

Geocoding – What it means and how it works in Risk Meter

Geocoding is the first step in almost all spatial analysis. Through this process an address is supplied and a Longitude and Latitude, also referred to as an X and Y, is generated for the address. There are many levels of accuracy associated with Geocoding and in this document we will refer to geocoding accuracy as Georesults.

Geocoding is dependent on addresses (junk in = junk out)

Interactive (One at a time)

There are a number of ways single addresses are sent to Risk Meter. Companies may submit using our Risk Meter On-Line product, they may pass addresses to Risk Meter through their own applications, or Risk Meter may get passed info directly from an agent or through a companies Point of Sale system. The reason that this is important is because a geocoder is always trying to produce the best possible candidate. The geocoder has a number of ways to match addresses based on input, however, that can work as a positive but may also have negative implications.

What this means in simple terms is as follows: A user types in the following address.

111 S Maine St, Anytown, NY 12345

The geocoder might find an excellent match for 111 Main St, Anytown, NY 12347. The question then becomes is the user talking about the same location.

The geocoder is constantly trying to fix misspellings, missed prefixes, suffix's, town names and zip codes. To further these corrections phonetics may be used at times. Many times people are under the assumption that an S5 automatically means a perfect candidate, however, due to the input, the address in question may have been significantly altered during this matching process and as a result, the candidate may not be the same location as the original input intended it to be.

The point is that an S5 is not automatically perfect. Although this will yield an excellent result the vast majority of the time, please keep in mind that the geocoder is again trying to find the best possible match against a given address and it is possible that an S5 may not be the exact address the user is looking for.

Batch

There are a couple ways addresses may be geocoded. Some companies look to do this operation in Batch mode. This is very helpful in running whole books or pieces of, e.g. by region or state. This file is delivered in the form of a database (Excel, Access, Text, or Flat File). The cleanliness of these files is often dependant on the care administered from the Database owner. Some companies have strict standards when gathering Property location addresses. Others are much less strict.

In some cases it is preventable, and in others, it is not. Here are some of the issues that may result in poor addresses

1. Restrictions on the length of fields due to the vintage of systems.
2. Free flow text, e.g. "1/2 mile from the corner of the oak tree"
3. Address 1 and Address 2 info in the same field
4. Alternative garage locations put in randomly
5. No street numbers
6. PO Boxes allowed
7. Lack of consistency in abbreviations

Georesults can be looked at and are based on a hierarchy of levels from 5 being the most accurate to 1 being the least accurate. See notes below on the 6.

At a high level, this is how Georesults can be interpreted.

The S category indicates that the record was matched to a single address candidate. The second position in the code reflects the positional accuracy of the resulting point for the geocoded record, as indicated below.

S5 (Most Accurate) - single close match, point located at the street address position
S4 - single close match, point located at the center of shape point path
S3 - single close match, point located at the ZIP+4 centroid
S2 - single close match, point located at the ZIP+2 centroid
S1/Z1 (Least Accurate) - single close match, point located at the ZIP Code centroid
SX - single match, point located at the street intersection

There is also an additional Georesult that can provide excellent match, however, in order to use this Georesult, the following conditions must be met.

S6/Z6 (Most Accurate) - single close match, point is located at point Zip centroid

As stated above an S1 is the least accurate where as the definition of an S6 indicates that this is the most accurate, so how is this possible?

When 5-digit Zip codes have no area, they are represented as dots on a map rather than polygons and have no geographic extent defined in terms of street segments. These point Zips include Post Office box Zips and Unique Zips (single site, building or organization).

If the user inputs an address which includes a Street Address, City, State and Zip Code for a location, and the Georesult returned is an S6, this indicates that the location is at a point Zip Code that have a known Longitude and Latitude (e.g. no interpolation). HOWEVER, if an address supplied contains a PO Box for the physical address, the S6 should not be used. The reason is simple. An S6 indicates that the location is located at a Point Zip code location where the Longitude and Latitude are known. What this means is that is a user inputs an address which contains a PO Box, the S6 indicates that the Longitude and Latitude are that of the Post Office and obviously there are no streets in a Post Office.

How it works in Risk Meter

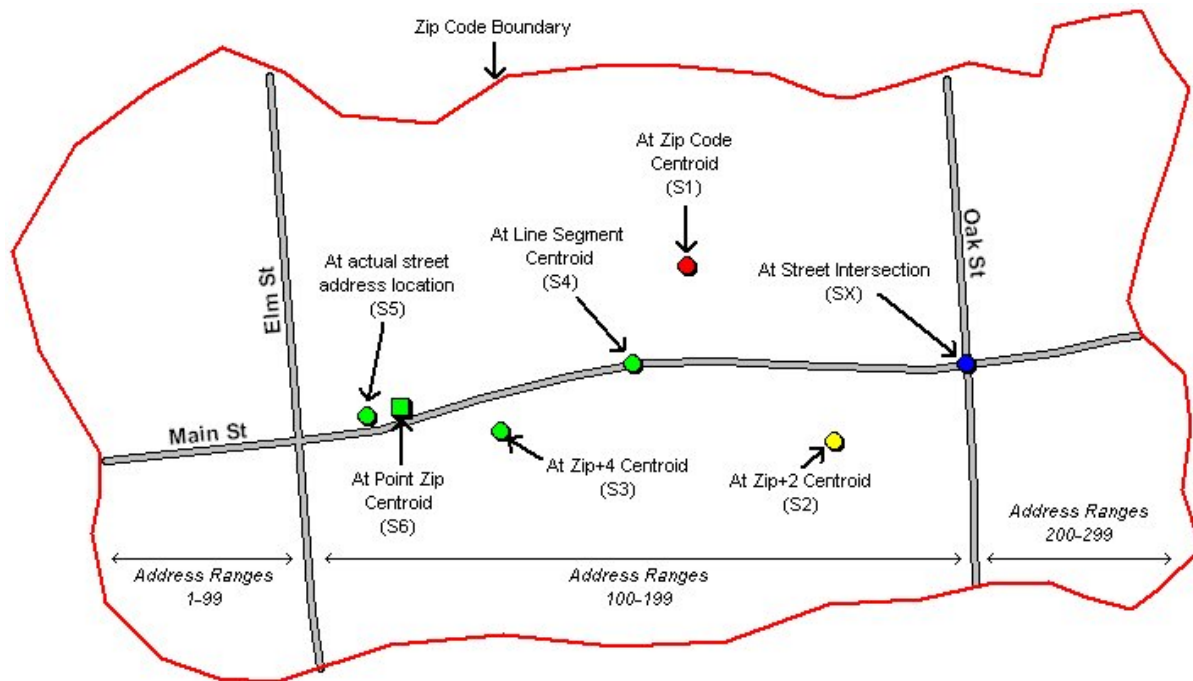
Geocoding works through something called Interpolation. Interpolation is a mathematical algorithm and is defined as: A mathematical procedure which estimates values of a function at positions between listed or given values. Interpolation works by fitting a "curve" (i.e. a function) to two or more given points and then applying this function to the required input.

To put it simply, geocoding uses interpolation to create Longitude and Latitude coordinates for an address based on underlying street geometry to derive an X, Y Value.

The following image will be used as an example.

See Figure 1 on the following page

Figure 1



The following section will discuss each Georesult and explain what pieces of the address matched and what subsequent Georesult would be returned. This Map contains the following features:

Zip Code Boundary – 12345

Streets – Main St, Elm St & Oak St

Address Ranges for Main St – 1-99, 100-199, 200-299

Georesults – S6, S5, S4, S3, S2, S1 and SX

The Addresses entered into Risk Meter to illustrate Georesults will be:

111 Main St, Anytown, NY 12345 for Georesults S5, S4, S3, S2, S1

111 Main St, Anytown, NY 12346 for Georesult S6 and Z6

Main St & Oak St, Anytown, NY 12345

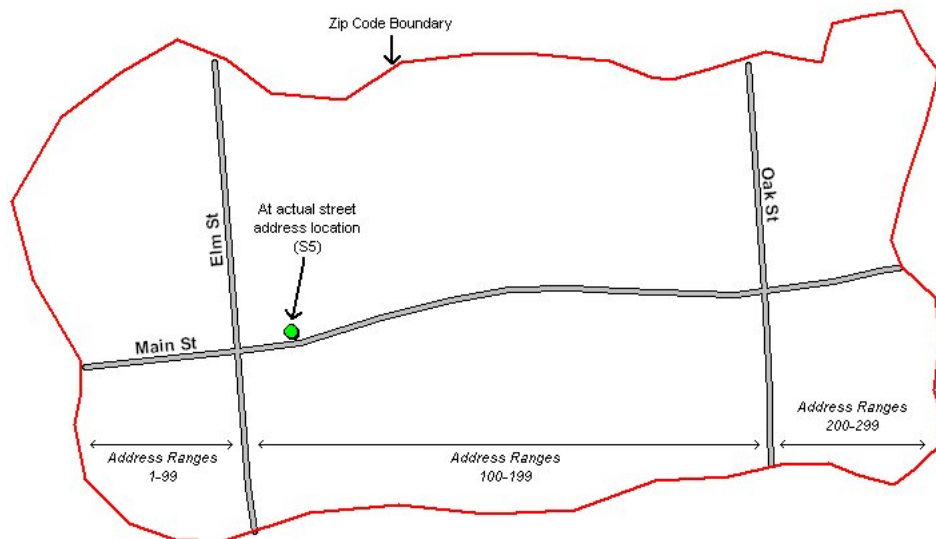
This address is for illustration purposes only. 12345 is an actual Zip Code for Schenectady, NY although the streets are fictional and created to illustrate Geocoding.

The following pages will explain the different Georesults and the subsequent placements of the points which represent a home, office, or building. These pages will contain what pieces of an address match and subsequently the Georesults returned. Further, below this information, there will be some answers to commonly asked questions relating to the Georesult.

S5 – (Most Accurate) - single close match, point located at the street address position. This is an exact match on Street number, Street name, City and Zip Code.

Matching Element in BOLD RED

111 Main St, Anytown, NY 12345-6789



The placement of this point listed above is derived by the geocoder interpolating 111 from the Address Range 100-199 in Anytown, NY 12345. If you look at the placement, you will see that this position is approximately 1/10 of the way up Main St on the left hand side. The geocoder understands that the Address Range for 100-199 is between Elm St and Oak St. If the address were below 100 e.g. 1-99, the point would be placed to the Left of Elm St, if the address were above 200 e.g. 200-299, the point would be placed to the right of Oak St.

Commonly asked Questions

- Q:** How precise is the S5?
A: An S5 has a horizontal Accuracy of +/- 167 ft. which meets Federal Mapping Accuracy standards. This means that in a sample group, at least 90% of sampled points are within this range.
- Q:** Will the Geocoder always produce a Zip+4?
A: No, although this is a CASS Certified Product there are many instances that the +4 will not be appended.
- Q:** Are these actual house locations e.g. There are 10 houses on my street with number 7-28 does the geocoder know how many houses are there?
A: No, again geocoding uses interpolation and looks at address ranges associated with a line segment. If the street only has 10 houses but address ranges of 1-99 it will put 7 near the beginning of the street and 28 will be placed approximately 1/4 of the way along the street segment although 28 may be the last house on the street. In order to determine whether a house number is “real” a user may want to use another piece of software which has DPV which stands for Delivery Point Validation. Again, even DPV will not change the placement given the logic of the geocoder.
- Q:** Why is address being corrected to something entirely different?
A: The geocoder is constantly trying to fix misspellings, missed prefixes, suffix's, town names and zip codes. To further these corrections phonetics may be used at times. There are many aliases, for town/cities, Zip codes and even streets. The address supplied may being corrected to postal standards that the Geocoder uses via the USPS as opposed to common neighborhood names. State Route 156 = Shore Road, Coral Gables = Miami, etc. 12347 may be a post office as opposed to a geographic zip code boundary of 12345.

Geocoding with Risk Meter

Page 4 of 10

©CDS Business Mapping, LLC., 03/2006

All Rights Reserved

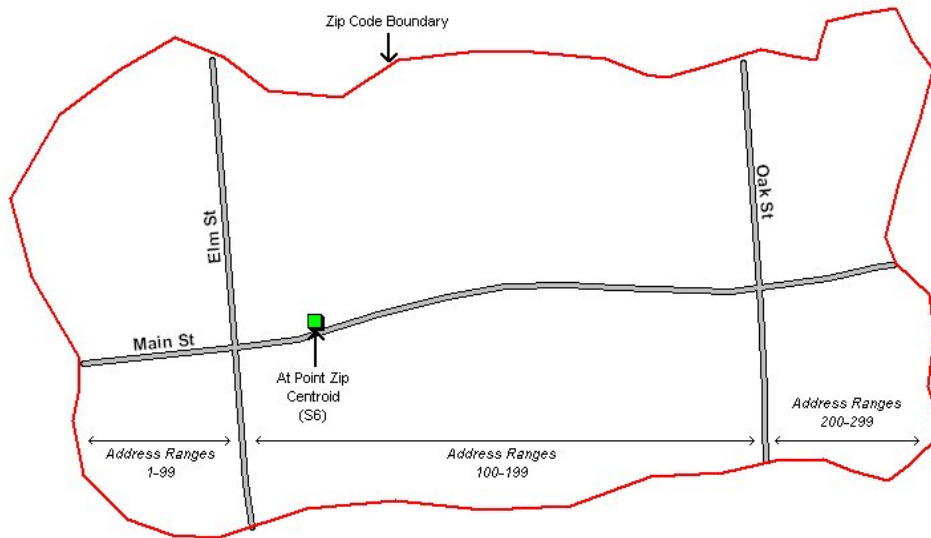
www.cdsys.com

Mapping For The Next Millennium

S6/Z6 - (Most Accurate) - single close match, point is located at Point Zip Centroid. This is an exact match on a Point Zip code where the point has a known X, Y location e.g. GPS derived or Satellite imagery derived Longitude and Latitude.

Matching Element in BOLD RED

111 Main St, Anytown, NY 12346 or **12346** whereas the Longitude and Latitude are not interpolated



Again, an S6 is a superior match to an S5 when the address given contains the Address, City, State and Zip Code. *If a PO Box is given the S6 should not be used*, because the known X, Y is that of a Post Office location which again contains no Streets.

The placement of the point listed above is not derived from Interpolation, instead it is a known X, Y position from a GPS device or derived from satellite/aerial imagery.

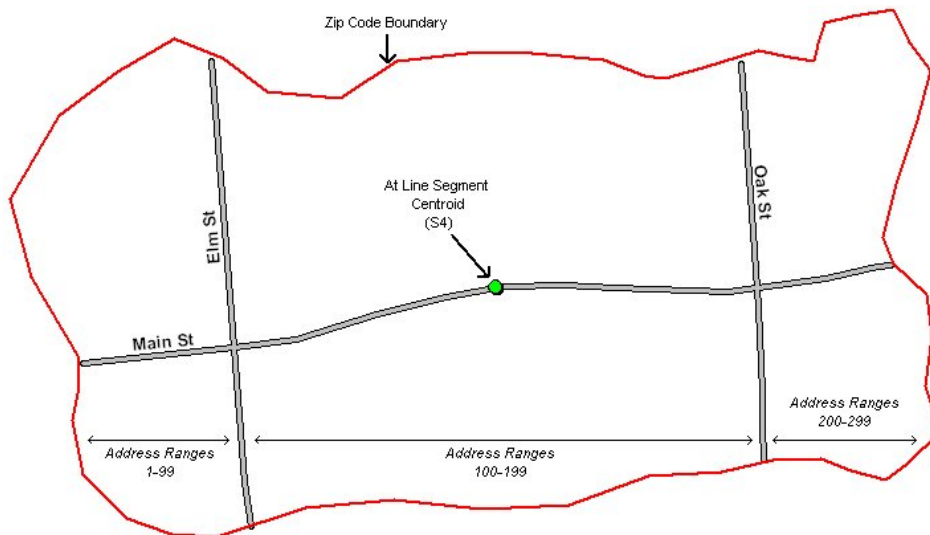
Commonly asked Questions

- Q:** Isn't an S6 better than an S5?
A: An S6 can be considered more precise than an S5 if the user supplies an address, city, state and Zip. If a PO Box is allowed to be passed this Georesult will put the location at the Post office location.
- Q:** What the difference between an S6 and a Z6?
A: From the standpoint of Longitude and Latitude, the answer is none. An S6 and a Z6 will put the location at the same point. However, the way they get there is different. For example: Let's say that 111 Main St is the address for BIG, Inc. BIG sends and receives so much mail that the post office designates a zip code just for BIG. Due to its' size, in both the eyes of the post office and the general public, a Longitude and Latitude is determined for 111 Main St. Therefore if someone were to enter just 12346 they would receive an Z6 because it is doing this via a Zip Code matching technique. However, if someone were to input the address of 111 Main St, Anytown, NY 12346, the geocoder would yield an S6 result.

S4 - single close match, point located at the center of shape point path. This is an exact match on the Street Name, City, and Zip Code but *not the Street number*.

Matching Element in BOLD RED

111 **Main St**, Anytown, NY 12345-6789



An S4 is returned when the address entered corresponds to a street that does not contain address segments. The geocoder will place the point at the center of the line segment. In the case illustrated above, imagine that Main Street did not have any address ranges of 100-199 and as a result, the geocoder would place the point at the Centroid of the Line segment. If you look at the placement, you will see that this position is approximately 1/2 of the way up Main St and located in the center of the road.

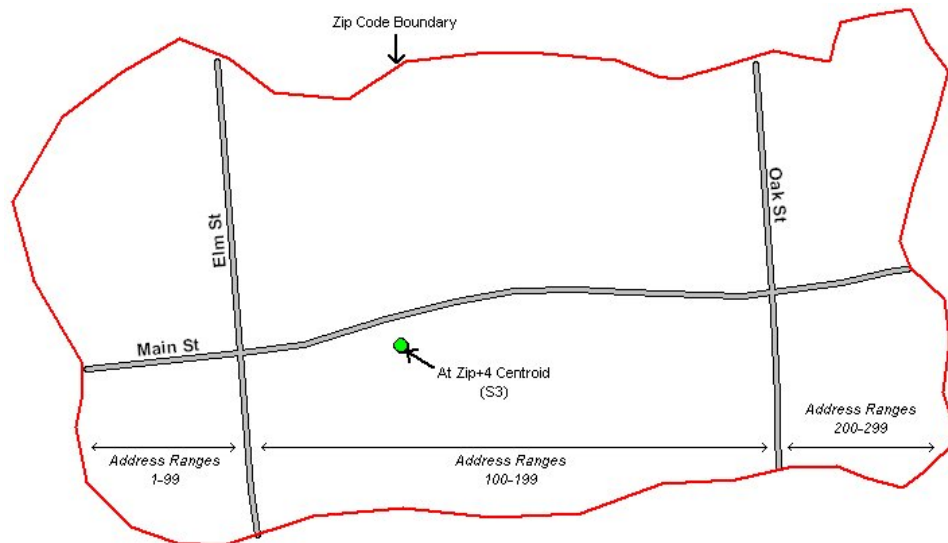
Commonly asked Questions

- Q:** How precise is an S4?
A: The Answer is dependent on the geography of the area. Much of the accuracy of the S4 is dependent on the line segment in question. If the S4 is generated off a street in a rural location, the accuracy can be off by hundreds if not thousands of feet. However, if the S4 is determined off of a line segment in a more concentrated area, the answer could be as about as accurate as an S5. The S4, is not nearly as common as the S5 and therefore the user should feel very comfortable with this result.
- Q:** Will I get an answer for an S4 Result?
A: By default, the Risk Meter will run all tests, except Flood Zone determinations on an S4. Based on customer need, Risk Meter can return results for all addresses provided the customer understands the accuracy levels of the corresponding Georesults.

S3 - single close match, point located at the ZIP+4 centroid. This is an exact match on the associated Zip+4 of the address, e.g. a 9 digit Zip Code match.

Matching Element in BOLD RED

111 Main St, Anytown, NY **12345-6789**



In this case, the geocoder has found a single close match for the address, however, the matching address could not correspond to the Street Geometry. Therefore it can not Interpolate a placement along Line segment (e.g. a road). Therefore, the point is placed at the Centroid of a Zip+4 Location. This may be an individual building, office park or very small group of houses, approximately 10.

The S3 is comparable in accuracy to the S4 if not stronger. An end user is more likely to see an S3 as compared to an S4 but again this should be considered a very strong match.

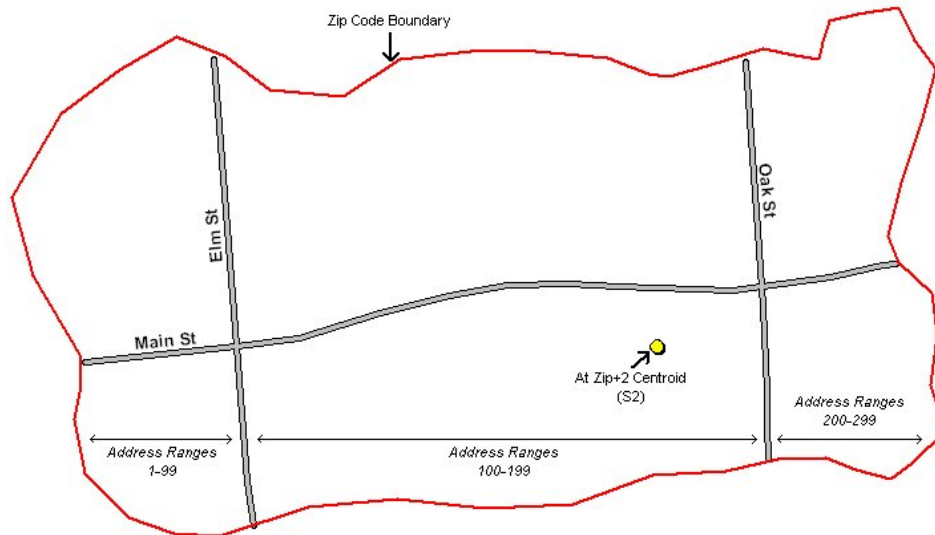
Commonly asked Questions

- Q:** How precise is an S3?
A: An S3 is again a very strong match. An S3 is generated when there is a known Zip+4 location but the street geometry is not developed enough to place this with certainty along an address range on a street segment. An S3 may be generated for a small cluster of houses, possibly an office park, or other industrial site. Although there is not a standard accuracy range, this stands to be a very strong match.
- Q:** Will I get an answer for an S3 Result?
A: By default, the Risk Meter will run all tests, except Flood Zone determinations on an S3. Based on customer need, Risk Meter can return results for all addresses provided the customer understands the accuracy levels of the corresponding Georesults.

S2 - single close match, point located at the ZIP+2 centroid. In this case, the geocoder can find only a Zip+2 Centroid.

Matching Element in BOLD RED

111 Main St, Anytown, NY **12345-6789**

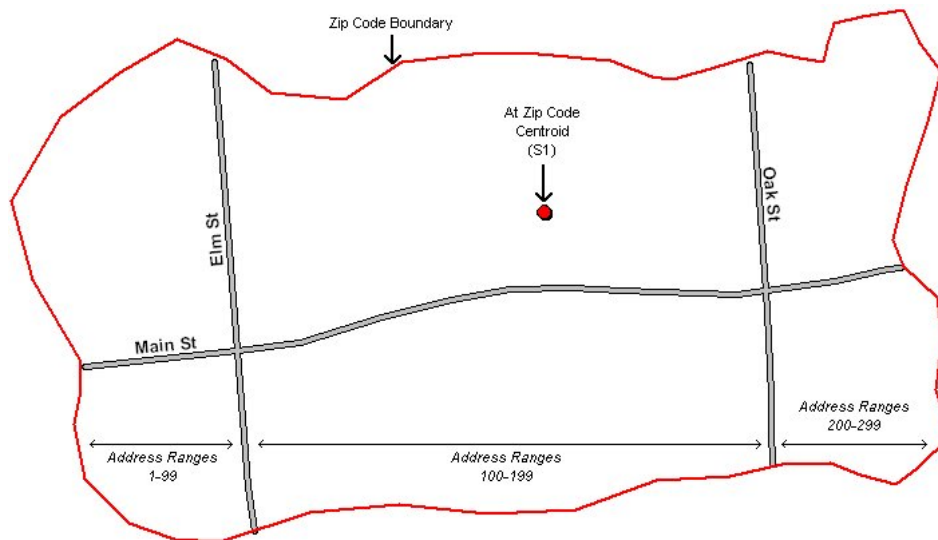


This is a weighted point amongst known Zip+4 centroids. In looking at the illustration above, we give a very simple imaginary Zip Code. Actual Zip Codes are not as small and many times are more geographically complex. Zip+2 points do not fall into predefined order, e.g. start in the NW portion of a Zip Code and proceed SW. The +/- horizontal accuracy can vary tremendously and therefore may not be suitable for some applications.

Commonly asked Questions

- Q:** How precise is an S3?
A: With an S2 the confidence of the accuracy begins to deteriorate. An S2 is amongst known Zip+4 centroids.
- Q:** The address was corrected with a Zip+4, why doesn't the geocoder return at least an S3?
A: Assigning a Zip+4 has to do with the CASS ability of the geocoder. However, because a Zip+4 is known for an address, does not mean that the Zip+4 Centroid is known or available. There are many million Zip+4 addresses, however, there are many times where the Zip+4 centroid is not known and hence can not be returned.
- Q:** Will I get an answer for an S2 Result?
A: By default, the Risk Meter does not return test results for a Georesult of S2. The user would instead see the message "Unacceptable Georesult: S2 (XX) where XX is the test e.g. SL=Shoreline. However, Based on customer need, Risk Meter can return results for all addresses provided the customer understands the accuracy levels of the corresponding Georesults.

S1 (Least Accurate) - single close match, point located at the ZIP Code Centroid. This Georesult is returned when the only part of the address that could be matched is the Zip Code. As a result, the geocoder places the point at the geographic center of the Zip.



This is the least accurate of all Georesults. By default, the Risk Meter does not return test results for a Georesult of S1. The user would instead see the message “Unacceptable Georesult: S1 (XX) where XX is the test e.g. SL=Shoreline

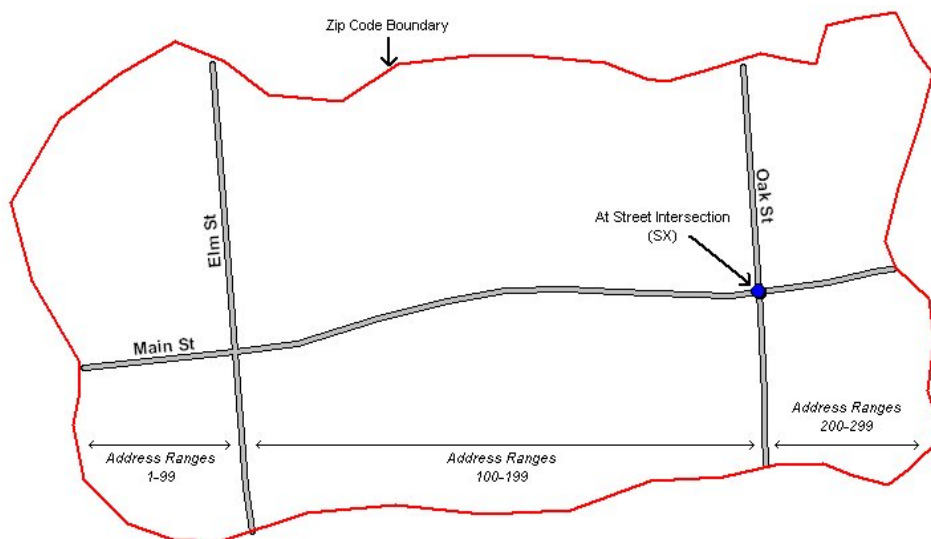
Commonly asked Questions

1. **Q:** How precise is an S1/Z1?
A: The S1 is based on the geographic centroid of the Zip Code boundary. In most cases, people would rarely want to know attributes of a home, office or commercial structure based on this result. The reason being is that the answers returned may have little to do with structure in question.
2. **Q:** The address was corrected with a Zip+4, why doesn't the geocoder return at least an S3?
A: Again, assigning a Zip+4 has to do with the CASS ability of the geocoder. There are cases where the Geocoder can find an excellent match for the address but lacks street geometry, Zip+4 centroids and therefore the best that can be produced is the S1/Z1.
3. **Q:** What is the difference between the S1 and the Z1?
A: From the standpoint of Longitude and Latitude, the answer is none. An S1 and a Z1 will put the location at the same point. However, the way they get there is different. An S1 looks at all the components of the address: Address, City, State and Zip, whereas the Z1 can only match the Zip Code piece.
4. **Q:** Will I get an answer for an S1 Result?
A: By default, the Risk Meter does not return test results for a Georesult of S1/Z1. The user would instead see the message “Unacceptable Georesult: S1 (XX) where XX is the test e.g. SL=Shoreline. However, based on customer need, Risk Meter can return results for all addresses provided the customer understands the accuracy levels of the corresponding Georesults.

SX – single match, point located at the street intersection. This Geocode is returned two street addresses are entered in conjunction with City, State & Zip Code.

Matching Element in BOLD RED

Main St & Oak St, Anytown, NY 12345-6789



This is the only case where omitting the street number is suggested. In the example listed above the geocoder would return an SX if the address entered was Main St & Oak St, Anytown, NY 12345. This point would be placed at the intersection of these two streets. It should be noted that depending on the address ranges associated with the street segments, that more than one candidate may be returned. This is due to the fact that there may be up to four unique address segment combinations. E.g. the grid network of New York City.

An SX is a very strong match so long as the following conditions are met. The first being that you have a known cross street to enter and the second being the physical location is in close proximity, e.g. a few hundred feet to the intersecting streets.

Commonly asked Questions

- Q:** How precise is an SX?
A: The SX is based on the intersection of 2 street segments, so in practical terms, the precision is excellent. However, the SX can only be returned when a user passes in 2 streets, along with a City, State and Zip Code.
- Q:** If I know the house number of one address but not the intersecting street, should it be passed?
A: This is the only case where omitting the street number is recommended. The reason again has to do with address ranges associated with intersections.
- Q:** When should the intersection be used?
A: An intersection is helpful when the user receives an S2 or S1. The problem is that in many cases, users lack this piece of information. Many times a user may get an S2 because the Street database lacks address ranges and Zip+4 info. However, the street segment may be available but unusable in the traditional sense. In this case, if a user has access to the cross street, the user may get an SX which may be very suitable for the placement and subsequent values returned.