Geocoding Standards

Prepared by:

GIS Services
Office for Information Resources
312 8th Avenue North, Suite 1600
Nashville, TN 37243-0288

Last modified: 11 January 2007

Table of Contents

Executive Summary

This document represents accepted criterion whereby address collection and correction is best implemented to facilitate an accurate and usable geocode of such information. These standards are based on geoprocessing tools within the Environmental System Research Institute's (ESRI) geographic information system (GIS) application framework, and from other nationally recognized standards (cf. *Other Resources*).

Specific examples of potential addressing problems that may result in a non-geocode are provided. This may serve as a guide for GIS users and/or data custodians to improve or correct addressing information stored in a database.

The address matching process is basically 3 steps: Standardization, Matching and Scoring. Standardization consists of parsing the address string into its respective address elements. Next, all of the separated and standardized address elements are compared to existing data to find all of the combinations that possibly match. Finally, a score is computed for each combination of parts that match and then the highest score(s) are returned to the user.

The "matching" algorithms used by the search, allow for alternate names, alternate spellings, bad punctuation, and other typographical errors. These, however, have their limitations. The matching process limits the number of records it needs to search by restricting blocks of records to specific values, and by weighting the importance of other values. Misspellings and the use of punctuation generally slows or even inhibits this process.

1.0 Address Elements

Street addresses are comprised of several addressing elements. To ensure a proper geocode of street address information, it is important to understand these elements and how they are parsed in the geocoding engine.

NUMBER

Example: 120 W MAIN ST ; 345 AVENDIA C ; 156 SCENIC PKW SE ; 5870 HWY 200

The street number is typically what is referred to as the house number. This identifies the point along the street at which the address resides.

DIRECTIONAL

Example: 120 W MAIN ST ; 345 AVENIDA C ; 156 SCENIC PKW SE ; 5870 HWY 200

The directional consists of the road bearing, identified with the cardinal or ordinal points of the compass. To separate these values from actual road names (e.g., "NORTH ST"), these should always be abbreviated as follows: N, S, E, W, NW, NE, SW, SE.

TYPE

Example: 120 W MAIN ST ; 345 AVENIDA C ; 156 SCENIC PKW SE ; 5870 HWY 200

The street type is a descriptor that identifies the type of street. These can come after (as is common) or before the actual street name.

NAME

Example: 120 W MAIN ST ; 345 AVENIDA C ; 156 SCENIC PKW SE ; 5870 HWY 200

The street name is the primary name given to the street. Commonly, "HWY 200" is thought of as the street name, but notice in the above examples, that "HWY" is not a part of the street name.

2.0 Address Standardization

When collecting address information, the key to a successful geocode is standardization. Correctly identifying the address elements within the address string is key to effective geoprocessing. Following these principals will alleviate the pitfalls often associated with establishing a spatial reference to address data.

2.1 Spacing

Spacing is the key designator in distinguishing the various address elements in a particular address string. With the exception of street names with ≥ 2 words, spaces provide an easy parser for the eye and the geocoding engine. The correct format is one (1) space between each of the address elements. Run-on words can cause confusion; for example, it is hard to tell if the address string "NORTHWEST ST" should actually be "N WEST ST". Similarly, the address string "HWY200" does not separate the distinct address elements of street type ("HWY") and street name ("200").

2.2 Inclusion

It is important that all necessary address elements that are known be included in the address string. Omission of street types or directionals could result in mis-matched records when taken through the geoprocessing application. For example, the address string "123 MAIN" might return viable options to include "MAIN ST", "MAIN DR", "MAIN BLVD", "MAIN RD". The street directional is also important, whereas there may be a "MAIN ST" and a "W MAIN ST" within the same target area.

2.3 Exclusion

Certain values should always be excluded from the address string. Secondary address information (e.g., apartment, suite, building, or floor information) should never be included with the address string. This information should be captured in a separate field (spreadsheet or database) or delimited as separate from the address (ASCII text file). It may be necessary or important to retain this information for further analysis within the GIS, and as with all data, should be standardized. A list of appropriate abbreviations for secondary address information can be found in Appendix B.

2.4 Spelling

Correct spelling is a must. Spell checking must be performed to ensure that there is no accidental overlap between different streets with similar spellings. "MAIN ST" is different from "MAINE ST" and could geocode to a completely different part of the target area returning an incorrect location.

2.5 Punctuation

Typically, to expedite the geocoding process, there should be no punctuation used in the address string (the only exception is the hyphen (-)). Programming languages make use of punctuation and this can cause inaccurate results if the data being processed also contains these values. For a list of punctuation marks to avoid, reference Appendix C. A special case arises for street names composed of initials, as in "S. P. MCCLANAHAN RD". Street names that contain initials should be converted to the format "S_P_MCCLANAHAN RD", using the underscore (_) as a separator.

2.6 Abbreviation

Abbreviations can be a hindrance or an advantage depending upon the elements of the street address to which it is applied. Negative impacts result from abbreviations that are not standardized throughout the dataset, or abbreviations that equate to other elements of the address string. Therefore, the distinction between the address elements within the address string is vital.

2.6.1 Street Directional

Street directionals should be abbreviated to avoid confusion with the street name. Acceptable values are as follows: N, S, E, W, NW, NE, SW, SE

2.6.2 Street Type

Street types can be abbreviated, but only if consistency is followed. For a list of acceptable street type abbreviations, reference Appendix A.

2.6.3 Street Name

The street name should never be abbreviated. The exaggerated case is exemplified in the address string "457 N N St": "NORTH" is the prefix directional and is abbreviated correctly, but the street name is also "NORTH" and the abbreviation causes confusion. "457 N NORTH ST" is the preferred form.

2.7 Special Cases

2.7.1 Alpha-Numeric Street Numbers

Typically found in Wisconsin, this address style pairs number and letter designations together in the street number place. These should never be separated by a space. An example might be "S4W35020 BAKER ST".

2.7.2 Fractional Street Numbers

This type of address contains fractional modifiers for the street number. Although part of the street number place in the address string, the fraction should be separated from the whole number by a single space. In the example "12 1/2 MAIN ST", unless the space was provided for clarity, the geocoding engine would interpret the whole number as "121".

2.7.3 Hyphenated Street Numbers

Typically found in New York, Hawaii and parts of California, this type of address pairs two whole numbers together with a hyphen. An example might be "1210-4 MAIN ST". A cautionary note here: It is advisable to not confuse this type of address style with the

insertion of an address range in the street number place. "1210-4" does not interpret to "1210-1214" as an address range. Address ranges that might be valid for a particular parcel or lot should not be recorded in the address string, but in a separate field or delimited as separate from the address string within your collection data. Each individual address should have its own record or line item regardless of whether it is co-located with other addresses.

2.7.4 State, County, Local Highways

For these types of addresses, the USPS recognizes several possibilities for composition. We recommend following these guidelines:

Example	GIS Services Standard
BYPASS	BYP
COUNTY	CORD
EXPRESSWAY	EXPY
FARM TO MARKET	FM
HIGHWAY	HWY
INTERSTATE	I
ROUTE	RT
STATE HIGHWAY	STHY
STATE ROUTE	SR
TENNESSEE HIGHWAY	STHY
TOWNSHIP	TWHY
US HIGHWAY	USHY

2.8 Zones

Separate from the address string, but equally important to the geocoding process, is the zone information for each address. Since any given area may have duplicate addressing (e.g., several "MAIN ST" addresses with the same street numbers), it is necessary to uniquely identify each address. Therefore, classify each address by a zone. The most frequently used zones are ZIP Code or City / Community designators. This information is very important, especially if the geocoded addresses cover a large area (i.e., county or state). If you have other zonal information, please contact us to see if it can be used in the geocoding process.

3.0 Other Resources

For more information on address standardization and geocoding, please reference these additional resources:

United States Postal Service (USPS) Publication 28

Online
http://pe.usps.com/text/pub28/welcome.htm

PDF
http://pe.usps.com/cpim/ftp/pubs/Pub28/pub28.pdf

USPS Online Abbreviation Standards
http://www.usps.com/ncsc/lookups/abbrev.html

Federal Geographic Data Committee (FGDC) Addressing Standards
http://www.fgdc.gov/standards/projects/FGDC-standards-projects/street-address/index_html

Understanding Geocoding
http://webhelp.esri.com/arcgisdesktop/9.1/index.cfm?ID=1734&TopicName=What%20is%20an%20address%3F&rand=815&pid=1733

Address Geocoding with ArcGIS
http://www.gsd.harvard.edu/gis/manual/geocoding/index.htm

4.0 Glossary of Terms

These definitions are quoted directly from the on-line GIS dictionary provided by ESRI and available at http://support.esri.com/index.cfm?fa=knowledgebase.gisDictionary.gateway .

address

A designation of the location of a person's residence or workplace, an organization, or a building, consisting of numerical and text elements such as a street number, street name, and city arranged in a particular format.

address data

Data that contains address information used for geocoding. Address data may consist of one individual address or a table containing many addresses.

address element

One of the elements that comprise an address. House numbers, street names, street types, and street directions are examples of address elements.

address range

Street numbers running from lowest to highest along a street or street segment. Address ranges are generally stored as fields in the attribute table of a street data layer. They often indicate ranges on the left and right sides of streets.

address standardization

The process of breaking down an address into elements and converting those elements with standard abbreviations or spellings. For best practices, this process applies to preparing the reference data and address data for matching.

algorithm

A mathematical procedure used to solve problems with a series of steps. Algorithms are usually encoded as a sequence of computer commands.

geocode

To assign a street address to a location.

A code representing the location of an object, such as an address, a census tract, a postal code, or x,y coordinates.

geocoding

A GIS operation for converting street addresses into spatial data that can be displayed as features on a map, usually by referencing address information from a street segment data layer.

geocoding engine

An entity in the geocoding framework that drives the geocoding process.

geoprocessing

A GIS operation used to manipulate GIS data. A typical geoprocessing operation takes an input dataset, performs an operation on that dataset, and returns the result of the operation as an output dataset. Common geoprocessing operations include geographic feature overlay, feature selection and analysis, topology processing, raster processing, and data conversion. Geoprocessing allows for definition, management, and analysis of information used to form decisions.

Appendix A

USPS Standard Street Suffix Abbreviations

Primary Suffix Name	USPS Standard
ALLEY	ALY
ANNEX	ANX
ARCADE	ARC
AVENUE	AVE
BEACH	BCH
BURG	BG
BURGS	BGS
BLUFF	BLF
BOULEVARD	BLV
BEND	BND
BRANCH	BR
BRIDGE	BRG
BROOK	BRK
BOTTOM	BTM
BYPASS	BYP
BAYOO	BYU
CIRCLE	CIR
CLUB	CLB
CLIFF	CLF
COMMON	CMN
CORNER	COR
CAMP	CP
CAPE	CPE
CRESCENT	CRE
CREEK	CRK
COURSE	CRS
CAUSEWAY	CSW
COURT	CT
CENTER	CTR
COURTS	CTS
CURVE	CUR
COVE	CV
COVES	CVS
CANYON	CYN
DALE	DL
DAM	DM
DRIVE	DR
DRIVES	DRS
DIVIDE	DV
ESTATE	EST
EXPRESSWAY	EXP
EXTENSION	EXT
FALL	FAL
FIELD	FLD
FALLS	FLS

Primary Suffix Name	USPS Standard
FLAT	FLT
FORD	FRD
FORGE	FRG
FORK	FRK
FOREST	FRS
FERRY	FRY
FORT	FT
FREEWAY	FWY
GARDEN	GDN
GLEN	GLN
GREEN	GRN
GROVE	GRV
GATEWAY	GTW
HARBOR	HBR
HILL	HL
HILLS	HLS
HOLLOW	HOL
HEIGHTS	HTS
HAVEN	HVN
HIGHWAY	HWY
INLET	INL
ISLAND	IS
ISLE	ISL
ISLANDS	ISS
JUNCTION	JCT
KNOLL	KNL
KEY	KY
KEYS	KYS
LAND	LAN
LOCK	LCK
LODGE	LDG
LOAF	LF
LIGHT	LGT
LAKE	LK
LAKES	LKS
LANE	LN
LANDING	LND
LOOP	LOO
MALL	MAL
MEADOW	MDW
MEWS	MEW
MILL	ML
MILLS	MLS
MANOR	MNR
MISSION	MSN

Primary Suffix Name	USPS Standard
MOUNT	MT
MOUNTAIN	MTN
MOTORWAY	MTW
NECK	NCK
OVERPASS	OPA
ORCHARD	ORC
OVAL	OVA
PARK	PAR
PASS	PAS
PATH	PAT
PIKE	PIK
PARKWAY	PKW
PLACE	PL
PLAIN	PLN
PLAZA	PLZ
PINE	PNE
PRAIRIE	PR
PORT	PRT
PASSAGE	PSG
POINT	PT
POINTS	PTS
RADIAL	RAD
RAMP	RAM
ROAD	RD
RIDGE	RDG
ROADS	RDS
RIVER	RIV
RANCH	RNC
ROW	ROW
RAPID	RPD
REST	RST
ROUTE	RTE
RUE	RUE
RUN	RUN
SHOAL	SHL
SHORE	SHR
SKYWAY	SKW
SUMMIT	SMT
SPRING	SPG
SPUR	SPU
SQUARE	SQ
SQUARES	SQS
STREET	ST
STATION	STA
STRAVENUE	STR
STREETS	STS
TERRACE	TER

Primary Suffix Name	USPS Standard
TURNPIKE	TPK
TRACK	TRA
TRACE	TRC
TRAFFICWAY	TRF
TRAIL	TRL
THROUGHWAY	TRW
TUNNEL	TUN
UNION	UN
UNIONS	UNS
UNDERPASS	UPA
VIADUCT	VIA
VISTA	VIS
VILLE	VL
VILLAGE	VLG
VALLEY	VLY
VIEW	VW
VIEWS	VWS
WALK	WAL
WAY	WAY
WELL	WL
WELLS	WLS
CROSSING	XIN
CROSSROAD	XRD

A more extensive list and other abbreviation standards can be found at: http://www.usps.com/ncsc/lookups/abbrev.html

cf. USPS Publication 28

Appendix B
USPS Standard Secondary Unit Abbreviations

Secondary Unit Designator	USPS Standard
APARTMENT	APT
BASEMENT	BSMT **
BUILDING	BLDG
DEPARTMENT	DEPT
FLOOR	FL
FRONT	FRNT **
HANGAR	HNGR
LOBBY	LBBY **
LOT	LOT
LOWER	LOWR **
OFFICE	OFC **
PENTHOUSE	PH **
PIER	PIER
REAR	REAR **
ROOM	RM
SIDE	SIDE **
SLIP	SLIP
SPACE	SPC
STOP	STOP
SUITE	STE
TRAILER	TRLR
UNIT	UNIT
UPPER	UPPR **

** Does not require Secondary Range Number to follow.

This list and other abbreviation standards can be found at:
http://www.usps.com/ncsc/lookups/abbrev.html

cf. USPS Publication 28

Appendix C
Non-Uniform Punctuation in Address Field Values

These punctuation marks should not be used when editing address field values.

Description	Punctuation
Ampersand	&
Apostrophe	‘
Asterisk	*
At	@
Back Slash	\
Braces Open	{
Braces Close	}
Bracket Open	[
Bracket Close	]
Caret	^
Colon	:
Comma	,
Dollar Sign	$
Double Quotes	“
Ellipsis	...
Equals	=
Exclamation Mark	!
Greater Than	>
Less Than	<
Number Sign	#
Parenthesis Open	(
Parenthesis Close	)
Percent Sign	%
Period	.
Pipe	|
Plus	+
Prime	`
Question Mark	?
Semi-Colon	;
Virgule	/
Tilde	~
Underscore	_

Appendix D
USPS Abbreviations for Spanish-Language Addresses

Spanish Prefix	Standardization	English Translation
AVENIDA	AVE	Avenue
CALLE	CLL	Street
CAMINTO	CMT	Little Road
CAMINO	CMT	Road
CERRADA	CER	Closed
CIRCULO	CIR	Circle
ENTRADA	ENT	Entrance
PASEO	PSO	Path
PLACITA	PLA	Little Plaza
RANCHO	RCH	Ranch
VEREDA	VER	Small Path
VISTA	VIS	View

composed from the USPS Publication 28