Using the VDAT geocoding Script

This document describes the process for using the Red Cross VDAT Java Script to geocode address information. Geocoding converts address information to points that can be placed on maps.

If you plan to use geoprocessing tools it is essential that locations

## Step 1: Prepare RCCollect data for geocoding

1. Download the RCCollect Layer and as a CSV file.
2. Open the file in excel and select all
3. Do a find and replace (with blank spaces) to remove the following:
	1. Quotes (single and double)
	2. Slashes (backward only)
4. Save as a CSV file.
5. Go to script at URL [www.lanktonschmidt.com/VDAT\_Geocoder.html](http://www.lanktonschmidt.com/VDAT_Geocoder.html)

## Step 2: Using the VDAT geocoding script

1. Enter google API key. Follow instructions on the site if you do not have a google key yet
2. Click on the Browse button and locate your CSV file.
3. Select the Fixed Format VDAT DDA (assuming you are doing DDA processing)
4. Enter the distance between measured point and geocoded point at which a reviewer would want to take a look. When it occurs the note column of the output will have a “dist=2345 feet”. Geocoders can and incorrectly identify properties so anything over 2.5 miles ( 2.5 \* 5280 ft./mile is about 10,000 feet) should be reviewed. If you enter zero, all entries will have distance traced out.
5. Press the Purple Geocode button
6. As the script is processing, you will see the count of addresses processed and the number of addresses to be processed - Processed Line 5 / 809
7. You may receive a message similar to the following text. {Operation failed on Line 317 because of invalid charater in {318,{9DB1B7BB-5AEA-4F88-BD5E-DCF0DFD7FCB6},,7/5/2018,638-19,Polk County IA 7/18 Floods,DDA,flood,,Single\_Family,Yes,Yes,,16,Unknown,96,HVAC and water heater submerged,Minor,,,,,,Minor\_Mechanical,,,,,,,,,,,,,,,,,,,,,,1312,,NW 78th St,IA,Clive,,NW78th St and University,19153\_Polk,15R08 - Iowa Region,15080 - ARC of Central Iowa,"Garage had 36"""" water in addition to residential flooding",Cathy Olson,7/5/18 7:30 PM,5312618,7/5/18 7:49 PM,5312618,-93.72412207,41.60171494 }. This message indicates that on line 317 there was an invalid character. If you look through the text you will see that the {“ after Garage had 36”} was not removed. Return to the csv file in excel and make sure that all quotes are removed. There may be another special character that is not allowed and you may need to go to line 317 in excel to review the entries for special characters.
8. While the script is running, you will see a running display of three lists. An example is shown below. The leftmost window is your output data (text is purple). The other windows are statistics that show the quality of the data. These statistics will be described later. When the geocoding is complete, you may want to copy these statistics so that you have them. They will not be available after refreshing the browser.

Totals, accuracy (shows number of each accuracy rating

202 10 Rooftop (most accurate)

61 7 Interpolated (suspect unless in rural areas)

1 4 Point might be in the city, but not near house

2 2 No Idea where property is located.

Totals, error\_in\_feet (summary of all distances between measured and geocoded points)

0 0-100

3 10000 +

1 2000-4000

1 4000-10000

User, Success, Duplicates (Surveyor accuracy statistics)

Bob,Good=258/272,Dup=1

Jim,Good=12/56,Dup=5

 9. When complete, you will receive the message like - Done! Processed 807 lines!

## Step 3: Capturing Geocoded data

1. To save the output, select the contents of the purple text and copy into A1 cell of a new spread sheet.
2. If not already done, use the Text to Columns feature, from the ribbon, to convert it to different columns. During the conversion, make sure to switch the data types for House number, apt number, and zip code to text instead of numbers.
3. To save the statistics, select the contents of each window and copy them elsewhere. They are formatted as csv is you copy it into a spreadsheet if needed.

## Step 4: Understanding the output

1. Anything in upper case was arrived by reverse geocoding the survey point. All else is in lower case regardless of what was in the survey.
2. Anything in the r\_? fields in CAPS is there because the address did not accurately geocode and the survey point was reverse geocoded. This is mainly provided as a hint because survey points are not generally on the property in questions.
3. Anything in the r\_? fields that is in lower case is there because the script found the address, but did not match exactly what was in the survey. You should review the changes and make sure they make sense.
4. Look in the “note” column for “dist=x feet” which indicates the distance between the survey point and geocoded point. If this value is over 10,000 feet, the entry needs to be reviewed.
5. Look in the “note” column for “duplicate(x)”. The Script records all “y” column values and when it detects another instance.
6. Look in the “note” column for “ReverseGeocoded”. When this occurs you know the address information as the result of reverse geocoding a measured point. This only occurs when no address information was provided. If most of your data had address, this should be a red flag as no address information could have been a data entry error. This should only be allowed if all entries were provided without address or you know for certain the points were over properties.

## Step 5: Cleaning up the data

## During the next section, we will review address and investigate ones that do not look right. we will reference both Bing and Google maps. Each mapping platform some advantages you should be aware of.

## Both:

Allows you to enter address or lat/long information and have it find it on the map

Allows you to copy corrected address and lat/long from places it finds

## Bing:

Allows you to put multiple address at a time. Nice feature for comparison. Geocoder is more tolerant of misspelled street and city names.

## Google:

Allows you to just press a button over a house to get its address. This is helpful to get a sense of what street names and house numbers are for nearby homes.

The Script uses Google’s geocoder, so use it to verify an address will be properly located by the script?

Fixing poor data Examples:

# The first pass

Included with this write-up is a filed called "Reviewing Script Output Example.csv" which should be feed into the script. The output of the script is "reviewed" as described below.

35. (7929 New Rider St, North Charleston, SC)

This entry had an accuracy of 2 which means Google does not know where the property is located. The first step to resolve this is to place the measured point (32.94356509, -80.05140137) and the address on a Bing Map. The measured point is very close to the address so I'm guessing Google got confused by the misspelling of "Ryder" as "Rider". Bing appears more tolerant of misspelled street names. When I enter the corrected address (7929 New Ryder Rd, North Charleston, SC 29406) in Google, it finds the spot correct. To fix this, replace column "streetname" with value "New Ryder Rd" and add the value "redo" to the notes column so it gets re-geocoded on the second pass. If you want, you can also put additional comments in the note column like "misspelled name".

53. (7939 New Ryder Rd, North Charleston, SC)

This entry had the street name changed from r\_streetname="New Ryders Road" streetname="New Ryders Rd". The Accuracy is 10 and distance is 17 feet so I'm guessing this is a good address. Nothing else needs to be done.

176. (3002 Lee st Lee St, Edisto Island, SC)

This entry has a couple of problems. The first is a questionable houseno="3002 Lee st". We frequently see this because DA personnel start typing street name before moving the cursor over the correct location. The second is an incorrect city name that was changed from "Edisto Beach" to "Edisto Island". The accuracy is 7 and distance is 101 feet so I'm guessing this is the correct address. To fix this, change column houseno from "3002 Lee st" to "3002", add "redo" to the note column so it gets re-geocoded on the second pass.

197. (429 E Black Creek Rd, Florence, SC)

This entry has the string "dist= 45232319 feet" in the note Column and needs to be looked into. Start by putting both measured Point (-90 -77.43112564) and address on Bing. You will notice the -90 in the y component which is in Antarctica so it is not a valid value and can be ignored. Nothing needs to be fixed here. You could put a comment in the notes column indicating measured point invalid.

206. (606 Roosevelt St, Florence, SC)

This entry has dist= 4945 feet and is a 7 accuracy which could indicate a problem. The street name "Roosevelt" was replaced with "Roosevelt St" which is minor and accuracy was 7 so it should be ok. Place the measured point (34.21427114 -79.78919451) and address on Bing map. Bing could not find an exact match because it did not have the house number in the address it returned for the provided address. Google could not accurately find it either. If you look at the Google map, notice the point is not over a house (which is common with 7 accuracies). The next step is to use Google to reverse geocode a few points around the measured point. In this example, the house just left of the point picks 710 and it is at the end of the street as the street name changes as you go around the bend. The next house is 608 Dickman St and it is at the end of the street so could not have been 606. Most likely the DA person was mixing street names near their intersection. This is a common occurrence. Unfortunately, in this case we cannot determine what house they identified. The accuracy is 7, but we know it cannot accurately find the house. To fix this, all we can do is change the accuracy to 5 so people know the house is unknown. You could also add " Selected house not certain" to the note column. Nothing else needs to be changed.

222. (8076 Karl Rd, Loris, SC)

This address could not be geocoded because the accuracy was 2. In this case, the script reverse geocodes the measured point and places streetname, city and state differences in the r\_? fields in upper case to help you find the correct location. In this example, only the street name was different ("HENRY LEWIS RD"). To fix this, enter the measured point (34.01247507 -79.00370317) and the address in bing. Now, look around for entered street name "Karl Rd". Just below the point is a road called "Kerl Rd". Try using Kerl Rd instead of Karl Rd (8076 Kerl Rd Loris SC) and it finds the address. To fix this, change "Karl Rd" to "Kerl Rd" and insert "redo" in the note column.

320. (7 Olive Shell Ct, Spanish Wells, SC)

This is similar to the previous one, where the reverse geocoded was done, except the city varies as well as the street name. This is another common problem for DA personnel because city boundaries are not obvious without a map. Start by entering the measured point (32.20554766, -80.74814236) and address in Bing. Bing shows the address at the measured point to be (17 Demsey Ln, Hilton Head Island, SC 29926, US). Bing finds the address and it’s next to the measured point. However, when I look at Google map, the road does not even show up. Obviously Google needs to update it maps. You can see this road has recent construction on it when viewed via satellite. In this case, Google is never going to find it so we have to manually set the location and force accuracy to 6. Update the (y,x) fields with the Bing result (32.206204, -80.74874) and update the city to "Hilton Head Island. Do NOT put "redo" in the note column because we do not want the script to change any values because Google cannot locate the property.

524. (2311 Lester Rd, Dillon, SC)

This entry did geocode, but the street name was changed from "Lester Road" to "Lester Rd", the city changed from "Floydale" to "Dillon" and the accuracy was 7. Because the city changed, we need to verify this by entering measured points (34.32576113 -79.33409693) and address into Bing. Interestingly, Bing found both points close. Entering the same address into Google shows it goes to the wrong spot. By going to the spot on Google and reverse geocoding over the house Google cannot find the address so it probably does know about the property. To fix this, update the (y,x) (34.325783, -79.334015) locations from Bing maps and mark the accuracy as 6. Do NOT add "redo" to the note column.

526. (b 1st fl / 103 Cedar Reef, Harbor Island, SC)

This entry has "b" for its houseno which is invalid. There is no way to fix this.

986. (4987 Sequoia Ln, Georgetown, SC)

This entry has an accuracy is 2 so it could not be located. Start by entering the measured point (33.31004139,-79.33197361) and address into Bing to see what it returns. The address could not be located. Looking around when the measured point is located and just south of the point is a small road called Sequoia Ln. Google does not know the name of this street at all. However, if we use Google to reverse geocode some of the properties, we find most homes nearby all reference Fraser St. If we change the address to "4987 Fraser ,Georgetown, SC both maps correctly find the house. To fix this, update the street name to Fraser St and add "redo" to the note column.

1154. (3 Goettee Game Rd, Hardreville, SC)

This entry has an accuracy is 2 so it could not be located. Start by entering the measured point (32.2267725, -81.03601545) and address into Bing to see what it returns. Bing does better at finding the address, but is not located over a house. Selecting various points around Google maps show neither Goettee Game Rd nor Freedom Pkwy have address near "3". I'm guessing someone did not fully enter the houseno so there is no way to fix this.

1508. (105 Eaddy Street, Hemingway, SC)

While this entry has accuracy 7, the dist=392223 feet which is about 74 miles so the geocoded point is suspect. Enter both measured point (33.05079836 -80.2529679) and address in Bing. The address appears valid and the measured point as it is over house with that address. However, both Google and Bing appear to pick the same house, so I'm assuming the measured point is incorrect. This entry is good.

1661. (0 Plumbridge Lane, Hilton Head Island ,SC)

Accuracy of 4, but since the houseno is 0, there is no way to find this address. Leave it alone.

1939. (5 Chantilly Ln, Hilton Head Island, SC)

This entry had the city changed from "Hilton Head Isl" to "Hilton Head Island". Nothing needs to be changed. Accuracy is 10 and measured and geocoded points are only 107 feet so nothing needs to be done.

2879. (517 Clearview Dr, Charleston, SC)

dist= 51115 feet so we need to verify it's correct. Place measured point and address on Bing. It's clear the address geocoded location is where the property is located. Nothing needs to be done.

3467. (706 Not sure Pee Der Island Rd, Nichols, SC)

This entry has a low accuracy. Start by entering measured point (34.22518049, -79.18222929) and address (minus "Not sure") into Bing. Both Bing and Google find it. Notice the misspelling of (Pee Der Island Rd) to correct spelling (Pee Dee Island Rd). There is a house near there, but when I reverse geocode it in Google, it does not show an address. However, the measured point is close to where it geocodes so it is probably good enough. To fix this, change the houseno to 706 and enter "redo" into the note column.

3571. (100 701 S, Socastee, SC)

Low accuracy geocodes. Put the measured Point (33.69265636, -79.11557902) and address on Bing. Neither Bing nor Google and find the address. However, Bing shows the reverse point on a road called US-701 which is close. Tried changing address to (100 US-701, Socastee SC) but the point it found was 100 miles away. Tried changing address to (100 US-701, CONWAY SC) using the city provided in the r\_city field, but that did not work either. Basically, this route does not have a house number near 100. So, I look around Google maps for nearby streets (close to highway in question) to reverse geocoding to see what other road it could be called that has 100 street number and found a property at 100 Leona Ln, Conway, SC, 39527. So, to fix this change the streetname to "Leona Lane", city to "Conway" and add "redo" to note column.

4272. (3694 5 thru 8 Clay Pond Village Ln, Myrtle Beach, SC)

This entry geocoded accurately, but changed the streetname from (Claypond Village Lane) to (Clay Pond Village Ln) and city from (Brickyard Plantation) to (Myrtle Beach). Both Google and Bing located the same building which was near the measured point so we can assume it's the correct address.

4896. (29 Fortune Dr, Lane, SC)

This entry geocoded accurately, but changed the city name from "Greeleyvill" to "Lane". Since the distance is 203 feet, it is probably the correct location. Nothing needs to be done.

8368. (....)

This entry has no address information, but did have a measured value, so the script reverse geocoded the point and placed its results in call CAPS. Also, \the script added "ReverseGeocoded" to the note column. Nothing needs to be done. Since there is no way to tell if the address is correct, you have the choice to accept it or delete it.

The Second Pass.

We made several changes to address that we want to re-geocode the output. When we run the script on the recently edited sheets, only the entries marked with "redo" get re-geocoded. We will new review those entries.

35 Fixed. Accuracy from 2 to 10

53 Nothing changed

176 Nothing changed

222 Fixed. Accuracy from 2 to 10

986 Fixed. Accuracy from 2 to 10

3467 Fixed. Accuracy from 2 to 10

3571 Fixed. Accuracy from 2 to 10

8368 Address forced to proper case but note field still shows "ReverseGeocoded"