West Virginia Site Locator
and
West Virginia Street Locator

West Virginia GIS Technical Center
Final Report

Prepared August 2009

Revised June 2010

Revised February 2014

West Virginia GIS Technical Center
Morgantown, WV 26506
304.293.9462

Introduction: West Virginia is nearing completion of the Statewide Addressing and Mapping (WVSAM) project, a five year effort to update and address every structure in the state with a city styled address. The WVSAM project has resulted in the creation of many important GIS datasets, including point structure data and road centerline data. These two datasets have enabled the creation of a new geocoding system (address locators) for West Virginia. With the cooperation of WVGISTC, the Statewide Addressing and Mapping Board (SAMB), Department of Homeland Security and Emergency Management (WVDHSEM) Both point and street locator services are initially being developed as ArcGIS Server geocoding services. Once the services are functional, the online web applications, MapWV portal and WV FEMA Flood Hazards Mapping application, will be enabled to take advantage of the new services.

There are two locator services available from the WVGISTC, the WV Site locator and the WV Street locator. The WV Site locator is based on the point location of addressed structures, the WV Street locator is based on the addressed SAMB Street dataset. The WV Site locator service can offer a *direct match* of the physical address and location of the queried address. The WV Street locator service can provide an estimated match of the queried address. Both services will offer a highly accurate geocoding service as compared to any other commercial geocoding service available to West Virginia. Although the new locator services offer highly accurate returns, the current status of all addresses in the locator services are not 100% complete and will vary between counties. Because the addressing project is on-going and the data is constantly being updated, the WVGISTC plans to update the locator services quarterly with new data.

WVGISTC Development Process:

Identify Schema – Identified the required data schema of the locator services and the data schema (structure) of the data source.

Developed Test Locators - Built test locators with data from 6 counties. After successful application, WVGISTC built locators with data from the entire state. This allowed WVGISTC to document and develop the data processing steps.

Developed Scripts – The data source (SAMB) is multi-part, both geographically and in multiple data tables. The scripts automate the process used to join the data multiple tables, clean and edit data, and then build the locator services.

Developed Beta Locators – The WVGISTC made available the WV_Site and WV_Street beta locator services to select WV agencies for trial. The beta testing was intended to discover issues with the functionality of the services and to expose those agencies to familiarity of the tools.

Developed End Use – The final phase of building the locator services was integration into the web-based mapping sites developed by the WVGISTC; such as the FEMA Flood Hazard tool and MapWV.Gov . The integration utilizes the ArcGIS Server locator services but also incorporates custom programming to enhance the functionality of the services.

Locator Build Process: (Appendix A) for diagram.

WVSAM Data extract: The source data for the locators resides in an Oracle Database maintained by the office of WVDHSEM. The data is updated by local county E-911 administrators and is updated daily. WVGISTC worked with the original database contractor, Michael Baker Jr. Inc., to implement an extract tool that allows for the easy extraction of data from the WVSAM Oracle database. The extract tool operates as part of the WVSAM's Administration Web page as a button and takes 2-5 hours time to complete. The extract tool processes each county, extracts and compresses the data into a single file which is made available to download. The WVGISTC runs the extract tool quarterly to process the data and update the locator services.

WVSAM Data Processing: The raw data needs to be processed before it can be used as reference data for the locators. The extract tool extracts the raw data per county and separate data tables which need to be joined in order to create the locator reference data. Specifically each site and street feature (per file, per county) needs to be appended with the street name, zip code and city information, which resides in separate tables. In addition, the raw data contains numerous un-addressed sites that are removed during a cleaning step. (See table below) This cleaning process was necessary as it reduced the dataset size, which allows quicker processing time. These processes were automated through custom scripts written in python. The computation time for the automated process is between 4 to 5 hours and ultimately produces two shape files, one each for the sites and streets. After the script produces the two reference files, the site data goes through one more manual process that assigns a zip code and city to addressed sites that do not have one assigned.

WVSAM Sites April 2009 Raw	1,286,257
WVSAM Sites April 2009 Addressed	874,592

Data Processing: (Appendix B) The following is a schematic of the locator reference data processing steps. The script processes steps A-C on each county dataset. Step D appends the individual county site and street files into two datasets, WV_Site.shp and WV_Street.shp. Step E is not part of the automated process.

> Step A.) The first step joins two data tables, the individual county street_name_join.dbf table with the domain table street_name.dbf table. The join uses the primary key Street ID. The Street_Name.dbf table contains the street prefix, street name, street suffix and street suffix direction, which the join adds to the street_name_join.dbf table, which will then be joined to both the street and site files.

> Step B.) The second step processes the site data. The first process is to clean the county site data from non-addressed sites. This process reduces the size of the files which can speed the processing time both. The second process copies address numbers that may have been entered into the address label field and not the address number field. The third process joins the street information to the site data. The fourth process joins the city names to the site data.

> Step C.) The third step processes the street data. The first process joins the street information to the street data. The second process joins the city data to the street data. This process is done

twice because the street data uses a left and right city each.

Step D.) The fourth process step appends all county site files and county street files into a single site and street file respectively.

Step E.) The final process step is to integrate zip codes for sites that do not have zip code information included. This is done manually.

ZIP Code Process: Though most address points contain a zip code and Post ID (city location), there are some sites that do not have a zip code or post ID assigned. (See table below).   In order to make use of these sites in the locator service they must include zip codes. These sites could be processed by a couple of different processes. The first option considered was to reverse build a zip code boundary from the WVSAM sites and use the resulting file to intersect the sites; the other possibility was to use an existing zip code boundary to intersect the sites. After analysis, the WVGISTC used the existing zip code file to process the sites because the generated zip code file contained numerous errors.   (See Appendix C) The same process was tested to apply city information (Post ID), however boundary data is not available to use as a reference. The results were not 100% due to discrepancies of the zip code boundary data along state lines.

WVSAM Data: April 2009	Zip Code Present	Postal Local (Post ID)
WVSAM Addressed Sites	702,745 (80%)	710,899 (81%)
Post-Processed Sites	874,534 (99%)	n/a

WV Site and Street Locators: The locators are built manually in Arc Catalog using ESRI ArcGIS locator tools and published to ArcGIS Server. This process is not time consuming and therefore was not automated. There are two separate locators that are built in Arc Catalog using the Create Locator tools, one each for site addresses and street ranges. The following information conveys the locator setup and which fields in the table are used for reference.

WV_Site: The WV_Site locator is built using the "US One Address with Zone" with the following settings. (The Bold categories are required fields.)

Primary table: This data is the product of the SAMB data script, which produced a statewide file of sites, which was then post processed to populate blank zipcodes; the resulting file is used as the primary table.

House Number: [Address_Nu]
Prefix Direction: [Prefix_Dir]
Prefix Type: *not used*
Street Name: [St_Name]
Street Type: [Suffix_typ]
Suffix Direction: [Suffix_dir]
Zone: [Zip]

WV_Street: The WV_Street locator is built using the "US Streets with City, State and Zip" with the following settings. (The Bold categories are required fields.)

> **Primary table:** This data is the product of the SAMB data script, which produced a statewide file of streets, which was then post processed to populate blank zip codes; the resulting file is used as the primary table.

> > **House From Left:** [L_From_Add]
> > **House To Left:** [L_To _Add]
> > **House From Right:** [R_From_Add]
> > **House To Right:** [R_To_Add]
> > Prefix Direction: [Prefix_Dir]
> > Prefix Type: *not used*
> > **Street Name:** [St_Name]
> > Street Type: [Suffix_Typ]
> > Suffix Direction: [Suffix_Dir]
> > **Left City:** [Left_City]
> > **Right City:** [Right_City]
> > **State Abreviation:** [L_State]
> > **Left Zip Code:** [L_Zip]
> > **Right Zip Code:** [R_Zip]

Web Site Integration: The locator services were integrated into several web products of

WVGISTC, including the FEMA Flood Hazard Tool. The intent is to provide greater accuracy to the end users, specifically enabling access to the high spatial accuracy WV_Site locator. The point based locations are not available from any other commercial geocoding service.

Cross Reference: The WVGISTC improved the functionality of the locators operating inside the web-based applications; this enhancement will allow a user to enter an address with a city without the zip code. The locator tool will then perform a cross reference lookup of the city and automatically search for the correct corresponding zip code(s) from US Census (2007) zip code data. If multiple zip codes exist for the same area, the tool will find the best possible match using each of the zip codes and return the site with the highest match score.   This enhancement improves the chances of getting a match using the WV_site locator, which would require a zip code and ignore city data as entered.

Cascade Locator: The locators integrated into the FEMA Flood Hazard Tool uses multiple locator services. The address tool uses a cascading program which uses a hierarchical approach as the best available match is returned from the WV_Site, WV_Street or Google Locator Services. The input address from the user is entered into separate fields which allows the application to easily identify the street address, zip code and city. The input address and standardized address are returned and displayed to the user to verify correct entry. The standardized address is then submitted to the locators in a cascade fashion; the order is set below.

Cascade Locator Schematic

Input: Submitted Address
> Parse Input address: City / Zip Cross Check
>
>Return Standardized address, (contains validated city and zip code)

> Poll WV_Site: Standardized Address
>Match ? = 100%
> Yes = Return Address and Location Point
> No = Poll WV_Street: Standardized Address
> Poll: WV_Street / Match
> Poll WV_Street: Standardized Address
>Match ? = 60%
> Yes = Return Address and Area
> No = Poll Google Locator: Standardized Address
> Poll: WV_Street / Match
> Poll Google Locator: Standardized Address

Return Error Messages: Upon completion of an unsuccessful search, the application may return one of three errors to inform the user of a problem. These types of error messages may help a user identify and possibly correct the problem.

1.) "Not a valid input" – This error is returned if a user enters in only a city or only a zip code.
2.) "Address Not Found" – This error is returned if there are no valid returns from any of the locator services.
3.) "Not a West Virginia Address" – This error is returned if the address submitted contained a city or zip code outside of West Virginia.

WVSAM Site Derivative Products: (Appendix D) The WVSAM site data enables the creation

of numerous derivative products such as the example below, which demonstrates the generation of a new zip code polygon. However, this process conceptually can be applied to any attribute, such as emergency service zones or cities, creating a new addressed based polygon dataset. This process is not limited to attributes already assigned to the dataset, but can be added at any level that ties into the address of the site. For instance, the state tax office can calculate boundaries of incorporation, or local municipalities can determine actual voting precincts.

WV_Site and WV_Street Locators: (Appendix E) The WVGISTC offers both locators as an

ArcGIS Server Service. This means that the locators can be used directly in an ArcMap session and be used to geocode either a single address or table of addresses. This takes advantage of the existing toolset that ESRI has developed using either a find tool or specific geocoding interface which allows the user to interact with the data. Though the drawback of using the locator services in ArcMap is not able to take advantage of the zip code and city cross check that our web-based applications enjoy.

Summary: The new WV_Site and WV_Street locators made available by the WVSAMS data will be a valuable tool to assist many end users in multiple applications locating specific addresses. As more counties participate and contribute to the WVSAM system, the locator services will improve in parallel. As more people begin to use the locator services, they will be able to discover that the accuracy is much better from participating counties over other available commercial geocoding systems. In addition, the WVGISTC will continue to work with WVDHSEM, counties and end users to improve the functionality and availability of the locator services. If you have questions or comments about using the locator services please contact Kurt Donaldson or Kevin Kuhn at the WV GIS Technical Center.

Appendix A.) Process Flow Chart

| County E-911 | ⇄ | WVSAMS Data | Oracle Database |

| WVDHSEM |

Extract — WVSAM Extract

Pre-Processing — Python Script

Data Enhance — Zip and City Data Enhanced

Reference Data
wv_site.shp
wv_street.shp

WVGISTC

Zip Code Areas

City Areas

Custom Application

Manual Process — ESRI Arc Catalog

Create Locators

End Users

Locator Services — ESRI ArcGIS Server

ArcGIS Locator Service

Web Site Integration

MapWV   FEMA Flood Hazard

Appendix B.) Data Process Schematic

A.) Table prep:
   Join <u>Street Name Join.dbf</u> on [NameID] to <u>Street Name.dbf</u>

B.) Site:
   *County_*Site.shp
       Remove non-addressed records
           Select if [ADDRESS_NU] = 0 AND [ADDR_LABEL] = '0'
           Delete Selected
       Populate blank addresses in address number from address label
           Select if [ADDRESS_NU] = 0 AND [ADDR_LABEL] <> '0'
           Calculate value [ADDRESS_NU] from [Addr_Label]

   Join *County_*Site.shp to <u>Street Name2.dbf</u> on [StreetID]
   Join *County_*Site.shp to <u>Postal_LLC.dbf</u> on [PostID]

C.) Street:
   Join *County_*Street.shp to <u>Street Name2.dbf</u> on [StreetID]
   Join *County_*Street.shp to <u>Postal_LLC.dbf</u> on [L_Post_ID]
       Select if [R_post_ID] = ' ' OR Select if [R_post_ID] = '0'AND [L_post_ID] <> ' '
           OR [L_post_ID] <> '0'
       Calculate value [R_post_ID] = [L_post_ID]
           Left
           Right (The streets locator needs both L/R city)
D.) Append:
   Append *county_*site.shp to wv_site.shp
   Append *county_*street.shp to wv_street.shp

E.) Populated blank zip codes
   Select if "ZIP" = ''
   Analysis Tools>Overlay>Spatial Join

Appendix C.) Zip Code Enhancement

Enhance Zip Code: Using the SAMB sites, we were able to generate a thiessen polygon layer which contained the zip code as an attribute. This process essentially creates a polygon using the SAMB site as a centroid and the boundary is half the distance to the next point in all directions. From the thiessen polygons we performed a dissolve, leaving single polygons of a single zip code area. This method was tested and worked well for most areas, notably well for dense areas and included few missing points. However, this process failed in areas that have large clusters of missing data, where the generated zip code boundary did not have enough reference points, and therefore did not accurately reflect the true zip code boundary (as compared to the US Census zip code boundary). Because of the incorrect and incomplete divisions, these polygons were not able to transfer the correct zip code to the site, making this process invalid.
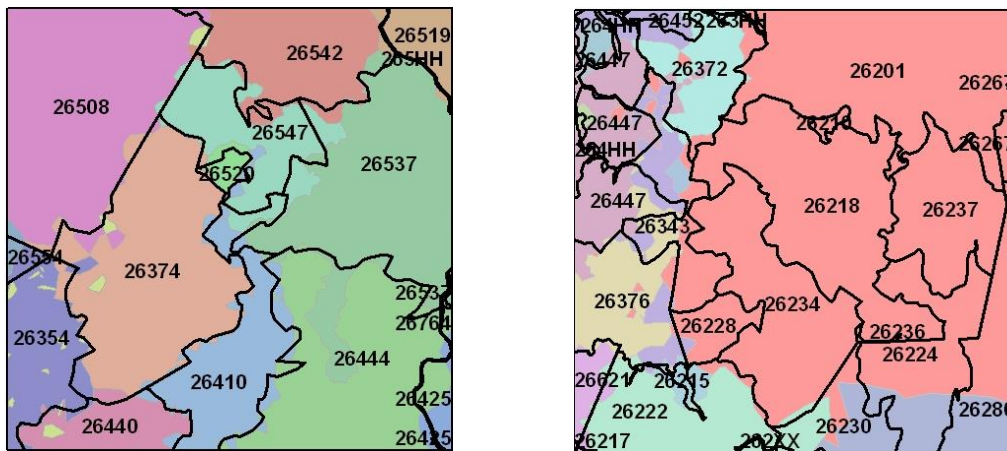
It should be noted that as the WVSAM data becomes more complete, the accuracy of the data will improve and it is feasible that this process can be used in the future. Using this process, we are able to discover areas of discrepancy between the US Census zip code file and that of the true geographical extent as listed by local officials. This process could therefore be used for correcting sites and streets missing zip codes and Post Ids. In addition, this process could be used to improve the existing zip code boundary file or generate a new one.

Appendix D.) WVSAM Addressing Data Derivative Products

Generate Area: This process was performed in ArcMap using various tools to create zip code areas polygon using the WVSAM Site data.

1.) Create Thiessen Polygon: This step creates polygons for every SAMB Site Point. The Output Field:ALL option is used to assign the appropriate attribute to the polygon, such as Zip code or Postal ID

>Analysis Tools > Proximity > Create Thiessen Polygons

>Input: SAMB Site / Output / Output Fields: ALL

3.) *Dissolve*: This step dissolves polygons based on an attribute, such as the Zip Code or Post ID, which creates a single polygon of like attributes. The option of "Create multipart features" may be used otherwise this may generate more than a single polygon of each attribute if it is separated by other polygons, IE you may have multiple polygons of the same attribute.

>Data Management Tools > Generalization > Dissolve

>Input / Output /Dissolve_Field: *f(f=zip code or f=posted)*
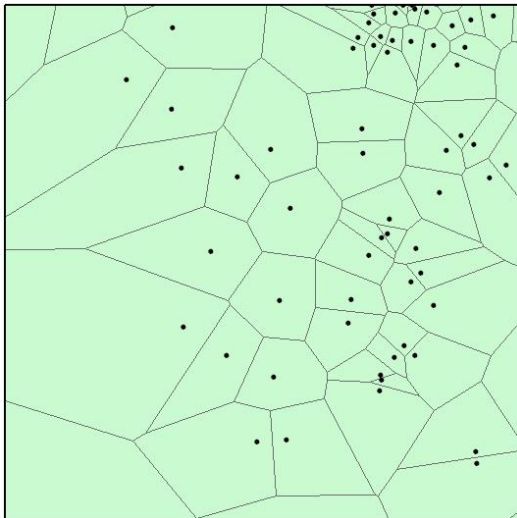
SAMB Sites and the generated Thiessen Polygons . In this graphic, you are able to visually see how a polygons boundary is determined by dividing the distance in half to each adjacent site.
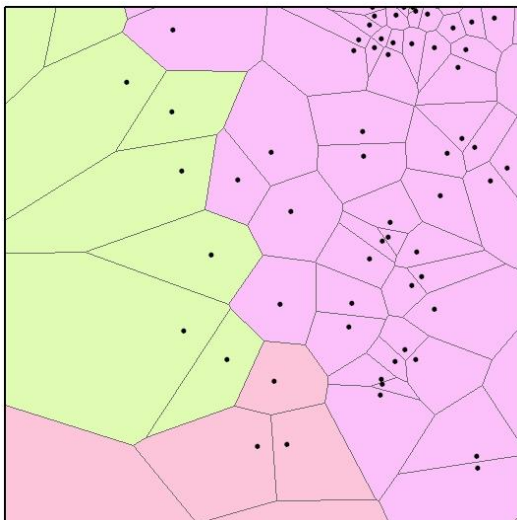
Thiessen Polygons displaying different zip codes by color from attribute information.
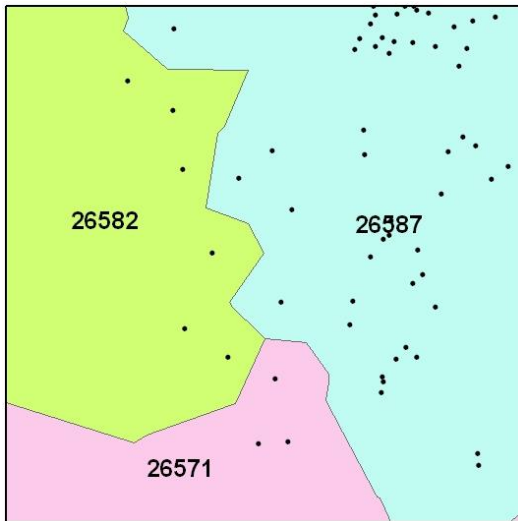
Thiessen Polygons dissolved by zip code attribute and displaying SAMB Sites
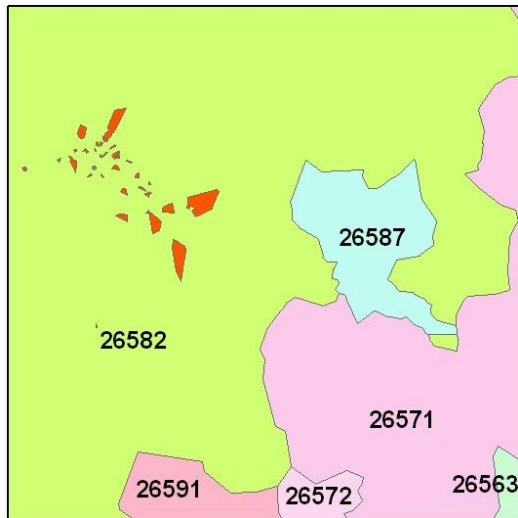
Generated Zip code polygons. The red polygons are the result of site locations that did not include a zip code attribute. When producing an actual zip code or postal local file, these could be merged into the larger polygon that surrounds them.
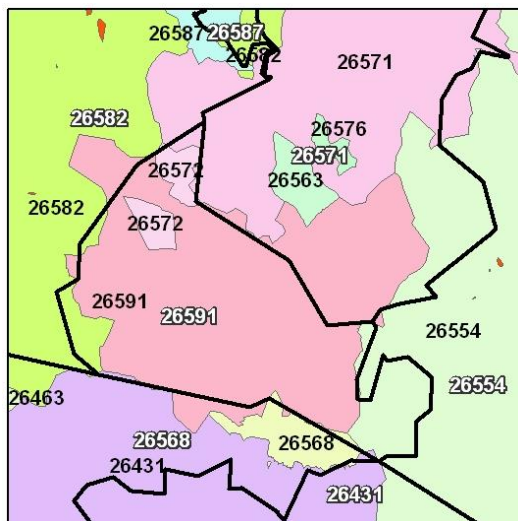
US Census zip code boundaries (Black lines) overlaid the SAMB generated zip code polygons (colored). From this graphic, you may see that there are major discrepancies in how areas are assigned zip codes versus the reported US Census zip code boundary file.

How to Setup the WV Locators From the WVGISTC: The locator services can be used to process single addresses or many addresses (bulk submittions) in ArcMap. Before you start geocoding, you must connect to the geocoding services to be used. Follow the instructions below to connect the WV Locators in ArcCatalog.
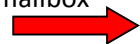
This guide is designed for Arc 10.2. For other versions or for extra assistance please consult the "Help" document "Working with address locators and geocoding " within ArcGIS.

ArcCatalog: add an ArcGIS Server

1. Launch ArcCatalog
2. Click GIS Servers in the table of contents area (see the illustrations below)
3. **Double click on the Add ArcGIS Server item**
4. **In the Add ArcGIS server dialog, select "Use GIS Services," then click Next**
5. **Enter the address for Server URL:** http://services.wvgis.wvu.edu/ArcGIS/rest/services/Geocode
   Click Finish
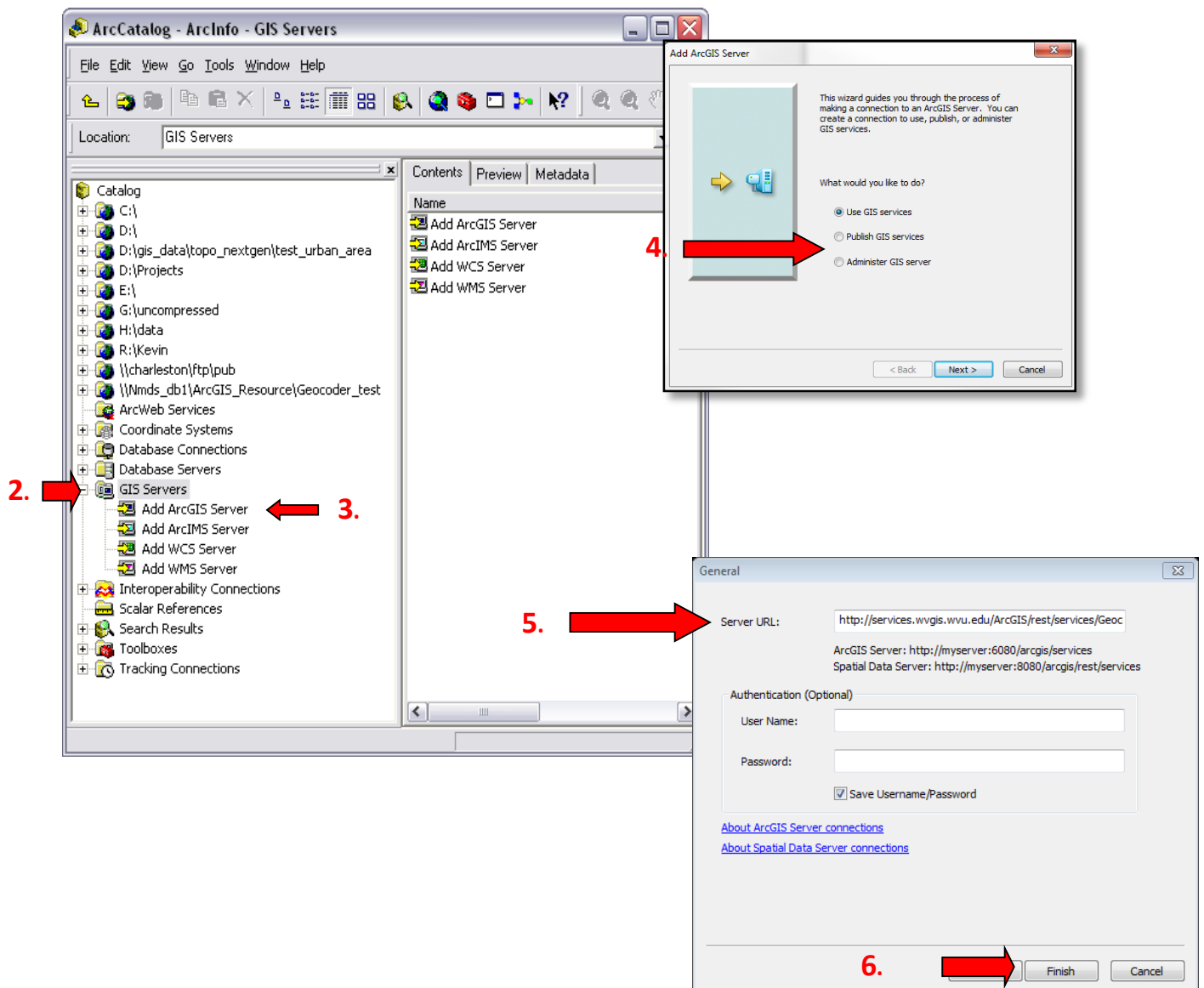
**ArcMap: add locator**

.

1. Launch ArcMap
2. Click > Customize > Toolbars> Geocoding
   To put the geocoding toolbar on your map document
3. Click > 1ˢᵗ Drop down menu or on the mailbox
   icon within the Geocoding Toolbar
4. Click> Manage Address Locators
5. **Click > Add**
6. **Select > GIS Servers**
7. **Select "http://services.wvgis.wvu.edu/ArcGIS/rest/services/Geocode"**
8. **Click Add**
9. **Select WV_geocode_Site and/or WV_geocode_Street**
10. **Click Add**
11. **Click Close**

3.

3.

5.

6

7

8.

9.
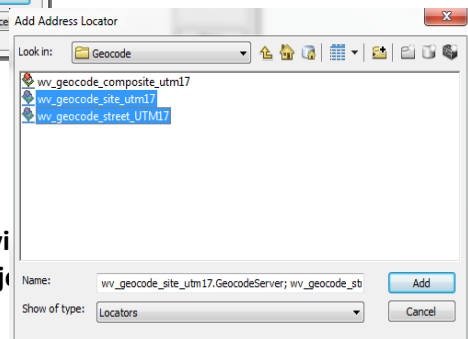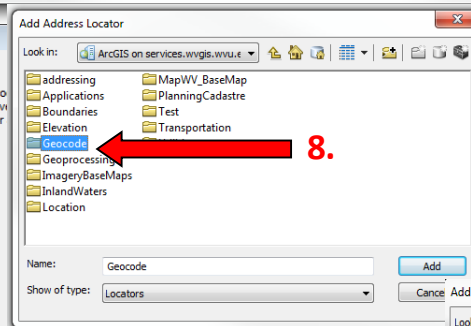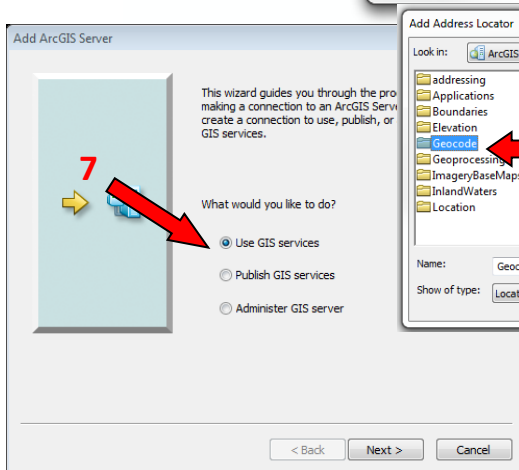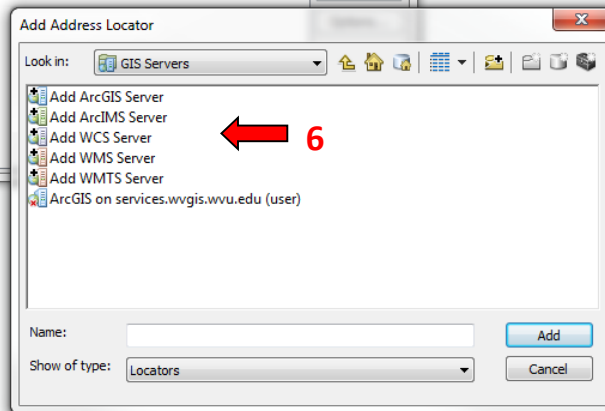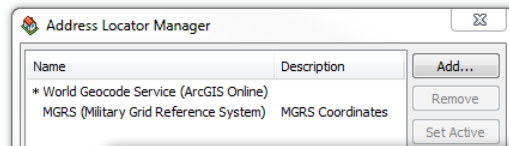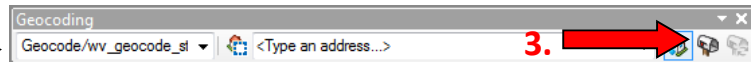
**How to Use the Locators: Once you have set up the locator servi
use them within ArcMap. You can either use a new arcmap proj**

Single Address: If you only need to locate a single address, The Find tool 🔍 **can be used to geocode.**

1. **Click the Find Tool**
2. **Click the Locations tab**
3. **Choose the locator wv_geocode_site or wv_geocode_street**
4. **Fill in Full Address**
5. **Click Find**
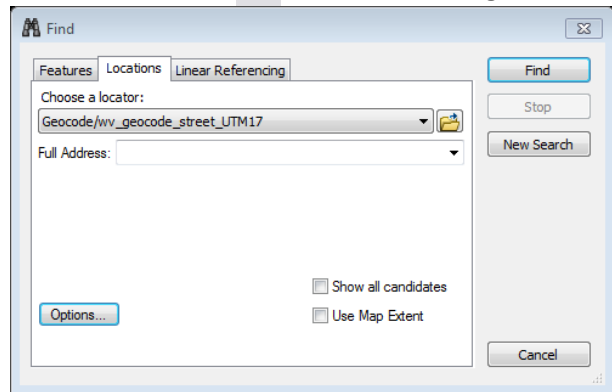
Bulk Addresses: Multiple addresses can be geocoded directly in ArcMap using t**able**s with the locator services.

1. Add a table to ArcMap
2. Click ArcToolbox
3. Click Geocoding
4. Click Geocode Addresses
5. Choose a locator, click OK
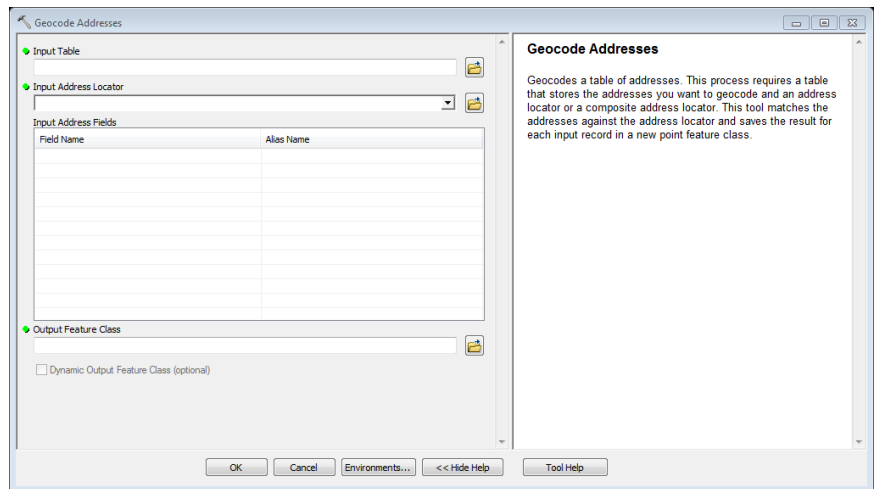
The table must comply with the minimum required fields for the selected locator (As listed below). The locator returns an event table in the ArcMap session. This event table can then be saved to a permanent shapefile. The locator service returns new fields (Listed below) associated with the event points created that allows you to determine the accuracy (or match quality) of the results.

Required fields:
WV_Geocode_Site_UTM17: street address, zip code
WV_Geocode_Street_UTm17: street address, city, state, zip code

Returned Results: (From the ESRI ArcMap Help) After a table of addresses is geocoded, the output feature class will contain a copy of the address fields from the address table that are used for rematching. The names of this set of address fields are prefixed with "ARC_". When an address is modified during rematching, the new address will be saved in these fields. In addition, here are the other fields created in the output feature class for the geocoding results:

- *Status* - A code indicating whether the address was matched. This attribute has values of "M" for matched addresses, "U" for unmatched addresses, and "T" (tied) for addresses for which there were more than one candidate with the best match score.
- *Score* - The match score of the candidate to which the address was matched.
- *Match_type* - A code showing how an adddress was matched. "A" for automatically matched or rematched. If the geocoded feature classes is rematched using the Interactive Rematch dialog box, you may see other values including "M" for manually matched or unmatched using the Match or Unmatch button. "PP" stands for Pick by Point. The address was matched to the click point using the Pick Address from Map tool in the Interactive Rematch dialog box with ArcMap. And "PA" stands for Pick by Address. That means the address was matched to closest address with the Pick Address from Map tool. You may group the results based on this attribute to show how the addresses were matched or use the grouping to select records for rematching.
- *Side* - The side of the street to which an address was matched—if the address locator that was used to match the table contains address information for both sides of the street. This attribute has values of "L" for the left side of the street, "R" for the right side of the street, or nothing if the address locator could not determine the side of the street.
- *Match_addr* - The address that the matched location actually resides based on the information of the matched candidate. For example, an input address "123 Main St N" is matched to a candidate with the suffix direction 'NW" and all other components matched correctly. The Match_addr field will contain "123 Main St NW" as the actual address that was matched.

    There are also some optional fields if the Output Fields options were checked in the address locator's property page:

- *Pct_along* - For address locators that use feature classes with polyline geometry as reference data, the percentage along the matched feature at which the matched address is located.
- *Ref_ID* - The ObjectID of the matched feature.
- *Stan_addr* - The standardized address.
- *X* - The x coordinate of the matched address.
- *Y* - The y coordinate of the matched address.