pitney bowes

Location Intelligence

Centrus® User Dictionary Writer

Version 1.0

User's Guide for Windows

April 2017

© 2017 Pitney Bowes Software Inc. All rights reserved.

Pitney Bowes Software Inc. is a wholly-owned subsidiary of Pitney Bowes Inc. Pitney Bowes, the Corporate logo, Centrus and “Every connection is a new opportunity” are trademarks of Pitney Bowes Inc. or a subsidiary. All other trademarks are the property of the respective owners.

The following trademarks are owned by the United States Postal Service®: CASS, CASS Certified, DPV, eLOT, FASTforward, First-Class Mail, Intelligent Mail, LACSLink, NCOALink, PAVE, PLANET Code, Postal Service, POSTNET, Post Office, RDI, SuiteLink, United States Postal Service, Standard Mail, United States Post Office, USPS, ZIP Code, and ZIP+4. This list is not exhaustive of the trademarks belonging to the Postal Service.

USPS Notice: Pitney Bowes Software Inc. holds a nonexclusive license to publish and sell ZIP+4 databases on optical and magnetic media. The price of the Pitney Bowes Software Inc product is neither established, controlled, nor approved by the U.S. Postal Service.

Pitney Bowes Software is a non-exclusive licensee of USPS® for NCOALink® processing. Prices for the Pitney Bowes Software products, options and services are not established, controlled or approved by USPS® or United States Government. When utilizing RDI™ data to determine parcel-shipping costs, the business decision on which parcel delivery company to use is not made by the USPS® or United States Government.

AD# 12.07

Centrus data products contained on this media and used within Centrus applications are protected by various trademarks and by one or more of the following copyrights:

Copyright © United States Postal Service. All rights reserved.

© 2017 TomTom. All rights reserved. This material is proprietary and the subject of copyright protection and other intellectual property rights owned by or licensed to TomTom or its suppliers. The use of this material is subject to the terms of a license agreement. Any unauthorized copying or disclosure of this material will lead to criminal and civil liabilities.

© 2017 HERE

Copyright © United States Census Bureau

© 2017 Claritas, Inc.

Table of Contents

1 - Centrus User Dictionary Introduction

Purpose of this guide 3
Contacting Technical Support 3

2 - Installing the Centrus User Dictionary Writer

Installation requirements 5
Verifying Java path 5
Installing UD Writer 5
Configuring the USA_DataManagerSettings.properties File 6

3 - Creating User Dictionaries

Understanding User Dictionary capabilities and requirements 8
Source data requirements 8
Required input fields 9
Optional (recommended) input fields 9
User Dictionary file names and formats 10
Additional User Dictionary considerations 11
Using User Dictionaries with address point interpolation 13

4 - Using the Centrus User Dictionary Writer

Running UD Writer 15

1 – Centrus User Dictionary Introduction

This chapter discusses the purpose and use of this guide, how its conventions are presented, and how to obtain assistance from Pitney Bowes.

In this chapter

Purpose of this guide 3
Contacting Technical Support 3

Purpose of this guide

Welcome to the User Dictionary Writer (UD Writer). This guide provides information on installing and using the UD Writer.

Contacting Technical Support

Pitney Bowes Technical Support Representatives work closely with your company's personnel to help you use your Pitney Bowes products. If you are having difficulty during the installation, before you contact Technical Support make a note of exactly what you are trying to accomplish as well as any error messages the system displays. Reporting complete details will enable the technical support representative to pinpoint the problem and resolve it quickly. Contact Technical Support via the web, email or phone for assistance:

- Create a case in our Online Case Management System at: http://go.pbinsight.com/online-case-management.
- The Technical Support e-mail address is software.support@pb.com.
- The Technical Support hotline number is 1-800-367-6950.

2 – Installing the Centrus User Dictionary Writer

In this chapter

Installation requirements	5
Verifying Java path	5
Installing UD Writer	5
Configuring the USA_DataManagerSettings.properties File	6

Installation requirements

To ensure proper installation, verify that you meet the following requirement:

- Java 1.7 or higher

Verifying Java path

Before installing the UD Writer, the correct Java path must be set.

1. Locate the **Environment Variables** dialog. The navigation to locate this dialog may be different depending on your version of the Windows operating system.
   On a Windows 8 system, this path would be:
   System (Control Panel) > Advanced system settings > Environment Variables
2. Under **System Variables,** check to see if the **JAVA_HOME** variable is defined; if not, click **New**.
3. On the **New System Variable** window, in the **Variable name** field, enter **JAVA_HOME**.

   In the **Variable value** field, enter the path to your Java installation, for example:
   C:\Program Files\Java\jdk1.8.0_101.
4. Click **OK** as prompted.
5. You must close and re-open all command windows before the JAVA_HOME setting takes effect

Installing UD Writer

1. Locate the **setup.exe** file in the UDWRITER folder. To run the install as an administrator, right-click on **setup.exe** and select "**Run as administrator**".
2. The UD Writer install wizard appears and will walk you through the installation process.

Note: A notice may open informing you that you need to verify that the USA_DataPath Properties file is configured properly.

Note: A JAVA_HOME window may open as well informing you of the selected Java path.

Configuring the USA_DataManagerSettings.properties File

Locate the USA_DataManagerSettings.properties file in the `/UserDictionaryWriter/config` folder.

By default the USA_DataManagerSettings.properties file appears as follows:

```
# Properties used by MapMarker UDCreator GUI
#
# Optional - The number of dictionaries to be loaded.  DEFAULT=1
DICTIONARY_COUNT=1

# Required - The path to the highest ranking dictionary (GeoStan data files)
# Note that DICTIONARY_PATH is required from 1 to DICTIONARY_COUNT.

# This information must be updated for your installation.  It specifies the
# location of the GeoStan data (using forward slashes).

DICTIONARY_PATH1=C:/Program Files (x86)/Centrus/data

# If us.z9 file is in a different folder than DICTIONARY_PATH1 then uncomment the
# following line and update to its path (using forward slashes).
# In addition, change DICTIONARY_COUNT above to 2.
#DICTIONARY_PATH2=C:/Program Files (x86)/Centrus/z9data


# Enter license file path (using forward slashes) and password

LICENSE_FILE=
PASSWORD=

########################################
#
#  These items should be left unchanged
#
LIB_PATH=dlls
#
# Option to return all GeoStan data elements in USA_UserCandidateAddress class
RETURN_EXTENDED_DATA=yes
#
# This optional property controls the number of GeoStan instances
# used to process requests in a multi-threaded environment.
#
GEOSTAN_INSTANCE_COUNT=1
```

You must edit the USA_Datamanager.properties file with the following information:

```
DICTIONARY_PATH1=//<path to US geocoder data files>
Z9_FILE=//<path to US geocoder data files>/us.z9
LICENSE_FILE=//<path to license file>
PASSWORD= <your password>
```

3 – Creating User Dictionaries

This chapter includes information on creating User Dictionaries, source data requirements and required fields, and other information specific to working with User Dictionaries.

Note: User Dictionaries are not for use with CASS geocoding.

In this chapter

Understanding User Dictionary capabilities and requirements	8
Source data requirements	8
Required input fields	9
Optional (recommended) input fields	9
User Dictionary file names and formats	10
Additional User Dictionary considerations	11
Using User Dictionaries with address point interpolation	13

Understanding User Dictionary capabilities and requirements

The capabilities of User Dictionaries and the basic requirements for creating them are as follows.

- All fields supported by normal street geocoding can be included in User Dictionaries.
- Landmarks and place names are supported in User Dictionaries. Postal or geographic centroid geocoding are not supported in User Dictionaries.
- User Dictionaries support address browsing using partial street names or landmarks and place names.
- GSDs are necessary to create the User Dictionary. This is because the GSDs have some internal structure that must be available when creating a User Dictionary.

The results from a User Dictionary are similar to that from the GSD. For address matches where the first letter of the match code would be 'S', a User Dictionary match has the letter 'J'.

Note: For GeoStan, the value of the GS_REC_TYPE is 'U'. Also, the enum GS_DATATYPE returns a new value for the User Dictionary record matches, see User Dictionary (GsFileStatusEx) for more information.

For example: SE9 is a match code for a match that comes from a GSD, while JE9 is for a match that comes from a User Dictionary. For more information, see the match code descriptions in your product manual.

Source data requirements

The source data for User Dictionaries includes street data but can also include place names and intersections.

To create a User Dictionary, your source data must conform to the following requirements:

- Source records must include required fields, and these fields are mapped during the User Dictionary creation process. If a value of a required field is empty for a particular record, then that record will not be imported into the User Dictionary. The MapInfo® table must contain specific fields. These input fields are described in “Required input fields” on page 9.
- Source records must be in a MapInfo table (TAB file).
- Segments must have two or more defined end points to be loaded into a User Dictionary. Segments without end points are ignored.
- Segments that make up intersections must have one or more end points in the intersection for the US geocoder to recognize it as an intersection. Source records can be either point objects or segments.
- Each row in the table is equivalent to a street segment.

Required input fields

You must specify the field names in the MapInfo table (TAB file) in order for the table to be translated into a User Dictionary. Certain fields are required and must be present in the MapInfo table. Other fields are optional, but are strongly recommended because there may be negative consequences if they are omitted. This is described in “Optional (recommended) input fields” on page 9. If any of the required fields are missing, a missing field error code is returned.

The following table describes the required input fields.

Required fields	Description	Maximum field length
Left start address	Start of address range on left side of street.	10
Right start address	Start of address range on right side of street.	10
Left end address	End of address range on left side of street.	10
Right end address	End of address range on right side of street.	10
Street name	Name of street.	30
State abbreviation	Two-character state abbreviation.	2
Left ZIP Code	ZIP Code for left side of street.	5
Right ZIP Code	ZIP Code for the right side of the street.	5

Optional (recommended) input fields

The Left and Right Odd/Even Indicator fields are used to specify whether the sides of the street segment contain odd or even address ranges. Although these indicators are not required for creating a User Dictionary, it is important to use the Odd/Even Indicators when your data contains odd/even address numbers.

When the Odd/Even Indicator is specified, but is inconsistent with address numbers, the indicator is set to Both.

When the Odd/Even Indicator is not specified and both Start Address and End Address have values, the indicator is set to Both, unless the start and end address numbers are the same number. In that case, the indicator is set to Odd if the address numbers are odd, and set to Even if the address numbers are even.

When the Odd/Even Indicator is not specified and both Start Address and End Address have values, the indicator is set to Both (odd and even).

Note: If your table contains Odd/Even indicator information, we strongly recommend that you use the Odd/Even indicator fields. These fields ensure that your geocoded addresses are located on the correct side of the street. Omitting the fields when your data contains Odd/Even information may produce incorrect results.

The following table describes the optional input fields.

Optional fields	Description	Maximum field length
Left Odd/Even indicator*	Left side of the street contains only odd or even address ranges (O=odd, E=even, B=both)	1
Right Odd/Even indicator*	Right side of the street contains only odd or even address ranges (O=odd, E=even, B=both)	1
City*	City name	28
Left ZIP+4 Code	4-digit ZIP+4 add-on for left side of street.	4
Right ZIP+4 Code	4-digit ZIP+4 add-on for right side of street.	4
Left Census Block	Census Block ID for left side of street	15
Right Census Block	Census Block ID for right side of street	15
Place Name	Place name	40

* These fields are highly recommended.

User Dictionary file names and formats

The following list specifies file name and formatting requirements:

- Each User Dictionary has a base name of eight characters or fewer.
- Each User Dictionary resides in its own directory.
- The maximum length of a path to a User Dictionary is 1024 characters.
- The ZIP Code range in the MapInfo table for a User Dictionary is unlimited.

Because each User Dictionary resides in its own directory, User Dictionaries may share the same name. However, it is generally good practice to use a unique name for each User Dictionary.

Some of the output files are tied to the base name. The other output files have constant names. For example, the output files for a dictionary called ud1 are the following:

```
postinfo.jdr
postinfo.jdx
lastline.jdr
post2sac.mmj
geo2sac.mmj
sac2fn_ud.mmj
ud1.jdr
ud1.jdx
ud1.bdx
```

If your data includes place names, the dictionary contains the following files:

```
ud1.pdx
ud1.pbx
```

The dictionary also contains these log files:

```
ud1.log
ud1.err
```

Additional User Dictionary considerations

See the following topics for more information when working with User Dictionaries.

Data Access License

You must still have a valid access license to the data contained in the GSD when you are geocoding against your User Dictionary. For example, if you create a dictionary of New York streets and addresses, you must purchase the New York or entire U.S. GSD.

Use without GSD data files

To utilize a User Dictionary without the use of GSDs, the files listed below are required:

- ctyst.dir - The USPS City State table
- parse.dir - Address elements and abbreviations dictionary

To perform postal centroid geocoding, in addition to a GSD or a User Dictionary and the files listed above, the following files are necessary:

- us.z9 - Postal centroid information
- cbsac.dir - Required only if county names or CBSA/CSA data are needed

CASS standards

You cannot geocode to CASS standards using a User Dictionary. This also means that the ParcelPrecision Dictionary cannot be used during CASS geocoding.

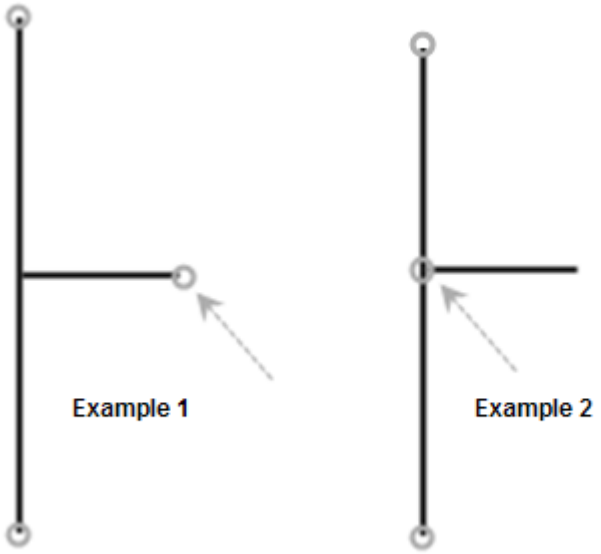
Address Range Order

The US geocoder determines the order of the address range based on a comparison of the start and end addresses. The comparison produces the following results:

- If the end is greater than the start, the range is ascending.
- If the start is greater than the end, the range is descending.
- If the start is equal to the end, the range is ascending.

Street intersections and User Dictionaries

When geocoding to street intersections with a User Dictionary, the US geocoder cannot recognize the intersections if one or more of the segments that make up the intersection does not have an end point at the intersection. This can happen when you create the User Dictionary from a customized street table in which some segments that terminate at intersections do not have end points (Example 1).



Example 1: Intersection in User Dictionary does not have end points for all segments; this will not be recognized by the US geocoder as an intersection.

Example 2: Intersection in TIGER-based GSD includes end points for all segments. The US geocoder geocodes to this intersection.

City lookup

The US geocoder relies on USPS data to determine addresses. If a new address was input, it might not have been recognized despite the address being valid if it was not yet valid according to the USPS. An example of an input address that would not match against a UD:

1 Second Street

Stickville, NY 11111

In this example, the city is fictitious and the zip is made up. This would fail to match even with a UD record having that city and that zip, because they are not found in the USPS data. But a user may possess a UD with such a city and zip.

When matching to a UD record, the US geocoder, if necessary, corrects the city name and/or zip code to the data that is in the UD record. The US geocoder is now able to obtain matches for non-USPS cities and zips that were prevented from succeeding or which required temporary workarounds.

Using User Dictionaries with address point interpolation

An important part of the process of creating a User Dictionary is to specify a mapping of fields from your source data. There are two main categories of data fields: required and optional.

Of the optional fields, there are two that have an impact on the address point interpolation feature. These are the "Left Odd/Even" and "Right Odd/Even" fields. If these are not populated, the results from address point interpolation is less accurate.

Please be aware that aforementioned fields are not populated by source data obtained via MapInfo StreetPro. You must modify the source TAB file by adding the "Left Odd/Even" and "Right Odd/Even" indicator fields, and create queries to populate them. Source data obtained from other products, or your own data, may have similar issues.

To add the "Left Odd/Even" and "Right Odd/Even" indicator fields to a source TAB file, you must add them and then run a series of SQL update queries to populate them. The fields should be filled in with “O” (odd), “E” (even), or “B” (both). Below are the steps for adding these fields:

1. Add two 1-char columns to your TAB file.

   Naming each column, for example, Ind_Right and Ind_Left.

2. Perform the following updates to populate these fields:

   - Update <tablename>

     Set Ind_Left="E", Ind_Right="O"

     Where From_Left mod 2=0 AND To_Left mod 2=0

   - Update <tablename>

     Set Ind_Left="O", Ind_Right="E"

     Where From_Left mod 2=1 AND To_Left mod 2=1

   - Update <tablename>

     Set Ind_Left="B", Ind_Right="B"

     Where From_Left="" AND To_Left=""

Note: These example queries are simplified for illustrative purposes. Your actual queries may need to be more complex.

4 – Using the Centrus User Dictionary Writer

Once you have installed the UD Writer and prepared your source data to meet the requirements for a User Dictionary, you can use the UD Writer to automate the User Dictionary creation process.

Note: Remember that in order to create a User Dictionary, you must edit the USA_DataManagerSettings.properties file to point to the appropriate US geocoder data files. See 2 - Installing the Centrus User Dictionary Writer for more information.

In this chapter

Running UD Writer 15

Running UD Writer

1. You may run the application by navigating to the appropriate directory from the command (cmd) line, and executing run.bat.

   You may also run the application by selecting **Start** > **All Programs** > **Centrus** > **Centrus User Dictionary Writer** > **Run UD Writer**.

1. When the MapInfo User Dictionary Utility window opens, select **USA** from the Country drop-down list, if not already selected.
2. Under Table Name, verify the file selected is correct. If not, click **Browse** to find the appropriate TAB file that contains the source data for your User Dictionary.
3. Click **Next** to continue.
4. Select the required columns. These columns are specified in the chosen TAB file.
5. Click **Next** to continue.
6. Select any optional columns, if you want them included in the User Dictionary. Since these are optional fields, you do not have to specify them.
7. Click **Next** to continue.
8. Enter a description under User Directory Description.
9. Specify the Output User Directory by clicking **Browse**. You must specify an empty directory.
10. Click **Finish** to create the User Dictionary.

    The UD Creation Progress window opens.

11. When complete, the UD Creation Complete window opens, showing a log file that summarizes processing and lists all the created files. If any errors occurred, these will be indicated in the Error File area of the dialog.
12. Click **Exit** to exit the MapInfo User Dictionary Utility.

After successfully creating the User Dictionary, the specified directory contains the files that comprise the User Dictionary (with .mmj, .jdr, .jdx, .sdx, and .bdx file extensions). The log file and error file are also stored in the same directory.



One Global View
Troy, NY 12180-8399
USA

www.pitneybowes.com

Support: +1 (800) 367-6950
Main: +1 (518) 285-6000
Fax: +1 (518) 285-7060

©2017 Pitney Bowes
All Rights Reserved