

Scan/US™

Geocoder

User's Guide

Version 8.x



Table of Contents

Using Scan/US Geocoder	3
Geocoder Setup Options.....	6
Appendix A: Address Hygiene.....	8
Appendix B: Address Database File Structure.....	9
Appendix C: Location Quality Codes.....	10
Appendix D: Location Status Codes.....	11



Scan/US, Inc.
Sales or Support: 800-272-2687 or 310-828-1450
6:30 AM-5:00 PM, Pacific time, Monday-Friday, excluding holidays
E-mail: sales@scanus.com or support@scanus.com
<http://www.scanus.com>

Eighth Edition

March 2009

Copyright

© Scan/US, Inc., 2001-2009. All rights reserved. No part of this publication may be reproduced, transmitted, transcribed, stored in a retrieval system, or translated into any language in any form by any means without the written permission of Scan/US, Inc.

Software License Notice

Your license agreement with Scan/US, Inc., which is included with the product, specifies the permitted and prohibited uses of the product. Any unauthorized duplication or use of Scan/US Geocoder in whole or in part, in print, or in any other storage and retrieval system is forbidden.

Licenses and Trademarks

Scan/US and the Scan/US logo are trademarks of Scan/US, Inc. Other company or product names used herein are for identification purposes only and may be trademarks of their respective companies.

ZIP Code, ZIP+4 are registered trademarks of the U.S. Postal Service.

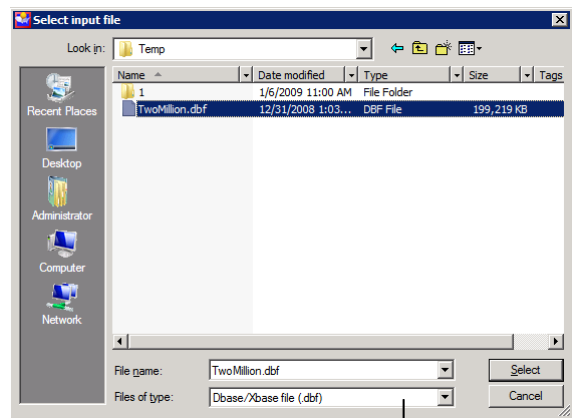
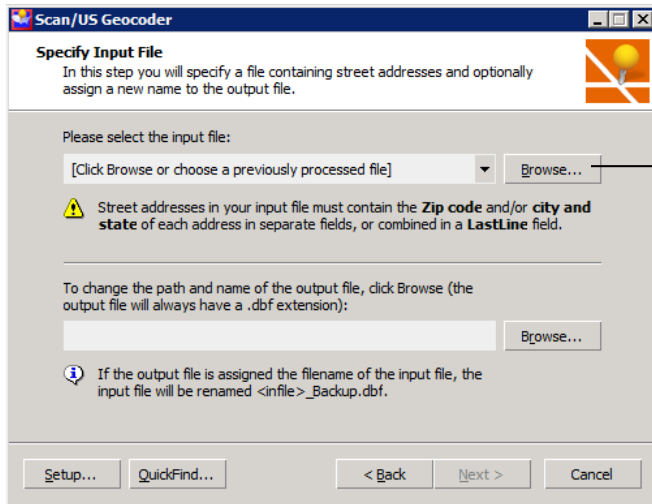
General Notice

Any use of other company or product names herein are for illustration purposes only and does not constitute an endorsement of the company or product by Scan/US, Inc. Scan/US, Inc. is not responsible for any problems and/or losses resulting from the use of any product(s) mentioned in this guide.

Using Scan/US Geocoder: the Scan/US Geocoder is an 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.

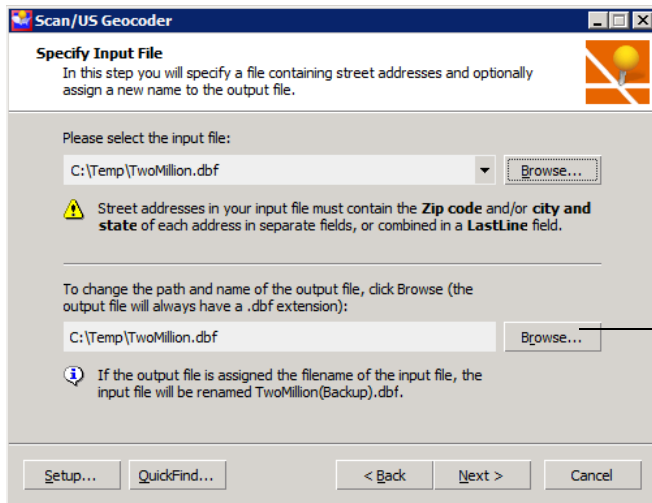
1. Geocoder Start Page: the welcome screen provides an overview of the address geocoding steps. The geocoder is designed as a wizard which will take you through the process step-by-step. Click “Next” to proceed to the input selection page.

2. Specify Input File: click “Browse” to select the address file to process. Geocoder supports dBASE, ASCII delimited-text, Excel, and MapInfo file formats. Use "Files of type" dropdown to change file types.



Change file types with this dropdown.

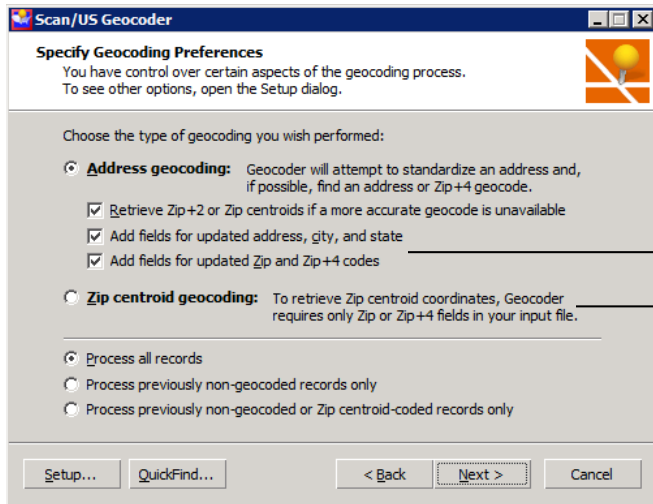
Once you’ve selected a file, the dialog will look like this:



“Output file” is what the geocoder will provide as the *end result* of the geocoding process—you can rename the output file. Click “Browse” to choose a different output file name and/or destination. All output files are written in dBASE format (.DBF) regardless of input file type. Your original input file is never overwritten by the geocoder. If the input is a dBASE file, it will be backed up to a new name—the string “(Backup)” is appended to the original file name.

Click “Next” to proceed.

3. Geocoding Preferences: choose various processing options. For additional coding options, click the “Setup” button (see page 6).

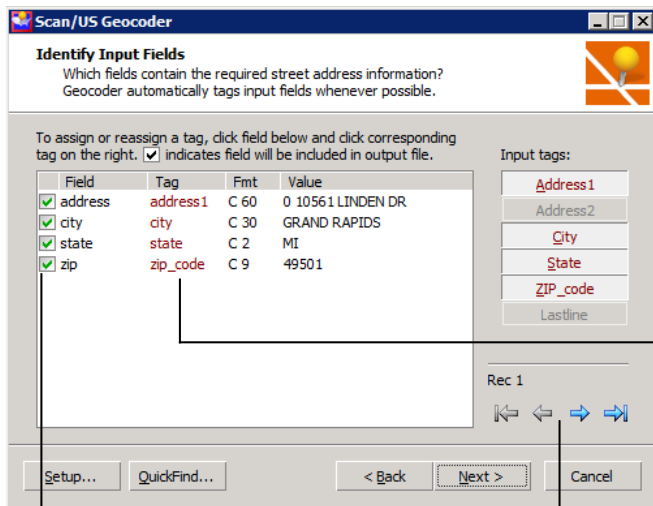


The geocoder is capable of detecting and correcting minor spelling mistakes and other errors commonly found in address files. These two “Add fields” options will trigger the geocoder to **append new fields** into the output file to hold corrected or standardized address and ZIP code information. The geocoder will not overwrite your original address/ZIP data in the file.

The “ZIP geocoding” option tells the geocoder to **ignore** street address information and only consider the ZIP code information. The geocoder will return only the 5, 7 or 9-digit ZIP centroid coordinates—whichever is found in your incoming ZIP code field.

Note: depending on the number of coding options chosen, the size of the output file can increase by 200-300% from the original input file size. Also note that the geocoder never overwrites the input file—instead, it backs up the input file and creates a *new file* to hold the output, thus at least doubling the storage requirement. 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.

4. 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:



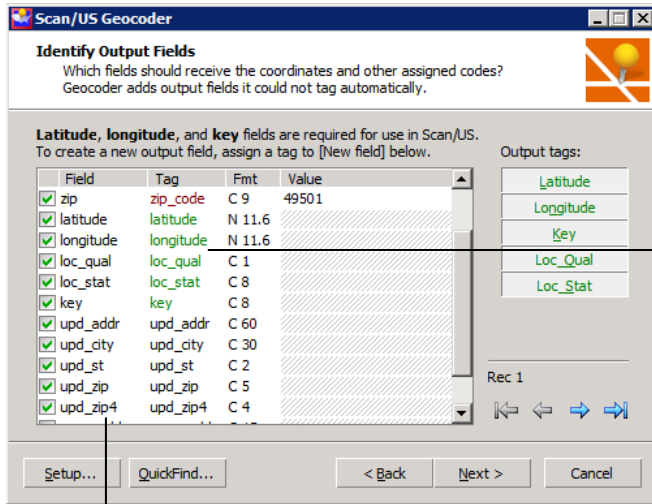
- **Address1** tag: assign to the input field that contains the street number and street name.
- **Address2** tag (optional): assign to the input field that contains miscellaneous address information such as suite or apartment number.
- **City** tag: assign to the input field that contains the city name.
- **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 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.

You can browse through the actual values in your database by using these arrow buttons to go forward and backward through the file.

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.

5. Identify Output Fields: assign tags to the geocoder’s outputs, such as *latitude* and *longitude*. Again, tags are used to tell the geocoder where to put the geocoder’s outputs. We recommend that you allow the geocoder to create all of the necessary output fields—thus ensuring correct formatting for all required fields. However, if your input file already contains one or more of the output fields, the assignments are as follows:

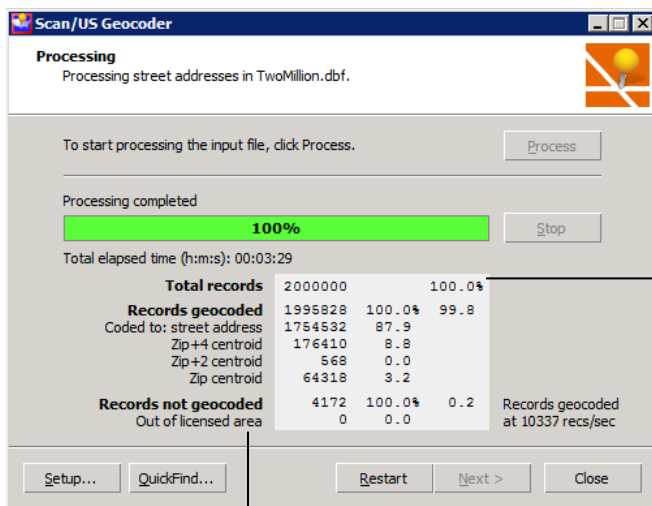


- **Latitude** tag: assign to field that is designed to hold the latitude value.
- **Longitude** tag: assign to field that is designed to hold the longitude value.
- **Loc_Qual** tag (Location Quality): the Location Quality code tells you how “good” the assigned coordinate is—whether it was an address match, a ZIP+4 centroid match, etc. See page 10 for more info.
- **Loc_Stat** tag (Location Status): the Location Status codes tell you how a match was made—whether elements of an address were corrected to make a match, or if no match, the type of error found. See details on page 11.
- **Key** tag: assign to the field that is designed to hold the 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, the geocoder will automatically add a key field to the output.

These new “upd_” fields will contain the corrected address and ZIP code information. This feature is only activated when you turn on the “Add fields” options in the Geocoding Preferences page.

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. If your input file does not contain the necessary fields, the geocoder will automatically append the required fields (recommended). To assign or reassign a tag, click a field in the list and then click a corresponding tag in the output 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.

6. Processing: you’re ready to code the file. Click “Process” to start coding.



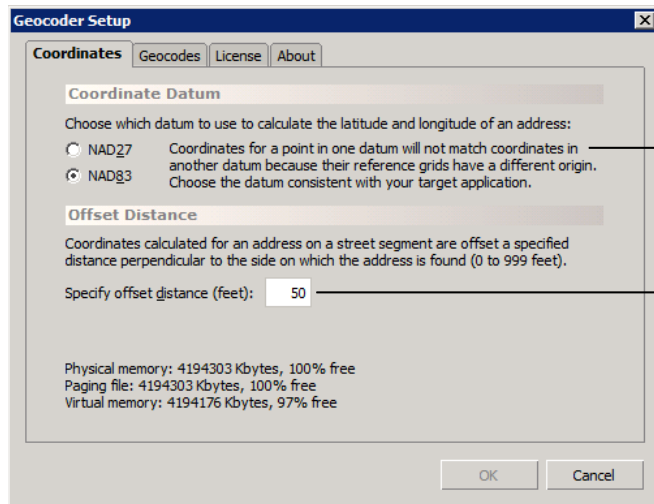
During processing, progress statistics such as total number of records processed, number of address matches vs. non-matches, etc. are shown here. The final summary is also written into a log file named GEOCODER.LOG located inside the “My Documents\Geocoder\Application Logs” folder.

The geocoder is sold with one of three types of license coverages—national, 1-state or 8-states. This “Out of licensed area” statistic shows how many addresses were outside of the coverage area.

After processing has been completed, you may press “Restart” to process another file or press “Close” to exit the geocoder.

Geocoder Setup Options: choose various processing options.

1. Coordinates: coordinate datum & building offset distance.



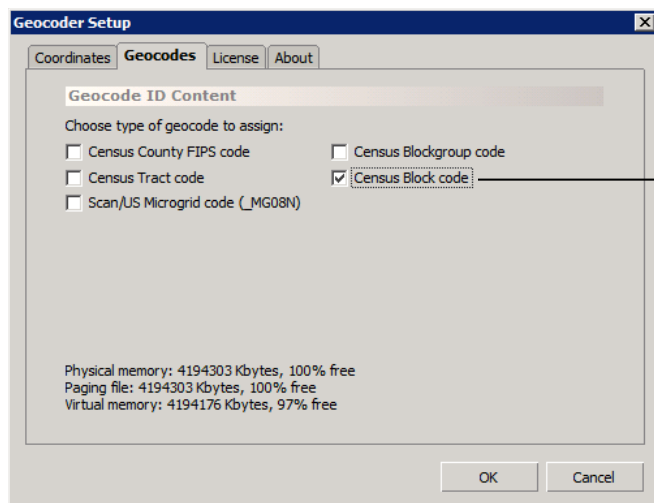
The default setting is NAD83. Scan/US for Windows, Ver. 8.x requires NAD83 coordinates. If you are using Scan/US version 7.6 or older, you will need to switch this setting to NAD27.

If you set the offset distance at zero feet, you'll essentially be putting the building right in the middle of the street. The greater the offset distance, the farther from the road's centerline you'll be placing the building. The default distance of 50 feet works well for most needs.

About Coordinate Datum: for locating objects in North America, our lat/lon coordinates are placed on a reference grid which is *calibrated* to a set of survey points gathered over the last century. This set of survey points is referred to as a “Datum”—specifically, “North American Datum 1927 (NAD27)” or “North American Datum 1983 (NAD83).” NAD83—used by all GPS devices—is a newer version of the datum which uses many more survey points, including satellite surveys.

Differences between NAD27 and NAD83 coordinates can be significant—in New England you can have a “NAD-shift” of up to 40 meters, and in California, the shift can be as much as 100 meters. Therefore, selecting the correct datum compatible with your mapping software is important.

2. Geocodes: select the geography code type to be appended to each matched address.

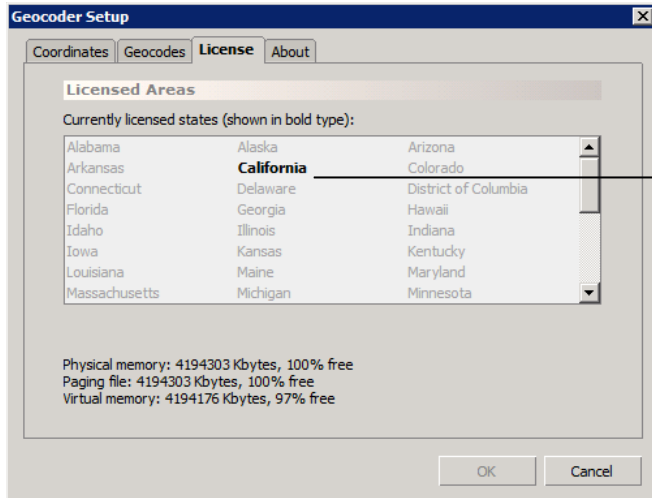


You may select any or all of the identifier types available on this page.

Please note, however, that each additional selection will **add another field** into your output file, therefore increasing the size of the final output: depending on the size of your input file, you'll need to take this under consideration if you are running on a PC with limited memory (RAM) since the entire output file needs to be held in memory during processing. On a memory-starved PC, performance will be severely affected if the machine runs out of physical memory and has to swap out to the system page file.

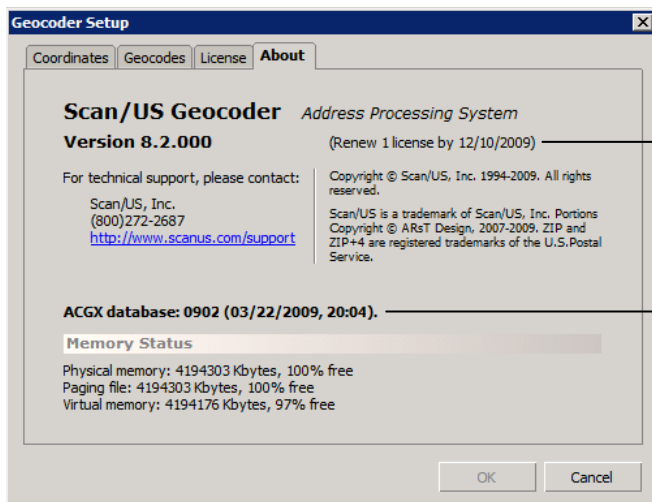
For intersection matches, the geocoder will only return blockgroup codes or above. For addresses that were only matchable to the 5 or 7-digit ZIP code centroid—which only have a confidence level of “within a few miles” of the actual location—the geocoder will return only the county code.

3. License: shows the list of licensed states in your current geocoder license file.



This list shows the content of your license file—your license’s coverage. The licensed states are shown in bold type. The grayed-out states are those that you have not yet licensed. The geocoder will process only those addresses covered by your license.

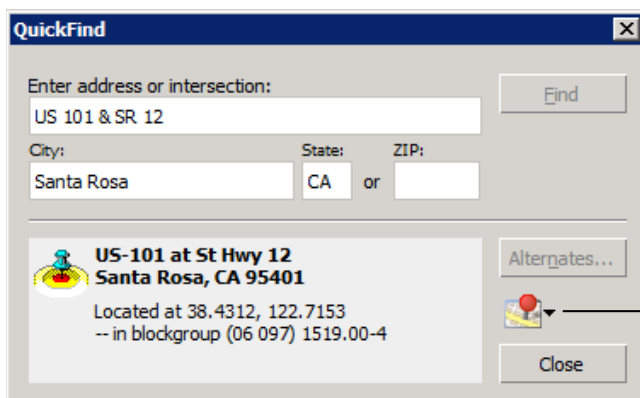
4. About: version number, license count/expiration date, tech support link and phone number, etc. You may be asked to provide some of this information during a technical support call, such as version number and ACGX database date.



Your license expiration date will appear here, as well as total number of licensed users.

This entry shows the version number and date of the geocoding reference database in your configuration.

Tip: The **QuickFind...** button—found in the geocoder’s standard wizard pages—provides a convenient way to look up a single address & display it on a web browser map (such as Google Maps).



Click this button to show the found location in Google Maps, Google Earth, Live Maps or MapQuest: requires an internet connection.

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, Scan/US 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:

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

Main St & 10th Ave*

Wilshire Blvd at 20th St*

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 or 2032 Main St. 2nd Flr	2032 Main St.
Los Angeles CA 90025 or Los Angeles, CA 90025	Los Angeles Los Angeles CA
90025-6119 or 90025 6119	CA 90025 CA 90025-6119

*While street intersections are not considered to be valid deliverable addresses by the USPS, the Scan/US Geocoder can geocode 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.

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 and output fields. You may add the required output fields using the appropriate application. Or, you can let the geocoder add them automatically (recommended). The field names are the defaults that Scan/US Geocoder looks for. You may use different names, and then before processing, tell Scan/US Geocoder which fields to process by assigning the appropriate input tags to the correct fields.

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

ASCII files must be delimited with *either* a comma or tab 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, and all field names should start with a letter & be no more than 10 characters long (with no blank spaces).

The .TXT file extension is expected to be Tab-delimited, and .CSV extension is expected to be Comma-delimited.

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

MapInfo .dat/.tab files.

Excel files must be compatible with Excel 97-2003 (.xls) or Excel 2007 (.xlsx). Only the first worksheet in the spreadsheet file will be processed.

File size limits Theoretically, dBASE records may contain up to 255 fields, and maximum number of records per table is one billion. Practically speaking, you are more likely to run out of memory long before you hit the theoretical maximum limits.

Performance For optimal performance, increase the amount of memory (RAM) in your computer. 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 “Last-line” 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.

Output Fields	Type	Length
Latitude**	Numeric	11 (6 decimals)
Longitude**	Numeric	11 (6 decimals)
Census_ID	Character	19
Loc_Qual**	Character	1
Loc_Stat**	Character	8

**Do not change these default names.

It is not necessary for you to have the output fields already defined in your input file. The geocoder will automatically detect their presence or absence, and it will create any necessary output fields it needs. In fact, we recommend that you allow the geocoder to create all of the required output fields—thus ensuring correct formatting for all required fields.

In addition to the above-listed output fields, the geocoder will create a “Key” field for unique identifiers if you do not already have one in your input file. Unique identifiers are required by all database programs, such as Scan/US for Windows.

Appendix C: Location Quality codes

The 1-character Location Quality codes—found in the “Loc_Qual” field of output file—show the type of match returned by the geocoder. There are six possible codes:

- **L**: Out of licensed area. You attempted to geocode an address outside of your license coverage.
- **E**: Error. Address could not be matched to anything in the reference databases.
- **0**: Address match. This is the highest quality match.
- **1**: 9-digit ZIP centroid match.
- **2**: 7-digit ZIP centroid match.
- **3**: 5-digit ZIP centroid match.

The two “process previously non-geocoded records” options on the “Geocoding Preferences” page make use of these codes to determine whether or not a particular address has been previously geocoded.

Appendix D: Location Status codes

The 4 or 5-character Location Status codes—found in the “Loc_Stat” 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 status 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).

(...continued)

Appendix D: Location Status codes (...continued from page 11)

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.



Scan/US Tech. Support: (800) 272-2687; 6:30 AM - 5:00 PM, Pacific Time; Monday through Friday, excluding holidays

