that the CLU polygon "should" reflect. How does this happen? When a user invokes changes through vertices editing on CLU polygon or a person going into the CLU dbf file and directly changing the value in the Calc_Acres field could lead to this problem. When the shape file is exported into the SDE environment, this is automatically corrected without question to provide the most accurate information on the each and every CLU. The way to avoid having this problem is to run the

“AC” button (the “Update System Acreage” button) in the 3.3 environment as a way to update this field.



The Update System Acreage Tool on the QC CLU Maintenance Tool Bar in ArcView 3.3

6. **CLU Overlap Feature Check:** Polygon overlaps occur if one or more edges of a polygon overlap another polygon. If overlaps are found, the tool continues with validation checks, the numbers of CLU’s that created the overlap are listed on the report in this section and the file will not be loaded into SDE. The user can use this report as a way to find the overlaps or the overlaps can be found using the “Check for Overlaps” button in ArcView 3.3. The ArcView tool takes you to where the problem exists but the user must fix the problem.



The Check for Overlaps Tool on the QC CLU Maintenance Tool Bar in ArcView 3.3

7. **CLU Multiple CLUID Feature Check:** This checks to see if there are duplicate CLUID numbers within the CLU shape file. If there are duplicate CLUID’s, they are reported but the file will not be converted. Run the “Check for Duplicate/Invalid CLUID’s.”



The Check for Duplicate/Invalid CLUID’s Tool on the QC CLU Maintenance Tool Bar in ArcView 3.3

8. **Multiple Tracts within Farm Feature Check:** This checks to see if a tract number occurs within more than one farm within the CLU shape file. If a tract number is found on more than one farm, the tool continues with validations and the tracts with this are listed on the report along with the corresponding farm numbers in this section. This can be avoided before running the data loader tool by running the “Tract/Farm Ratio” button in the 3.3 environment.



The Tract/Farm Ratio Tool on the QC CLU Maintenance Tool Bar in ArcView 3.3

9. **Unique CLU’s within Tracts Feature Check:** This checks to see if a CLU number occurs more than once within a given tract. Currently, this test excludes the number zero and allows for multiple zeros. If a CLU number is found to be duplicated within a same tract, the tool continues with validations and the CLU’s with this are listed on the report along with the corresponding tract



The Tract/CLU Number Ratio Tool on the QC CLU Maintenance Tool Bar in ArcView 3.3

10. **CLU Sliver Feature Check:** This validation checks to see if there are sliver polygons that are less than 0.005 acres in size within the CLU shape file. If a sliver polygon less than the prescribed size mentioned before, the tool continues with validation checks, the CLU’s with this error are listed on the report in this section, but the file is not loaded into SDE. The “Sliver Polygon’s” tool can correct these problems by taking to the areas of problems.



The Sliver Polygon’s Tool on the QC CLU Maintenance Tool Bar in ArcView 3.3

11. **The CLU Acreage Value Review Check:** This check reports on any of the CLU polygons that have had their system calculate acreage changed due to the SDE import process. When importing from a shape file, that has calculated acres reported only to the hundredths, into SDE, that will assign a precision reported to 8 decimal places to the right and has a spatial index from which may cause the slightest vertices shift, the result will be a change in one-one hundredths of an acre for a CLU polygon. Some number farther to the right of the decimal point has changed due to the spatial index moving the vertices ever so slightly so that the one-hundredths number is rounded up or down.
12. **CRP Attribute Structure Check:** Like the CLU Attribute Structure Check, this procedure checks the CRP table to determine if the proper fields exist in the file. Only valid fields will be brought into SDE. If there any extra fields beyond what is given in the following table, they will be skipped and not loaded into SDE. If there are any fields missing from the authoritative shape file (see below), the tool continues with validation checks, the missing fields are listed on the report in this section, but the file is not loaded into SDE. Field type and width are also checked.

CRP Table Attribute Structure

Alias	Type	Precision	Width	Indexed?
CLUID	FIELD_CHAR	0	36	false
Statedcd	FIELD_CHAR	0	2	false
Countycd	FIELD_CHAR	0	3	false
Contract	FIELD_CHAR	0	12	false
Acres	FIELD_DECIMAL	2	8	false
Expdate	FIELD_CHAR	0	8	false
Prac_nbr	FIELD_CHAR	0	5	false

13. **CRP Multiple CLUID Record Check:** This check tests for duplicate CLUID numbers within the CRP shape file. If there are duplicate CLUID’s, the records are deleted and reported. This will not abort the SDE load process but it is important to research the deleted records.

14. **CRP Orphaned Record Check:** This check identifies all the CLUID’s residing in the CRP table that do not have a match in the CLU shape file. This can happen by editing the CLU shape file and not having those changes reflected in the CRP table. The data loader procedure will not abort the SDE load but will delete the records and identify the CLUID that was deleted in the report.

It would be important to join the CLU shape file to the CRP table to find the orphans before the SDE load process. This way, you can straighten out the problem before the alteration to the file happens during the SDE load process.

15. **CRP Invalid Dates for the EXPDATE Field:** This process evaluates the field for a good date. If the data loader determines through a series of tests that the attribute is an invalid date, the expiration date is deleted in the data load process. This test does not cause the load process to abort.

Do not waste time prior to running the data loader to clean this. In the ArcGIS environment, use the data loader report to find the CLUID and the Contract number to re-enter the expiration date.

16. **The Wetland Attribute Structure Check:** Like the CLU Attribute Structure Check, this procedure checks the Wetland shape file to determine if the proper fields exist in the shape file. Only valid fields will be brought into SDE. If there are any extra fields beyond what is given in the following table, they will be skipped and not loaded into SDE. If there are any fields missing from the authoritative shape file (see below), the tool continues with validation checks, the missing fields are listed on the report in this section, but the file is not loaded into SDE. Field type and width are also checked.

Wetland Shape File

Alias	Type	Precision	Width	Indexed?
Shape	FIELD_SHAPEPOINT	0	6	False
Certdate	FIELD_CHAR	0	8	False
Wetlabel	FIELD_CHAR	0	100	False
Wetacres	FIELD_DECIMAL	2	8	False
Status	FIELD_CHAR	0	16	False
Statecd	FIELD_CHAR	0	2	False
Countycd	FIELD_CHAR	0	3	False

17. **Wetland Invalid Dates for the CERTDATE Field:** This process evaluates the field for a good date. If the data loader determines through a series of tests that the attribute is an invalid date, the certification date is deleted in the data load process. This test does not cause the load process to abort. The total number of invalid dates is reported within the data loader report.

Fixing Special Errors That Show Up In the Data Loader for FSA ArcGIS

Scope of Document:

This document is prepared to help Service Centers and State Offices understand special errors that may show up in the Data Loader Report that is not explained in the Data Loader User Guide.

The Errors:

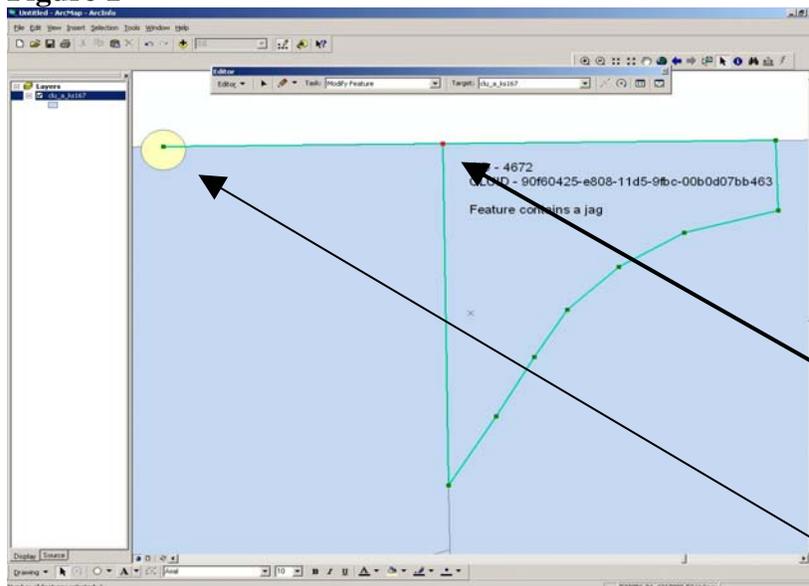
While preparing for the summer of 2004 Pilot Project, the GISC found 2 unique errors the Data Loader flags that the ArcView QC tools do not catch. Also, these errors are not something that shapefiles are should be concerned with because they can operate and perform the necessary business for FSA. ArcSDE, in contrast, does not like these errors and, thus, have been caught and need to be cleaned before converting the shapefile into an ArcSDE layer.

The 2 errors are:

1. Spatial errors with linestring or poly boundary is self-intersecting
2. The number of points is less than required for feature

These errors are found at the bottom of the reports and are very specific about the problem they are describing. **Linestrings** are situations where, at a corner of a polygon, two vertices are residing at the same x,y and a third vertex is situated to the side which creates 2 polylines occupying the same space (Figure 1 illustrates this problem).

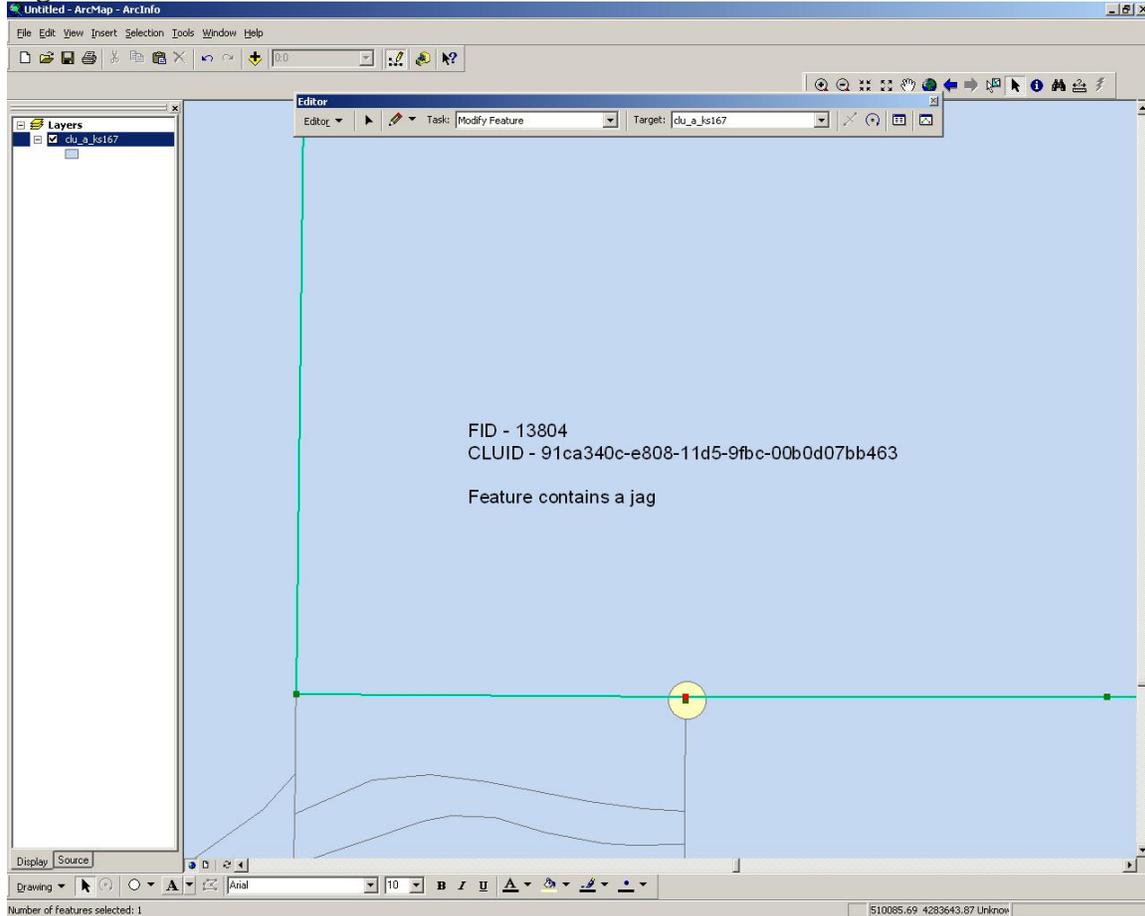
Figure 1



Shapefiles don't care if this situation exists. But in SDE, the database requires unique placement of all vertices and lines occupying the same space. The way to correct this problem is to delete one of the vertices sitting on top of each other. The user will also have to delete a second vertex which allows the polylines to occupy the same space.

Figure 1 shows a problem that will exist and is relatively easy to find for this type of error. In most instances, this error is very small and will require zooming in to the view. Figure 2 shows how these errors will generally show up.

Figure 2

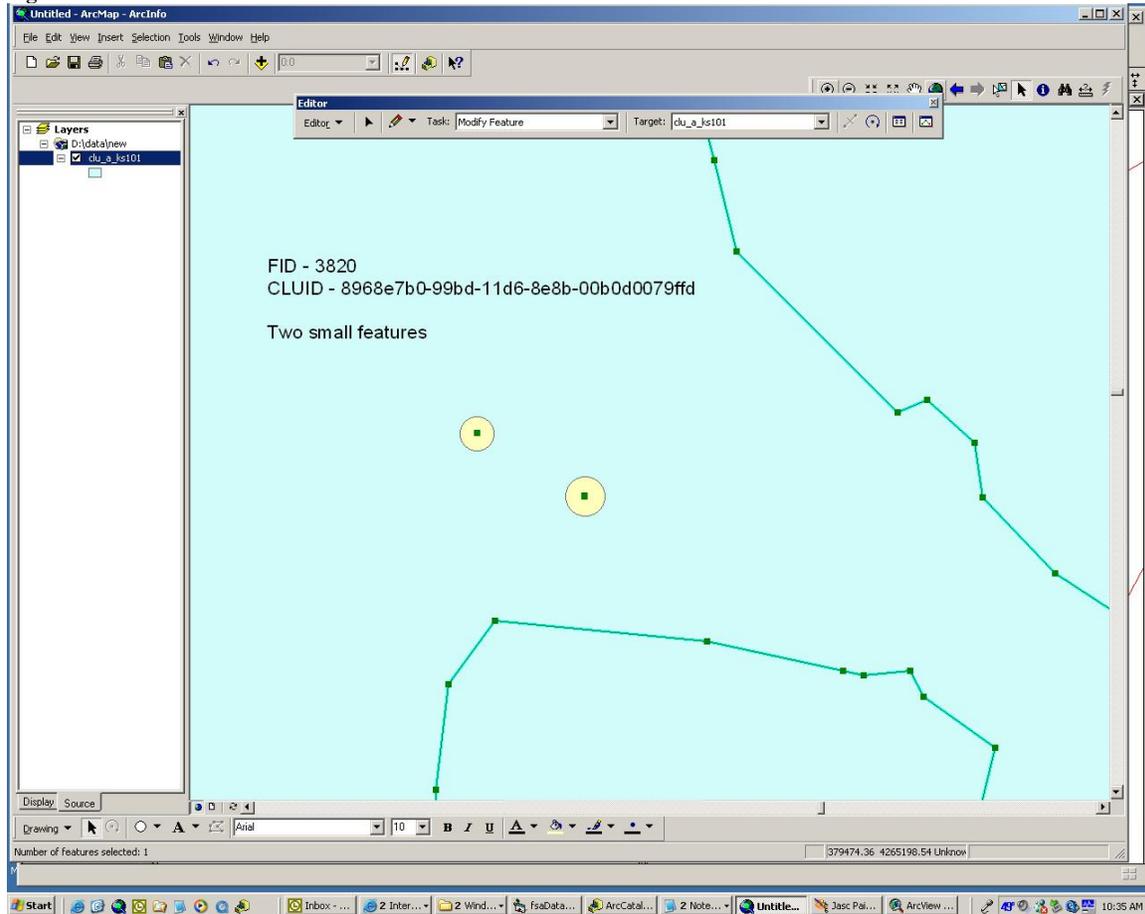


Fixing Linestrings: Although these figures are showing vertices editing in the ArcGIS environment, counties will have to perform the vertices edits in the ArcView world. It is advisable to have the digital copy of the Data Loader Report. A user can highlight the CLUID identified in the report and copy it over into the query builder window. The user will zoom to the selected polygon and start a vertice editing session (either from the ArcView Maintenance Tools or the ArcView out-of-the-box functionality). The user will have to pan around the edges of the polygon while zoomed in to catch these errors. It is always best to look at the polygon as a whole and look for concentrated areas of vertices. It is usually in those areas where the problem exists.

The **number of points is less than required for feature error** is a situation where a polygon only has 2 vertices and thus looks like a line (Figure 3). This is type of error happens when a digitizer or a CLU editor moves the mouse and accidentally left clicks very quickly. It's a simple error but finding them is difficult. They usually exist in the

middle of a polygon and will require a user to copy/paste the CLUID from the digital report to the query builder and then zooming to the selected polygon. Before zooming in further, start a vertice editing session. This will usually show extraneous vertices sitting in the middle of the polygon. Again, these errors are very small so be patient and look closely.

Figure 3



Fixing this error requires the user to place the cursor over the vertices and right click to delete the vertices. If the user is not zoomed in tight enough, it may look like nothing happened when this function is performed. In reality, one of the vertices has been deleted. Perform the function again and again until the entire object is gone.

This will take care of the offending errors and the data loader can be run again if all other problems have been fixed.