



GST 103: Data Acquisition & Management Lab Series

Lab 5: Spatial Data Quality

Document Version: **2013-07-31 (Beta)**

Organization: Del Mar College
Author: Richard Smith

Copyright © National Information Security, Geospatial Technologies Consortium (NISGTC)

The development of this document is funded by the Department of Labor (DOL) Trade Adjustment Assistance Community College and Career Training (TAACCCT) Grant No. TC-22525-11-60-A-48; The National Information Security, Geospatial Technologies Consortium (NISGTC) is an entity of Collin College of Texas, Bellevue College of Washington, Bunker Hill Community College of Massachusetts, Del Mar College of Texas, Moraine Valley Community College of Illinois, Rio Salado College of Arizona, and Salt Lake Community College of Utah. This work is licensed under the Creative Commons Attribution 3.0 Unported License. To view a copy of this license, visit <http://creativecommons.org/licenses/by/3.0/> or send a letter to Creative Commons, 444 Castro Street, Suite 900, Mountain View, California, 94041, USA.



The Center for Systems Security and Information Assurance (CSSIA), in partnership with the Network Development Group (NDG) is given a perpetual worldwide waiver to distribute per US Law this lab and future derivatives of these works.

Contents

Introduction	3
Objective: Spatial Data Accuracy	3
Lab Settings	3
1 Exploring Data Accuracy	4
2 Metadata	5
3 Data Aggregation	6
Conclusion	7
Discussion Questions	7

Introduction

This lab is part of a series of lab exercises designed through a grant initiative by the National Information, Security & Geospatial Technologies Consortium (NISGTC), funded by the United States Department of Labor in partnership with the Department of Education under the Trade Adjustment Assistance Community College and Career Training Grant Program (TAACCCT).

Your instructor may require that you provide screen captures and/or exported files. Please check with your instructor for the requirements specific to your class.

Spatial data is becoming more and more commonly available on the Internet. However, the accuracy of the data can be a concern. As we are experiencing a growth in data availability, we should choose our sources wisely. Data repeatability is key. When it comes to data accuracy, we must examine the metadata, as well as the spatial and attribute components. Metadata is becoming a significant data component and is a key factor in determining the completeness of data.

This lab includes the following tasks:

1. Exploring Data Accuracy
2. Metadata
3. Data Aggregation

Objective: Spatial Data Accuracy

This lab focuses on data and ensuring its accuracy. We will be looking at metadata and the standards of the metadata that some datasets require. We will also look at assessing the accuracy of data in order to decide if it is appropriate for use.


Lab Settings

Required Virtual Machines and Applications

Windows Machine User Account	Train
Windows Machine User Password	Train1ng\$

1 Exploring Data Accuracy

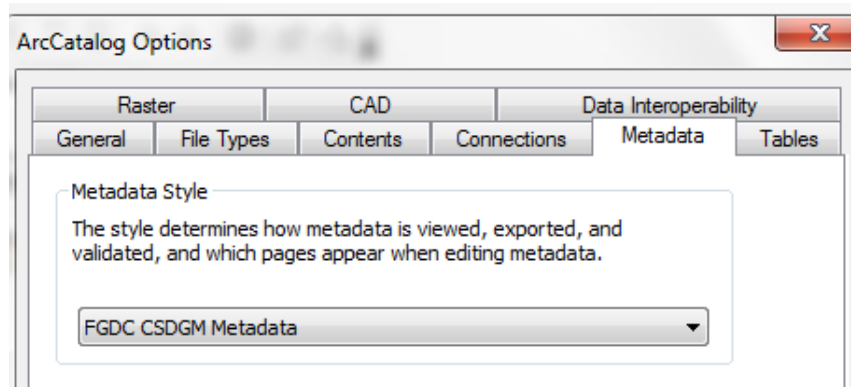
Data accuracy is an important concept for both spatial data as well as attribute data. Spatial and attribute data can be edited and changed, but what if there is a noted problem in the data? Can we overlook certain points in the data that we know have been captured erroneously? We must be aware of the errors that are inherent in the data and the fixes that are available.

1. Log into the computer, using the information provided in the Lab Settings section.
2. Click **Start->all Programs->ArcGIS->ArcMap 10.1**. Open to a blank map.
3. Connect to your *Lab 5* folder.
4. Add the **XYBuildings** shapefile and the **Campus** image to the map window from the *Lab 5* folder.
5. We have been given a shapefile with 5 building corners on campus, the data is not documented and there is an error. The error is noted in a text file found in the *Lab 5* folder.
6. Open the text file by browsing to it in Windows Explorer and double-clicking on XYBuildingsError.txt. The file should open in Notepad. Read the text file. There is an error in the data. However, this is an acceptable error because it has been noted and a correction has been given.
7. Open the attribute table for the XYBuildings shapefile, and compare the attribute data to the information given in the text file.
8. Select the **cp** point in the attribute table and **Zoom To the Selected Feature**. Place your cursor over the point and look at the XY coordinates in the bottom right-hand corner below the image. If listed in feet, you should notice that the values are not quite the same as those listed in the attribute table. In this particular instance, the attribute data is correct, but the spatial data is incorrect. In order to remedy this error, we will have to edit the vertices and move the point to the correct position.
9. Enable the **Editor** toolbar. Click on the **Editor** dropdown and click **Start Editing**.
10. Select the erroneous point (if not selected), right-click and select **Edit Vertices** (Editor toolbar).
11. Open the **Edit Sketch** property tab using the  button on the edit toolbar. In this tab, we can amend the coordinates given. Using the information from the text file, Enter the XYZ information and click **Finish Sketch**. Notice the point does not appear on the map.
12. The data about the error was erroneous. This is problematic. Can this point be trusted? Can this data point be trusted? Try to swap the X and the Y values from the text file around and see if the point is on the map.
13. When the X and Y values from the text file are swapped, the data point appears on the corner of a building. We should note this down so we can edit the metadata.
14. Save your edits.

2 Metadata

In this task, we will be looking at the metadata section of spatial data. When data is purchased, we expect to have a complete dataset. This includes spatial data, attribute data, and metadata. Metadata provides descriptive information that helps users to identify, interpret, and use the data. Metadata is a way for us to see what processing the data has undergone and how it has changed since it was captured. We can see the lineage and the geoprocessing history. Most importantly, we can see what is wrong with it. Not all datasets are free of errors. There may be a spelling mistake or an errant value that appears that will not represent actual data.

1. Open **ArcCatalog**.
2. Click on the **AirportPolygon** shapefile. Next, click on the **Description** tab and look at the data that has been entered into the fields. This is the item description. Go to **Customize->ArcCatalog Options**. Go to the **Metadata** tab and change the metadata style to **FGDC CSDGM Metadata**. Click **OK**.

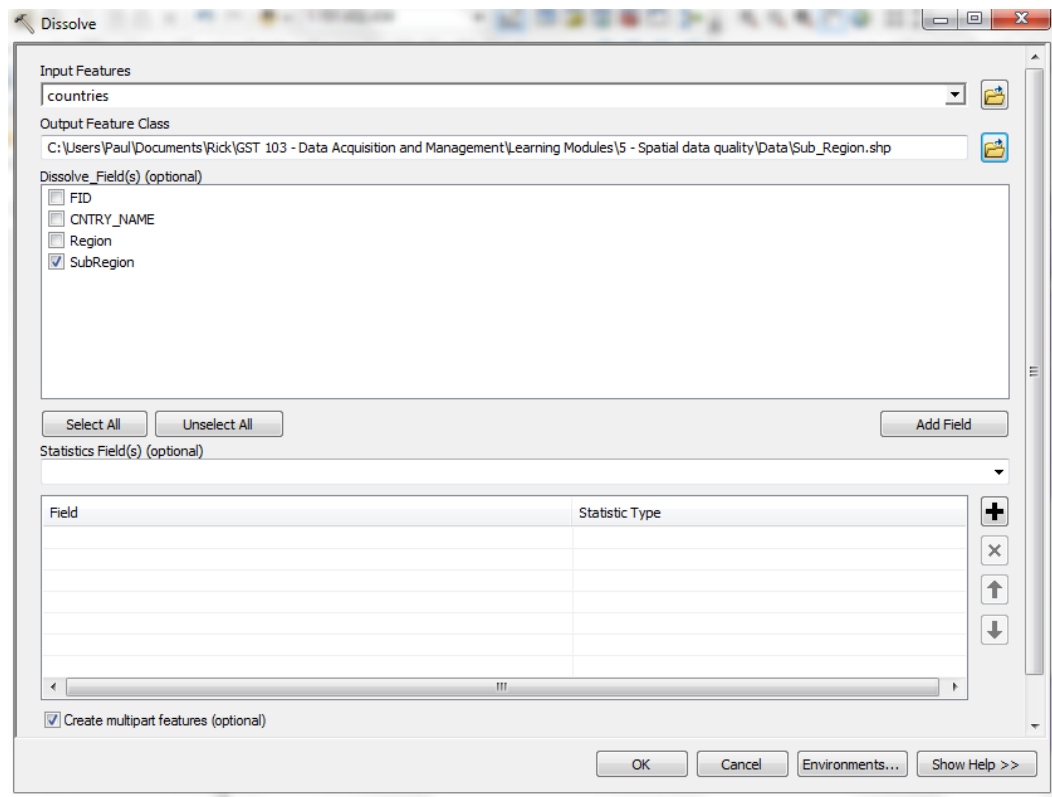


3. Scroll down through the metadata in the **Description** tab. You will see that this format or style of metadata stores more attributes about the data and is the style of the Federal Geographic Data Committee. It is more complete than the item description we just viewed.
4. Click on the **XYBuildings** shapefile and go to the **Description** tab. Click **Edit** at the top of the display window. You will be editing the first two sections of the metadata, the **Item Description** and the **Topics and Keywords** sections.
5. For the Title of the data, name it **Building Corner Points**, for the rest of the fields, use the lab description to make up some information about the data. If you are not sure of what kind of data belongs in a box, click on that box and a description will appear at the bottom of the page.
6. You have now edited the first two sections of the metadata. Save it and export it. When we export the data, it is exported into an XML. XML data may be imported if needed.

3 Data Aggregation

Data aggregation is the process where we join, merge, or abstract data to suit our needs. This can sometimes cause issues between datasets but it may also be done in such a way that will protect the information at a lower level.

1. Open **ArcMap** to a blank map.
2. We will be looking at a broad example of how this works. Add the **Countries** layer to the map
3. Using the search bar, find the **Dissolve** tool. It will be in the **Data Management** toolbox and the **Generalization** toolset in **Arc Toolbox**. Using the tool dissolves the countries. Your Input Features will be the **countries** shapefile. In the Dissolve Fields box, select **SubRegion**. Save the result as **SubRegion** in your Lab 5 folder. Click **OK**.



4. Give the tool a minute to work. The result is an example of data aggregation. All the data in the map will be mapped using those regions. If we have data that we want to add that is specific to a country we would run into some problems as the data could only be mapped to the entire region and not just the country. The data would have to be manipulated and massaged to work in this scenario. This is data aggregation.

Conclusion

Spatial and attribute data accuracy is important. Having a complete dataset and keeping track of all the edits and errors is also important. Metadata provides descriptive background information that helps users to identify, interpret and use the data. It supplies us with vital information regarding the use and resolution of the data.

Reading the metadata can be helpful in determining if the data is an appropriate choice for a project, before we try editing and manipulating the data. Data aggregation is a way to abstract the data and to remove data levels from the data. In Task 3, you changed the lowest level in the data from countries to sub-regions. This change could potentially create errors in our data. We need to be aware of the consequences of our actions when mapping data and how the accuracy is changed when we manipulate it.

Discussion Questions

1. Does metadata need to be provided for all datasets?
2. Why do we need a metadata standard?
3. How is data aggregation problematic in a real world mapping scenario?