BYTES of the BIG APPLE™ Street Name Dictionary (SND)

The Street Name Dictionary (SND) is a Geosupport System[1] data file containing the set of names of New York City geographic features that the system recognizes and (except for certain 'front-truncated street names') considers valid. The SND is updated continually and new releases are produced for user access several times per year.

Despite its file name, the SND contains not only street names, but also a wide variety of other geographic feature names, including the names of tunnels, bridges, rail lines, shorelines and geographic place names of various kinds.

It is important to note that New York City geographic names are meaningful only when the borough is identified, since features in several boroughs can have the same name. For example, all five boroughs have a street named Broadway.

All geographic feature names in the SND are in a standardized or 'normalized' Geosupport System format called the 'sort format'. A comprehensive description of Geosupport's normalizing algorithm and the sort format is beyond the scope of this document. The following are three notable characteristics of the sort format:

- 'Ordinal suffixes' are absent. Ordinal suffixes are the endings 'st', 'nd', 'rd' and 'th' often used at the ends of words in street names that contain numeric digits, such as '71st', '42nd', '23rd' and '107th'. For example, the SND contains WEST 42 STREET rather than WEST 42ND STREET.

- Words that consist of numeric digits are preceded by one or more 'alignment blanks'. The presence of the alignment blanks causes names containing numeric characters to sort appropriately (hence the term 'sort format'). Three alignment blanks precede a one-digit number, two precede a two-digit number and one precedes a three-digit number. The alignment blanks are in addition to the usual word-separating blanks. In the following examples, for typographic clarity, dashes represent alignment blanks while conventional spaces are used for word-separating blanks: EAST ---3 STREET, --12 AVENUE, EAST -125 STREET. For aesthetic reasons, alignment blanks are not included in any of the examples in this document outside of this paragraph.

- Standard street name words are abbreviated (such as ST for STREET, RD for ROAD, E for EAST) only to the minimum extent necessary to enable names to fit within the SND's 32-byte name fields. With the vast majority of SND names, no such abbreviating is necessary; however, standard abbreviations are used in some particularly long names.

[1] The Geosupport System is a comprehensive geographic processing application developed and maintained by the Department of City Planning's Geographic Systems Section. See the *Geosupport System User Programming Guide* for detailed system documentation.

The character set used for the names in the SND consists of the letters of the alphabet, the numeric digits 0 through 9, the dash (for example, in Manhattan name BEN-GURION PLACE), the ampersand (for example, in Manhattan name BARUCH INFORMATION & TECH BLDG), the apostrophe (for example, in the Bronx names O'BRIEN AVENUE and SAINT JOHN'S AVENUE), and the slash (for example, in Manhattan name 9/11 MEMORIAL MUSEUM).

The SND contains a single header record, followed by one data record for each normalized name of a geographic feature within a borough. The name is stored in the field Geographic Feature Name. Record positions 1-34, consisting of the one-byte filler in position 1, the Borough Code and the Geographic Feature Name, constitute a unique record key, which serves as the primary key in the VSAM version of the file. There are also three alternate VSAM keys (see the section on record keys).

Types of Geographic Features

The type of geographic feature that each name in the SND represents is identified by the field Geographic Feature Type. The SND contains the following types of New York City geographic feature names:

- The names of New York City's streets (including highways), tunnels and bridges

- Certain pseudo-street names, including BEND, CITY LIMIT and DEAD END, and certain Duplicate Address Pseudo-Street names (DAPS's), such as HILLSIDE AVENUE DOUGLASTON

- The names of certain non-street features, including some rail lines and shorelines

- Selected place names, of which there are three types:

  Addressable place names are place names that can be combined with address numbers to form addresses. Examples: the Manhattan names FEDERAL PLAZA and PENN PLAZA. Examples of the use of these names in addresses: 26 FEDERAL PLAZA, 2 PENN PLAZA.

  Non-Addressable Place names (NAP's) are place names that cannot be combined with address numbers to form addresses. Typical NAP's include the names of islands, parks, airports, bodies of water, building complexes, major named individual buildings etc. Examples: the Manhattan names EMPIRE STATE BUILDING, BRYANT PARK and CITY HALL, and the Queens name LA GUARDIA AIRPORT. Note that some NAP's (such as EMPIRE STATE BUILDING) refer to buildings that also have conventional street addresses.

  Intersection Names are names given to the intersection of two or more streets usually but not necessarily by the City Council. Intersection Names allow an intersection to be identified using one input "street" name instead of two as is

currently required. BEN KIMMEL SQUARE in the Bronx is an example of an Intersection Name.

- Front-truncated street names. These are names formed by deleting the word EAST or WEST from the front of a valid street name. The SND contains only front-truncated street names that are derived from selected Bronx and Manhattan street names beginning with EAST or WEST. Although front-truncated street names are not strictly valid, they are widely used. Manhattan examples: HOUSTON STREET (derived from EAST HOUSTON STREET and WEST HOUSTON STREET), 7 STREET (derived from EAST 7 STREET). The SND records representing front-truncated street names have a different layout from that of all other SND records.

Aliases and Street Codes

For many geographic features, the SND contains two or more aliases, that is, alternative names for the same geographic feature (or any part thereof) or spelling variants of the same name. One common case of aliases in the SND is names containing a numeric word between 1 and 10; these are all represented in two different spellings: with the numeric word expressed in digits (such as 6 AVENUE) and with the numeric word spelled out alphabetically (SIXTH AVENUE). As an example of a special case, 11 AVENUE/ELEVENTH AVENUE and 12 AVENUE/TWELFTH AVENUE in Manhattan are also represented in the SND both numerically and alphabetically. (Note that ordinal suffixes are absent when numeric digits are used and are present when numeric words are spelled out alphabetically.)

The following are examples of various types of aliases (in Manhattan except as otherwise noted):

- SIXTH AVENUE and AVENUE OF THE AMERICAS are aliases, since they are alternative names for the same street.

- SEVENTH AVENUE, FASHION AVENUE and ADAM POWELL JR BOULEVARD are aliases, since they are alternative names for the same street or (in the latter two cases) parts thereof.

- SIXTH AVENUE and 6 AVENUE are aliases, since they are spelling variants of the same street name. Other examples of spelling variant aliases are MAC DOUGAL STREET, MACDOUGAL STREET, MC DOUGAL STREET and MCDOUGAL STREET; BEN-GURION PLACE, BEN GURION PLACE and BENGURION PLACE; ST MARKS PLACE, SAINT MARKS PLACE and SAINT MARK'S PLACE; and (in the Bronx) O'BRIEN AVENUE, OBRIEN AVENUE and O BRIEN AVENUE.

There is a field in each SND data record for the Geosupport System Borough-and-10-digit Street Code (B10SC) that is assigned to the geographic feature name of that record. Despite this data item name, Geosupport 'street' codes are assigned not only to street names but to all types of names in the SND except front-truncated street names.

The existence of an alias relationship between two names in the SND can be readily identified from their street codes: Two names are aliases if and only if the first six bytes of their B10SC's, known as the Borough-and-5-digit Street Codes (B5SC's), are identical. It can also be readily determined whether two aliases for a geographic feature are valid for the same part (possibly all) of the feature: Two names are valid for the same part (possibly all) of a feature if and only if the first eight bytes of their B10SC's, known as the Borough-and-7-digit Street Codes (B7SC's), are identical.

For example, SEVENTH AVENUE, 7 AVENUE, FASHION AVENUE and ADAM POWELL JR BOULEVARD in Manhattan are all aliases, so their street codes all have the same B5SC value. In addition, SEVENTH AVENUE and 7 AVENUE are valid for the same part of the street (in this case, the entire street), so they have the same B7SC value. FASHION AVENUE is valid for a different part of the street (the portion in the Garment District of Midtown), so it has a different B7SC value. ADAM POWELL JR BOULEVARD is valid for yet a different part of the street (the portion north of Central Park), so it has yet a different B7SC value.

In the case of a complex and its constituent entities, the assignment of street codes is structured analogously to that just described for streets. The names of the entire complex and the names of its constituent entities are all treated as aliases of each other, since they are all names of the same geographic feature (the entire complex) or parts thereof (the constituent entities of the complex). "Being treated as aliases" means that the B10SC's assigned to these names all have the same B5SC value. Within the umbrella of this B5SC value, the entire complex has its own distinct B7SC value and each constituent entity has its own distinct B7SC value. For example, all of the names for Manhattan's Lincoln Center complex and its constituent entities have the same B5SC value. In addition, LINCOLN CENTER and LINCOLN CENTER FOR THE PFMG ARTS are alternative names of the same part of the complex (in this case the entire complex), so these two names have the same B7SC value. AVERY FISHER HALL and PHILHARMONIC HALL are alternative names of the same part of the complex (in this case, a particular building), so these two names have the same B7SC value, which differs from the B7SC values assigned to the names of the entire complex and to the names of the complex's other constituent entities.

Partial Names

A partial name is a name that is not itself a full geographic feature name nor a full front-truncated street name, and that can be formed by deleting one or more entire words (not portions of words) from the end of a full name. The full name from which a partial name is formed is called that partial name's progenitor. (This term is also used to refer to a full name from which a front-truncated name is formed.)

Consider the following examples, all in Manhattan. LEXINGTON and WEST 23 are partial names for the full street names LEXINGTON AVENUE and WEST 23 STREET, respectively. Both MALCOLM and MALCOLM X are partial names for the full street name MALCOLM X BOULEVARD. On the other hand, PARK AVENUE is not considered a partial street name for PARK AVENUE SOUTH, since PARK AVENUE is itself a valid Manhattan full street name.

Partial names are allowed only under the following conditions:

- The partial name's progenitor must be a conventional street name or tunnel name (Geographic Feature Type = blank, 'E', 'F', 'H', 'M' or 'T'), a bridge name (Geographic Feature Type 'B') or a front-truncated street name (Geographic Feature Type 'S').

- The partial name must be unambiguous in its borough; that is, it must have a unique progenitor in the borough. If a partial name is derivable from more than one full street, tunnel, bridge or front-truncated street name in the given borough, it is ambiguous in that borough and may not be used in Geosupport. For example, in Queens, MADISON STREET is the only full name that begins with the word MADISON. Therefore, MADISON is unambiguously a partial street name for MADISON STREET in Queens, and may be used in Geosupport. However, in Manhattan, the same name, MADISON, is ambiguous and may not be used in Geosupport, because it has more than one possible Manhattan progenitor (MADISON AVENUE, MADISON STREET and MADISON AVENUE BRIDGE, among others).

Data records other than for front-truncated street names (i.e., other than Geographic Feature Type 'S') have a field called Progenitor. The Progenitor is identical to the Geographic Feature Name.

Front-truncated Street Names

A customary, informal practice when referring to many street names in Manhattan and the Bronx that have the word EAST or WEST as their first word is to specify the street name without that first word. For example, the intersection of Broadway and WEST 42 STREET in Manhattan might be expressed as "the intersection of Broadway and 42 STREET", even though 42 STREET is, strictly speaking, an invalid Manhattan street name. We refer to such a name as a front-truncated street name. We call a valid street name from which a front-truncated street name can be derived a progenitor of that front-truncated street name. Thus, WEST 42 STREET is a progenitor of the front-truncated street name 42 STREET. (The term 'progenitor' is also used to refer to the full name from which a partial name is derived.)

The SND contains a limited set of front-truncated street names, which are represented by records with Geographic Feature Type = 'S'. The type 'S' records enable Geosupport to reconstruct the valid progenitor(s) of a user-inputted front-truncated street name so they can be included in the Similar Names List returned to the user. The criteria for the inclusion of front-truncated street names in the SND are as follows:

- The front-truncated street name must be an invalid name in the given borough, i.e., it must not be represented in the SND by an ordinary (non-type 'S') record. For example, THAMES STREET is a valid Manhattan street name, hence is represented by a non-type 'S' SND record. This precludes the representation of THAMES STREET by a type 'S' record as the front-truncated version of the valid Manhattan street name WEST THAMES STREET.

- Only street names (Geographic Feature Type = blank, ‘A’, ‘B’, ‘E’, ‘F’, ‘H’, ‘M’, ‘O’, ‘R’, ‘T’ or ‘Z’) may serve as progenitors, not other types of names such as bridge names, pseudo-street names, non-street feature names or place names.

- The borough must be Manhattan or the Bronx.

- Only street names that have EAST or WEST as their first word may serve as progenitors.

- Progenitors must consist of at least three words. This criterion prevents, for example, the Manhattan street name WEST STREET from serving as a progenitor, and therefore precludes STREET from being represented as a front-truncated street name in the SND.

- (Special exceptions) The valid Manhattan street names EAST END AVENUE, WEST END AVENUE and EAST RIVER DRIVE do not serve as progenitors (it is not customary to front-truncate these names), and therefore type ‘S’ records do not exist for END AVENUE or RIVER DRIVE.

Some front-truncated street names have only one progenitor and some have two. For example, there is no WEST 7 STREET in Manhattan; therefore, the Manhattan front-truncated street name 7 STREET has only one progenitor, EAST 7 STREET. The Manhattan front-truncated street name 8 STREET has two progenitors, EAST 8 STREET and WEST 8 STREET. Front-truncated street names are included in the SND even when they are ambiguous (have two progenitors).

Name	**Type of Name**
EAST 8 STREET	Full street name
WEST 8 STREET	Full street name
8 STREET	Full front-truncated street name with progenitors EAST 8 STREET and WEST 8 STREET

The layout of type ‘S’ records differs from that of non-type ‘S’ records. The contents of a type ‘S’ record are as follows:

- The field Geographic Feature Name contains the full front-truncated street name.

- The contents of the Borough Code field immediately preceding the Geographic Feature Name, the Numeric Name Indicator field and the Record Type field are the same as they would be for a non-type ‘S’ record.

- The Geographic Feature Type contains an ‘S’.

- There are two sets of fields containing information on the one or two progenitors of the given front-truncated name, labeled as Progenitor-1 and Progenitor-2. Each set consists of a First Word field containing a one-letter abbreviation for the progenitor’s first word, either ‘E’ (for EAST) or ‘W’ (for WEST); the progenitor’s Geographic Feature Type

code; and the progenitor's B5SC value. The fields for Progenitor-1 always contain values. If the front-truncated street name has only one progenitor, the fields for Progenitor-2 contain blanks; if it has two progenitors, the order in which the information is listed for Progenitor-1 and Progenitor-2 is arbitrary. Consider the following examples.

> The Manhattan front-truncated street name 7 STREET has only one progenitor, EAST 7 STREET. In the type 'S' record for 7 STREET, the First Word of Progenitor-1 contains 'E' (the abbreviation for EAST), the Geographic Feature Type of Progenitor-1 contains a blank (the Geographic Feature Type code for the street name EAST 7 STREET), and the B5SC of Progenitor-1 contains the B5SC of the street name EAST 7 STREET. The First Word, Geographic Feature Type and B5SC of Progenitor-2 are all blank.
>
> The Manhattan front-truncated street name 8 STREET has two progenitors, EAST 8 STREET and WEST 8 STREET. In the type 'S' record for 8 STREET, the two First Word fields contain 'E' and 'W' in some order, and the two Geographic Feature Type fields and the two B5SC fields contain the information for EAST 8 STREET and WEST 8 STREET in the corresponding order.

- The field Length of Progenitor(s) contains the length in bytes of the given front-truncated street name's progenitor(s). This length is always 5 greater than the length of the front-truncated street name itself (four bytes added for the word EAST or WEST and one byte for the word-separating blank that follows EAST or WEST). For example, in the record for the Manhattan front-truncated street name HOUSTON STREET, which has a length of 14 bytes, the Length of Progenitor(s) field contains the value 19 (= 5 + 14).

SND File Format

Previously there were two formats of the SND. The older format which was designed for the mainframe was referred to as the Mainframe Specific Work Area (MSW) format. The newer form which was designed to be useable on virtually all platforms is referred to as the Character Only Work Area (COW) format. Since the MSW format is no longer used, only the COW format is discussed in this document.

COW SND Record Keys

Key	Constituent Fields	Size	Positions	Used by Geosupport Function(s)
Primary	Filler, Borough Code, Geographic Feature Name	34	1-34	1, 1A, 1E, 1N, 2, 3, 3C, 3S
AIX1	Primary and Principal Name Flag & B5SC	8	35-42	D, 1A, 1E, 2, 3, 3C, 3S
AIX2	Indicator for Principal Name of Local Group, B7SC	9	36-44	DG, 1A, 1E, 2, 3, 3C, 3S
AIX3	B10SC	11	38-47	DN, 1A, 1E, 2, 3, 3C, 3S

COW FILE SPECIFICATIONS:

This file is maintained as an SQL database on a server. It is uploaded to the mainframe where it is converted into a DLL that facilitates searches corresponding to the four keys described above.
LRECL = 200.
Number of records = approximately 88,000; varies over time.

COW RECORD LAYOUT FOR HEADER RECORD:

FIELD	SIZE	POSITIONS FROM	TO	COMMENTS
Header ID – '0000SND'	8	1	8	
DATE OF FILE CREATION	6	9	14	yymmdd format
GEOSUPPORT RELEASE IDENTIFIER Also known as Cycle Number	4	15	18	e.g., "01B "
NUMBER OF RECORDS IN FILE (incl. header record)	8	19	26	Zero-filled
FILLER	174	27	200	

COW RECORD LAYOUT FOR NON-TYPE 'S' DATA RECORDS:

FIELD	SIZE	POSITIONS FROM	TO	COMMENTS
PRIMARY VSAM KEY:	34			
Filler	1	1	1	Always '1'
Borough Code	1	2	2	
Geographic Feature Name	32	3	34	Full Name
Indicator for Primary Name of Street	1	35	35	P or V
Indicator for Principal Name of Local Group	1	36	36	F or S
B10SC:	11			
Borough Code	1	37	37	
Five Digit Street Code	5	38	42	
Local Group Code	2	43	44	
Spelling Variation	3	45	47	
Filler	2	48	49	
Numeric Name Indicator	1	50	50	
Geographic Feature Type	1	51	51	
Length of Progenitor	2	52	53	
Progenitor (Full Street Name)	32	54	85	
Minimum Street Name Length	2	86	87	
Twenty Byte Street Name	20	88	107	
Horizontal Typology Type Code	1	108	108	Blank, G, US or R
Filler	92	109	200	

COW RECORD LAYOUT FOR TYPE 'S' DATA RECORDS:

FIELD	SIZE	POSITIONS FROM	POSITIONS TO	COMMENTS
PRIMARY VSAM KEY:	34			
Filler	1	1	1	Always '1'
Borough Code	1	2	2	
Geographic Feature Name	32	3	34	Front Truncated Name
Filler	15	35	49	
Numeric Name Indicator	1	50	50	
Geographic Feature Type	1	51	51	Contains 'S'
Length of Full Name	2	52	53	
Number of Progenitors	1	54	54	Either 1 or 2
Information on Progenitor-1:	16			
First Word of Progenitor-1 (abbreviated)	1	55	55	Either E=East or W=West
Geographic Feature Type of Progenitor-1	1	56	56	
B10SC of Progenitor-1	11	57	67	
Horizontal Topology Flag of Progenitor-1	1	68	68	
Filler	2	69	70	
Information on Progenitor-2 (if any):	16			
First Word of Progenitor-2 (abbreviated)	1	71	71	
Geographic Feature Type of Progenitor-2	1	72	72	
B10SC of Progenitor-2	11	73	83	
Horizontal Topology Flag of Progenitor-2	1	84	84	
Filler	2	85	86	
Filler	87	87	200	

FIELD DESCRIPTIONS

BOROUGH CODE

CODE	MEANING
1	Manhattan
2	Bronx
3	Brooklyn
4	Queens
5	Staten Island

B5SC OF PROGENITOR-1, B5SC OF PROGENITOR-2 (TYPE 'S' RECORDS ONLY)

The B5SC value(s) of the given front-truncated street name's progenitors, that is, the valid full street name(s) from which the front-truncated street name can be derived by deleting EAST or WEST as the first word. If there is only one progenitor, its B5SC is stored in B5SC of Progenitor-1, and B5SC of Progenitor-2 is blank. See discussion under entry INFORMATION ON PROGENITORS.

FIRST WORD OF PROGENITOR-1 and -2 (TYPE 'S' RECORDS ONLY)

This is a one-character abbreviation, 'E' for 'EAST' or 'W' for 'WEST', for the first word of the corresponding progenitor of the given front-truncated street name. See discussion under entry INFORMATION ON PROGENITORS.

GEOGRAPHIC FEATURE TYPE

This item indicates the type of geographic feature, an attribute of the geographic feature or the type of geographic feature name represented in this SND record.

blank	None of the below, e.g., either a name of a street that has no hyphenated house numbers and no part of which is within Edgewater Park, or a name of a tunnel, etc
A	Addressable place name
B	Name of bridge
C	Business Improvement Districts
D	Duplicate Address Pseudo-Street name (DAPS)
E	Street is entirely within Edgewater Park
F	Street is partially within Edgewater Park
G	Non-Addressable Place name (NAP) of a complex
H	All house numbers on this street are hyphenated
I	Intersection Name
J	Non-Physical Boundary Features
M	Some house numbers on this street are hyphenated, some are not
N	NAP of a 'stand-alone' geographic feature (not a complex

	or a constituent entity of a complex)
O	Shore Line
P	Pseudo-street name (BEND, CITY LIMIT, DEAD END and their aliases)
R	Rail line
S	Front-truncated street name
T	Tunnel
U	Miscellaneous Structures
X	NAP of a constituent entity of a complex
Z	Ramp

GEOGRAPHIC FEATURE TYPE OF PROGENITOR-1, GEOGRAPHIC FEATURE TYPE OF PROGENITOR-2 (TYPE 'S' RECORDS ONLY)

The Geographic Feature Type code(s) of the given fro10nt-truncated street name's progenitor(s), that is, of the valid full street name(s) from which the given front-truncated name can be derived. If there is only one progenitor, its Geographic Feature Type code is stored in the field Geographic Feature Type of Progenitor-1, and Geographic Feature Type of Progenitor-2 is blank. See discussion under entry INFORMATION ON PROGENITORS.

HORIZONTAL TOPOLOGY NAME TYPE CODE

G	Generic
R	Name applies to a roadbed
U	Undivided
Blank	Name applies to a generic road and possibly a roadbed as well

INDICATOR FOR PRINCIPAL NAME OF LOCAL GROUP (NON-TYPE 'S' RECORDS ONLY)

F	Principal name of its local group
S	Not the principal name

Each local group of names (see the *Geosupport System User Programming Guide* for a discussion of local street name groups) has one name designated as the 'principal name' of that group. The principal name is the name returned by Geosupport System Function DG. The SND record for the principal name is identified with an 'F' in this Indicator field; the SND records for the secondary names (the non-principal names) in the local group are flagged with an 'S'. All Names in a local group have the same B7SC.

INFORMATION ON PROGENITORS (TYPE 'S' RECORDS ONLY)

A front-truncated street name's 'progenitors' are the valid full street name(s) from which the front-truncated street name can be derived by deleting the progenitor's first word, which must be EAST or WEST. A front-truncated street name can have either one or two progenitors. Type 'S' records contain two sets of fields containing information on the given front-truncated street name's progenitor(s). The second set of fields contains values only if the given front-truncated

street name has two progenitors. In each set, there are fields for the one-character abbreviation of the progenitor's first word ('E' for 'EAST' or 'W' for 'WEST'); the Geographic Feature Type of the progenitor; and the Borough-and-5-Digit Street Code (B5SC) of the progenitor (MSW) or the Borough-and-10-Digit Street Code (B10SC) of the progenitor (COW).

NUMBER OF PROGENITORS (TYPE 'S' RECORDS ONLY)

Contains either a '1' or '2', indicating how many progenitors the given front-truncated street name has.

NUMERIC NAME INDICATOR

N	Progenitor is 'numeric'
blank	Progenitor is purely alphabetic

Indicates whether the Progenitor is 'numeric', i.e., contains at least one numeric character (the digits 0-9). (Note that, according to this definition, the street name THIRD STREET is not considered numeric.) Whether a name is numeric affects how the Geosupport System normalizes that name.

PRIMARY STREET NAME FLAG (NON-TYPE 'S' RECORDS ONLY)

P	Primary name
V	Non-primary name

Among all the alias names for a given geographic feature (i.e. for a given B5SC value), one record (i.e. one full name) is flagged as the primary name of the feature. The primary name is the one returned by Geosupport Function D.

PROGENITOR (NON-TYPE 'S' RECORDS ONLY)

The name in this field is the same as the name in the Geographic Feature Name field. This field is a vestigial remnant from when partial street names were included in the SND.

PROGENITOR FIRST WORD-1 and PROGENITOR FIRST WORD-2 (TYPE 'S' RECORDS ONLY)

This is a one-character abbreviation ('E' or 'W') for the possible first word of the given front-truncated street name's progenitor. See discussion under entry INFORMATION ON PROGENITORS.