Scan/US® Geocoder

User's Guide

Version 2024

Scan/US Geocoder

Scan/US Geocoder will append latitude-longitude coordinates and Census and other geocodes to your address file.

Geocoder will help you complete the following steps:

- Open a file containing street addresses.
- Identify input fields containing the required address components: street address, city, state and/or zip.
- Check input fields to pass through to the output record.
- Choose optional geographic codes to append in addition to geocodes that are automatically appended.
- Process the street addresses.

To continue, click Next.

About... QuickFind... ? < Back Next > Cancel

Scan/US, Inc.

Sales or Support: 800-272-2687 or 310-828-1450

9:00 AM-5:00 PM, Pacific time, Monday-Friday, excluding holidays

E-mail: sales@scanus.com or support@scanus.com

https://www.scanus.com

Eleventh Edition

February 2024

Copyright

Software License Notice

Licenses and Trademarks

General Notice

Using Scan/US Geocoder: the Scan/US Geocoder is a Windows desktop application designed to append latitude/longitude information to U.S. addresses contained in database files. Once lat/lon coordinates have been attached to addresses, you will be able to display those addresses as individual points in mapping applications such as Scan/US for Windows and use them for analytical purposes such as trade area determination, customer origination studies, competition maps or batched site analysis.

1. **Geocoder Start Page:** the welcome screen provides an overview of the address geocoding steps. The geocoder is designed as a wizard which will take you through the process step-by-step. Click "Next" to proceed to the input selection page.

2. **Specify Input File:** click "Browse" to select the address file to process. Geocoder supports ASCII delimited-text (**.csv** comma-separated or **.txt** tab-or-pipe-separated; first line must be a header), Excel (**.xlsx** and **.xls**; addresses must be in first worksheet, and first row must be a header), and dBASE (**.dbf**) file formats. Use "Files of type" dropdown to change file type filter.
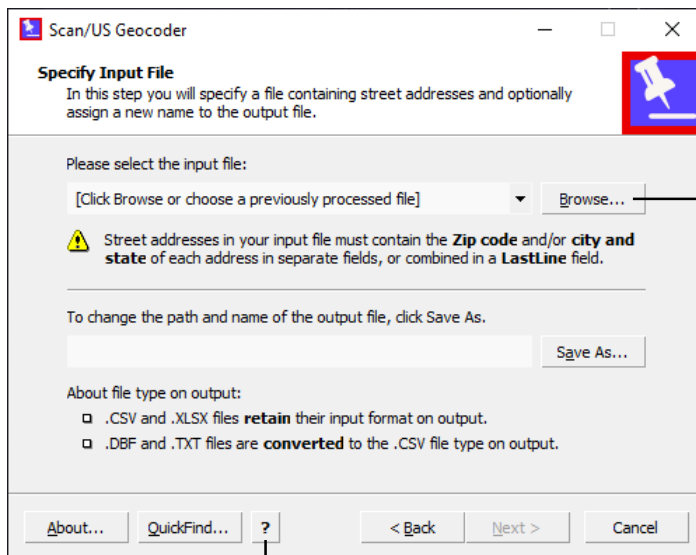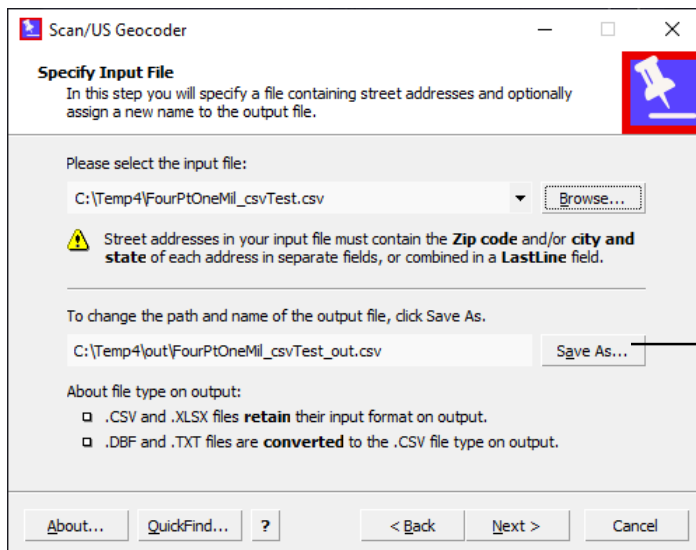

Click this button to open user guide in web browser
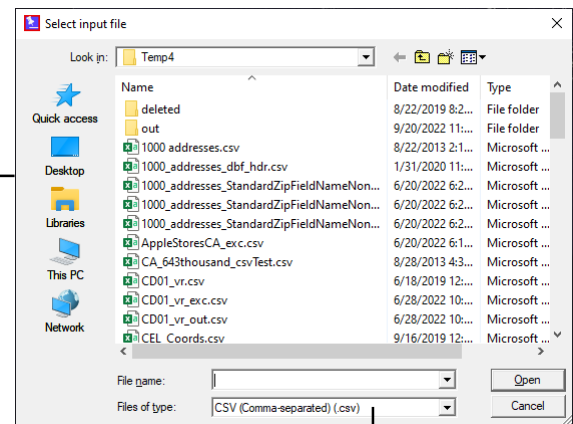
Change file types with this dropdown

Once you've selected a file, the dialog will look like this:

"Output file" is what the geocoder will provide as the end result of the geocoding process—you can rename the output file. Note that your original input file is never overwritten by the geocoder—the geocoder will automatically add "**_out**" at the end of your input file's name for the output file, and that the **output file is written into a separate ..\out subfolder inside the input file's folder**. Click "Save As" to choose a different output file name and/or destination.

CSV, XLSX and XLS input files are written in the original input file's format. TXT and DBF input files are converted to .csv on output.

Click "Next" to proceed.

3. **Identify Input Fields:** match the appropriate fields from your input file to the geocoder's function tags. Tags are used to tell the geocoder where to find specific parts of an address in your file:


The checkmarks in this column indicate that the field will be included in the output file. Uncheck any fields you don't want carried over into the output file.
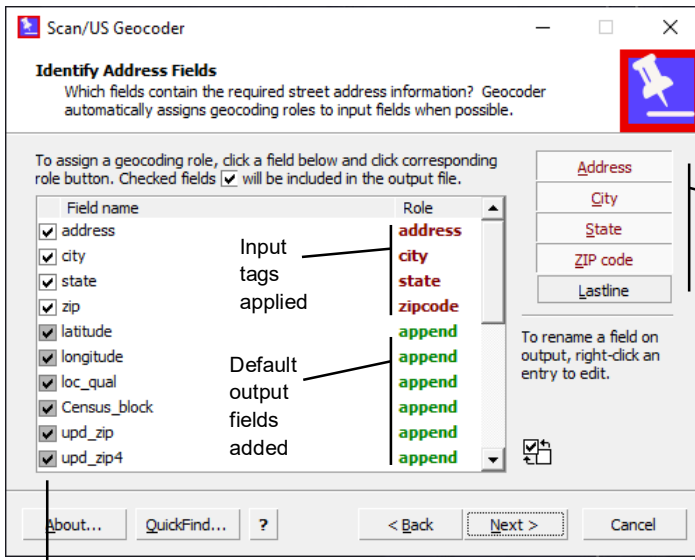
- **Address** tag: assign to the input field that contains the street number and street name.
- **City** tag: assign to the input field that contains the city name.
- **State** tag: assign to the input field that contains the two-letter postal state abbreviation.
- **ZIP_code** tag: assign to the input field that contains the 5-digit or 9-digit postal ZIP code.
- **Lastline** tag (optional): it's possible to put City, State and ZIP code information into a single field. If your file is in this format, then assign the Lastline tag to this field (in this case, don't use the City, State & ZIP_code tags).

The geocoder will scan the field names in your input file and attempt to automatically assign the appropriate tags to the incoming fields. In most cases, the geocoder will successfully assign tags automatically so you probably will not be required to manually assign tags unless you've used very unusual field names.

To assign or reassign a tag, click a field in the list and then click a corresponding tag in the input tag list on the right side of the dialog. If a tag has already been assigned to a field, it will require two clicks on the tag—once to unassign it, and the second click to assign it to the new field.

Default Output Fields: the geocoder has six default output fields—they are listed below. You cannot change or opt out of these default outputs—they are the minimum required outputs. If your input file already contains fields named "latitude" and "longitude," the geocoder will automatically rename those fields to "input_latitude" and "input_longitude" in the output file (to preserve your original input values):

- **Latitutude**: holds the latitude value.
- **Longitude**: holds the longitude value.
- **Loc_Qual** (Location Quality): the Location Quality code tells you how "good" the assigned coordinate is—whether it was an exact address match, a ZIP+4 centroid match, etc. See Appendix C on page 9 for more details.
- **Census_block**: Census Block FIPS code.
- **upd_zip**: updated 5-digit ZIP code.
- **upd_zip4**: updated 4-digit Plus4 code.
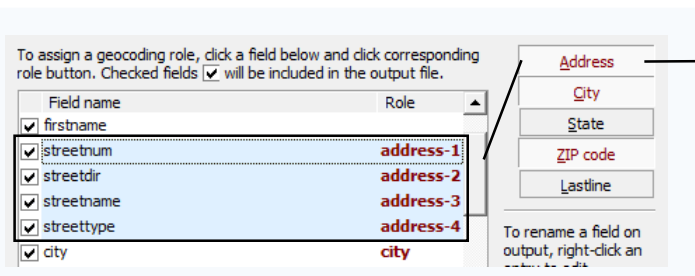
Note: In some address files, the different components of an address field—street number, predirectional, street name, street type and postdirectional—are separated into different fields. You can Control-click to select and highlight multiple fields, then click the "Address" role tag to have the geocoder concatenate the values from those fields to create a standard "Address" field for processing. You must make sure that the fields are present in the correct order in your input file.

If you scroll down further, you will see additional, optional output fields that you can add to your output file:
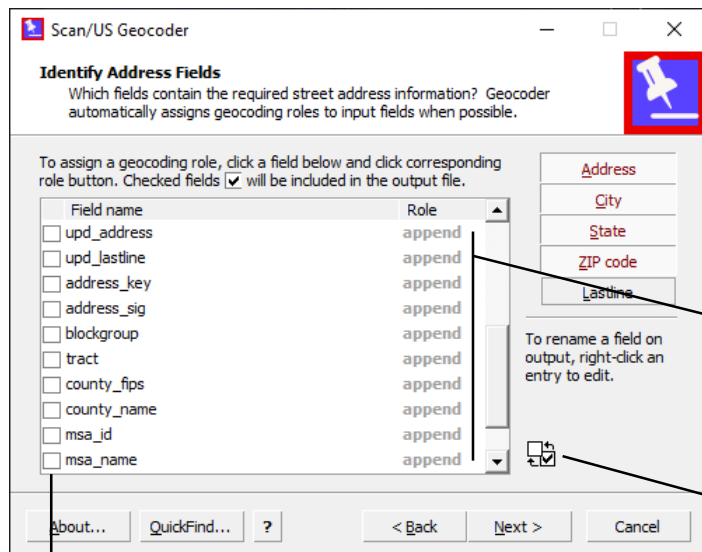

Note: depending on the number of optional output fields chosen, the **size of the output file can increase by 200-700%** from the original input file size. If you are only working with small files—say, a few thousand addresses—the additional storage requirements are minimal. However, if you are processing large files, you should plan for the added storage requirements by making sure that you have enough free hard disk space on the output file destination drive.

For example, the sample input file FourP-tOneMil_csvtest.csv shown in this example is 159MB, but the output file expands to 1.1GB when all of the optional output content is select-ed—**that is nearly 700% of the original size**.

Select/Deselect all optional items at once with a single click with this button

Turn on the checkmarks for these optional output items to include them in the output file
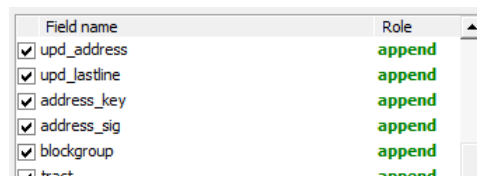
The following are the additional output content you can append to your output file:

- **upd_address**: the corrected street address, such as corrected street name and/or street type—may in-clude corrected spelling, corrected street type, etc.

- **upd_lastline**: the corrected city, state abbreviation and ZIP code—may include corrected spelling.

- **address_key**: a "technical" output that presents the updated address in the format: state-abbreviation|zip|city|street-address. Sorting the file on the address key results in the maximum "localization" of the records; that is, records that pertain to a locale show adjacent in the file.

- **address_sig**: a "technical" output. It contains, along with the state abbreviation and ZIP code a char-acter string with letters representing the class of tokens resulting from parsing the street address. Most users will probably never have a use for this output. However, we may ask you to include this field in case of a coding problem.

- **blockgroup**: Census Blockgroup FIPS code.

- **tract**: Census tract FIPS code.

- **county_fips**: county FIPS code.

- **county_name**: county name fully spelled out.

- **msa_id**: MSA FIPS code.

- **msa_name**: MSA name fully spelled out.

- **microgrid**: Scan/US MicroGrid code.

- **unique_key**: adds database keys—unique identifiers. Examples of such unique identifiers are customer IDs, social security numbers, driver's license numbers, store numbers, etc. If you do not have a unique ID field in the input file, you should add a key field if you plan to map these locations in a GIS.

4. **Processing**: you're ready to code the file. Click "Process" to start coding.
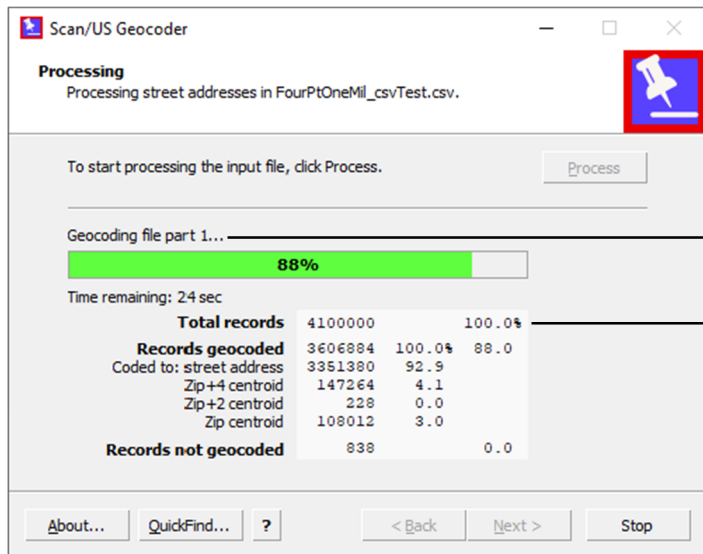
If the file is large, the geocoder will split the file into multiple parts while processing.

During processing, progress statistics such as total number of records processed, number of address matches vs. non-matches, etc. are shown here. If a large file has been split into multiple parts for processing, you will not see a true "Total records" count until the last portion of the file has been reached.

The final summary is also written into a log file named GEOCODER.LOG located inside your "\Documents\Geocoder\Application Logs" folder.

After processing has been completed, press "Close" to exit the geocoder.

§

Some additional details are below:

- While street **intersections** are not considered to be valid deliverable addresses by the USPS, the geocoder can code intersection entries in this format: two street names with their types and directionals joined by "&" or "at" in the address field. You must, of course, also provide the city and state info with the intersections.

  Example 1: Santa Monica Blvd & Ocean Ave, Santa Monica CA
  Example 2: Wilshire Blvd at 20th St, Santa Monica CA

- For intersection matches, the geocoder will only return blockgroup codes or above. For addresses that
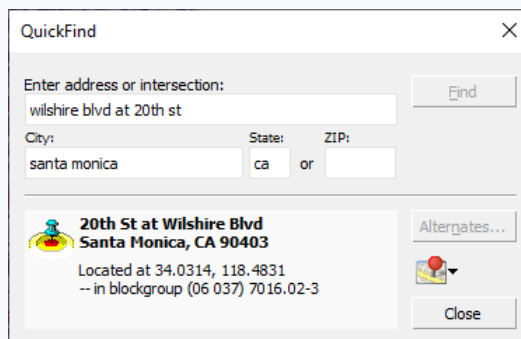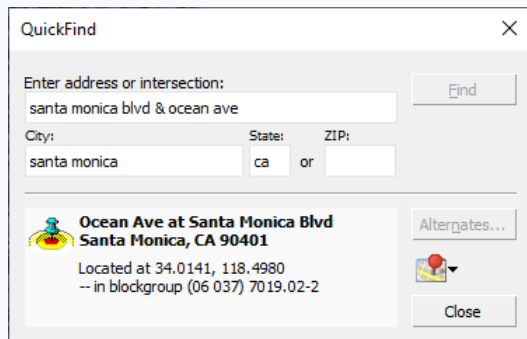
were only matchable to the 5 or 7-digit ZIP code centroid—which only have a confidence level of "within a few miles" of the actual location—the geocoder will return only the county code.

- **Addresses that cannot be coded** will not be included in the standard _out output file. Instead, you will find a separate exception file with **_exc** in the file name in the output destination folder. For example, if your input file is named MyInputFile.csv, then your exception file would be named **MyInput-File_exc.csv**.

- **ZIP-only**: a lot of field-collected addresses tend to be incomplete—with many only containing a ZIP-code. Geocoder now returns a ZIP-centroid coordinate for these incomplete addresses as a last resort.

Appendix A: Address Hygiene (or, "How to get the best match rates")

Maintaining a clean address list is critical to ensuring that every address in your list is assigned a high-quality coordinate. The quality of the addresses in your file determines how well they are geocoded. Scan/US Geocoder can correct most minor misspellings, missing or incorrect directionals and street types, as well as incorrect or missing ZIP codes. However, if an address has an incorrect street number, or a number of errors throughout, the geocoder may not be able to assign a coordinate.

A "clean" address includes the following:

- Primary unit designator, such as a house or building number. Secondary unit designators (such as apartment or suite numbers) may be part of an address, but they are not needed for geocoding.

- Street name spelled correctly

- Street type (e.g St, Ave, Blvd, etc.)

- City name spelled in its entirety.

- Two-letter State abbreviation (e.g. CA, UT, NY, etc.)

- A correct 5-digit ZIP code.

- All necessary pre- and post-directionals (e.g. North, South, East, West, NW, SE etc.) A missing directional may result in incorrect coding.

Important Because Scan/US Geocoder uses the U.S. Postal Service (USPS) directory files and the U.S. Census Bureau's TIGER database (with street names standardized to the USPS names) to match your addresses and geocode them, you can use the USPS guidelines to clean up your addresses. Publication 28, Postal Addressing Standards, is available from your local post office. Or, call 1-800-238-3150 for more information.

Also, when you abbreviate the street type, city or state name, make sure that you follow the USPS guidelines for abbreviations.

File Structure A clean database file also includes appropriately formatted data fields that Scan/US Geocoder can process. A recommended file structure and field descriptions are on the next page (Appendix B).

The quality of the addresses in your file determines how well they are geocoded. Include all the necessary information: (each line represents a field)

Do	Don't
2032 Main St, 2nd Flr	2032, 2nd Flr Main 2032Main
Los Angeles, CA 90025	Los Angl LosAngeles,CA CA900 9025
110 N Main St NE	110 North Main 110 North 110 Main
Plummers Landing, KY 41081	Plummers Landing, 41081 KY41081

A clean file also includes appropriate fields that Scan/US Geocoder can process. In the illustration below, each gray rectangle represents a single field.

Do	Don't
2032 Main St, 2nd Flr	2032 Main St *
2032 Main St 2nd Flr	
Los Angeles CA 90025	Los Angeles
Los Angeles, CA 90025	Los Angeles CA
90025-6119	CA 90025
90025 6119	CA 90025-6119

*Even though this "Don't" example—address components in separated fields—is now supported in this version of the geocoder, we still don't recommend it as a standard practice.

Appendix B: Address Database File Structure

When Scan/US Geocoder processes your file, it looks for input fields that contain your customer addresses, then tries to find a matching address in the U.S. Census Bureau's TIGER street database and the U.S. Postal Service directory files. If a match is found, Scan/US Geocoder fills in output fields with the coordinate and Census ID codes.

Below is a description of the necessary input fields. The field names are the defaults that Scan/US Geocoder looks for. You may use different field names, and then tell the geocoder which fields to process by assigning the appropriate input tags to the correct fields.

Scan/US Geocoder supports the following three formats for your address file:

ASCII files must be delimited with either a comma or pipe | character. The first line of the ASCII file must contain field names separated by a delimiter (a "header" line). Field names must contain only letters, numerals or underscores. All field names should start with a letter, and field names may not contain blank spaces.

The .TXT file extension is expected to be pipe-delimited, and .CSV extension is expected to be comma-delimited.

dBASE/xBASE files must be compatible with dBASE III+ or dBASE IV.

Excel files must be compatible with Excel 97-2003 (.xls) or Excel 2007 or newer (.xlsx). Only the first worksheet in the spreadsheet file will be processed. The first row of the worksheet must contain field names.

File size limits the standard geocoder license limits you to a maximum input file size of 200MB. A premium geocoder license allows unlimited file sizes.

Performance For optimal performance, increase the amount of memory (RAM) in your computer and use an SSD drive. On 32-bit Windows, Geocoder can use up to 2GB of application memory, and on 64-bit Windows, Geocoder can utilize up to 4GB of memory (though a 32-bit app, it is "large address-space aware").

Input Fields	Type	Length**
Address	Character	60
City	Character	30
State	Character	2
ZIP_code	**Character**	5
Plus4	**Character**	4

**Your input file's field lengths may be different. These lengths are recommended. Also, the ZIP_Code and Plus4 fields may be combined into a single field that is at least 9 characters long. Or, you may combine City, State, ZIP and ZIP+4 information into a single "Lastline" field at least 42 characters long.

Please note that **ZIP_Code** and **Plus4** fields are defined as **Character field types** even though their contents are composed solely of numerals—this is to prevent the loss of leading zeroes for those ZIP codes that start with zeroes, such as Connecticut ZIPs.

The advice about field type and field length mostly only applies to dBASE (.dbf) files since ASCII-delimited files and Excel files are variable-length, freeform files. **Excel users**, however, must be aware that normal Excel behavior removes leading zeroes in cell values unless the cells are formatted as character—*if you are not aware of this, you will definitely lose leading zeroes from many East Coast ZIP codes*.

Appendix C: Location Quality codes

The 4 or 5-character Location Quality codes—found in the "Loc_Qual" field of output file—show a more detailed summary of the methods used to arrive at a match.

Here are a few example codes:

AS00 - Perfect address match.

AX103 - Very close intersection match where the second street's qualifiers were imperfectly matched.

ZB94 - Address was matched to a USPS range without a TIGER match, so Zip9 centroid was used. There is a block code present. In the input, a ZIP was present but it was not used or the place name was spelled incorrectly, and the base name was spelled incorrectly.

Below is a listing of the meaning carried by each of the characters (bytes) in the quality code:

First byte:

A - address range segment or intersection match.

Z - zip5, zip7 or zip9 centroid. Sometimes addresses will match a USPS range that has no TIGER match. In that case the zip9 centroid will be used if available.

E - no match.

If first byte is A, the second byte represents the street range match:

S - perfect match of street number within street range.

Z - perfect match of street number within street range, but segment had zero distance (single point) so no setback could be made.

I - an exact match could not be found, but a range within 200 of the address number was found. Also, although the exact match was not found, the address was within the range of the lowest to the highest addresses for the street.

N - an exact match could not be found, but a range within 200 of the address number was found. The address was outside the range of the lowest to the highest addresses for the street.

O - an exact match could not be found, but a range with opposite parity (even if the address was odd and vice versa) was found.

T - an exact match was found when the first two numbers of the address were reversed.

X - represents an intersection.

If the first byte is A, the third byte represents the quality of the match:

0 - perfect match (internal match score of 1.0).

1 - very close match (internal match score of .95 or greater).

2 - close match (internal match score of .9 or higher).

3 - relaxed match (internal match score of .8 or higher).

Appendix C: Location Quality codes   (…continued from page 9)

If the first byte is A or Z, the fourth byte represents the reasons for an imperfect match due to something other than the range match (in the case of an intersection - AX - the fourth byte is for the first street and the fifth byte is for the second street):

0 - perfect match.

1 - ZIP code was present but a better match was found using the place name, or the place name was spelled incorrectly.

2 - Base name was misspelled.

3 - Qualifiers were imperfectly matched.

4 - 1 & 2.

5 - 1 & 3.

6 - 2 & 3.

7 - 1, 2 & 3.

If the first byte is Z, the second byte represents the type of census code:

B - block (for zip9 centroid matches).

C - county code (for zip5 and zip7 centroid matches).

If the first byte is Z, the third byte represents the centroid type:

5 - five digit zip centroid.

7 - seven digit zip centroid.

**9** - nine digit zip centroid.

If the first byte is E, the second byte represents the reason for no match occurring:

**1** - address not found.

2 - lastline/ZIP not found.

If an address generates a multiple match possibility, the first possible match will be used. An "M" will be appended to the end of the status code when multiple matches were found.

Appendix D: Error messages

These are the most common error messages you will see:
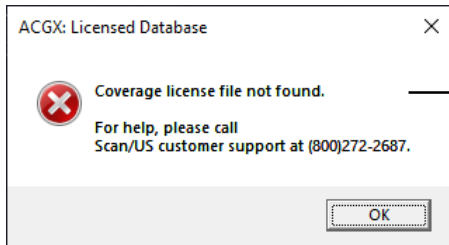
When you buy a license for the Scan/US Geocoder, you are given two license file installations. The geocoder expects to find two license files. **ACGX.lic** is a **coverage license file** that contains all the states you have licensed—either national, custom 8 states, or a single state. This error means that you have not yet installed the coverage license.
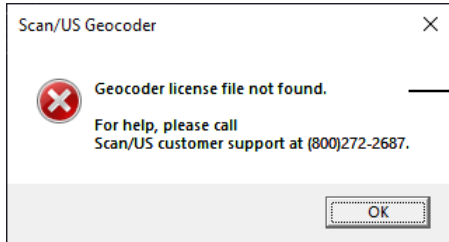
When you buy a license for the Scan/US Geocoder, you are given two license file installations. The geocoder expects to find two license files. **Geocoder.lic** is an **expiry license file** that contains your license expiration date. This error means that you have not yet installed the expiry license.
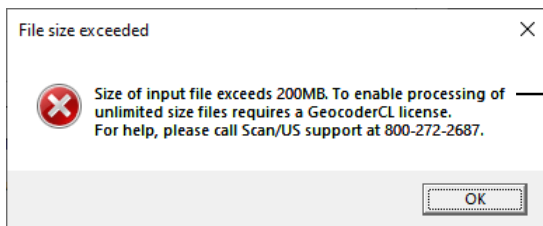
The standard geocoder license file limits you to a maximum input file size of 200MB. A premium geocoder license will remove this file size limit. Please call 800-272-2687.

Remember the 4.1 million record example file shown in the instruction pages? That file is only 159MB, so it would be pretty rare for most users to see this message—only high-volume users would ever see this error.

Scan/US Tech Support: (800)272-2687; 9:00 AM—5:00 PM, Pacific Time; Monday through Friday, excluding holidays