

A vertical blue sidebar on the left side of the page, containing several white icons: a lightning bolt striking a city skyline, a network diagram with nodes and lines, a 3D city model, a dashed line forming a curve, a multi-story building, a mathematical formula $= \sqrt{\sum_{i=1}^n (L_i - \bar{L})^2}$, a hurricane, a stylized sun with wavy lines, and a flooded street with a car partially submerged.

APPLICATIONS OF BUILDING- LEVEL DATA FOR IMPROVED EXPOSURE DATA QUALITY AND CATASTROPHE RISK MANAGEMENT

Sanborn CitySets[®] Data

INTRODUCTION

The need for insurers and reinsurers to obtain a better understanding of where their exposures are located, as well as the building-level attributes associated with those locations, beyond what has traditionally been captured for businesses exposed to catastrophes, has been growing rapidly. Having key locations on an account geocoded at five-digit ZIP Code or even street-level resolution is no longer good enough. The need also goes beyond just gathering location-level data for property lines of business.

The issue of obtaining improved exposure data is nothing new. After a series of significant natural catastrophe events in the recent past, companies realized that existing methods of identifying the location of their exposures were insufficient, as were the availability and quality of the building attributes data used for risk modeling, often resulting in poor underwriting and portfolio risk management decisions as well as an aggregation of exposure in hazardous and vulnerable areas. A similar change is happening today as companies look to capture data at the building-level for property and workers compensation lines of business in urban areas.

RMS' response to market demand has included improving the quality of geocoding resolution and the availability of building-level attributes data in urban areas of the U.S. by collaborating with the Sanborn Map Company to implement detailed city-center databases of address and other building attribute data into RMS products. Sanborn CitySets[®] data includes information on all structures in the city centers of the largest U.S. cities. The coverage areas selected were based on an evaluation of exposure density from catastrophic risk perspective, as well as an assessment of terrorism risk. The addresses and building names in this database have been integrated into the geocoding engine to allow RMS clients to geocode their locations to specific buildings, providing a clear understanding of where their exposures are located. In addition, locations that geocode to the street or five-digit ZIP Code level are flagged if they are located in a Sanborn downtown business district, providing a quick proxy of exposure subject to risk as well as an indication that building-level geocoding is available with better address data. The data includes location-level building attribute information. Sanborn data is updated with every new release of RMS products.

Applications of Sanborn CitySets[®] Data

- Improved geocoding accuracy
- Enhanced exposure data
- Improved risk analysis
- Higher resolution accumulation management
- Improved event response

IMPROVED GEOCODING ACCURACY

A key benefit of Sanborn® CitySets data is the ability to geocode to specific buildings in urban areas. Locations are not only assigned a latitude/longitude coordinate, but building-level geocoding stamps a unique building identifier on each location that is geocoded at the building level. The coordinate assigned to this building identifier is based on the centroid of the building. Thus, all valid addresses for a building are stamped with the same building identifier and are geocoded to the same latitude/longitude coordinate.

Sanborn CitySets data meets National Map Accuracy Standards (NMAS), ensuring positional accuracy of these coordinates. In terms of horizontal accuracy, NMAS requires that 90% of the covered features be within 3.3 feet (1 meter) of their true position. In contrast, the positional accuracy of street address geocoding can be measured by the positional accuracy of street segments in the underlying data. For example, how many street segments are within 40 feet (12 meters) of their true ground position? Prioritized based on population, counties that make up 50% of the U.S. population have an average of 90% of the street segments within 40 feet (12 meters) in the GDT data. Expanding this analysis to counties with 80% of the population, less than 50% of the street segments are within 40 feet (12 meters).

Street-level geocoding also includes another source of uncertainty because street address geocoding uses interpolation to assign an address to a certain position on a street segment, as illustrated in this simple example using the address 50 Main Street. For street-level geocoding, the underlying data quantifies the length of the street segment. If the street segment runs from 1 to 100, then an address with 50 would be placed at the mid-point of the street segment. However, there are often assumptions made about the length of the street segment. If the underlying data assumed that the street segment was from 1 to 100, but the actual street segment was from 1 Main Street to 200 Main Street, then placing the address at the mid-point would be inaccurate. Another way to measure this uncertainty is to estimate the maximum error possible from interpolation. City blocks in major population centers average 500 feet (150 meters), thus the interpolation error in urban centers could be as high as half of a city block, or 250 feet (75 meters). Interpolation by itself introduces uncertainty that is not found in building-level geocoding.

As an example, consider the assignment of hazard data to a specific location. RMS often reports back a confidence area. A confidence area is a region that represents the uncertainty in the location of a point on the map. It accounts for the inaccuracies that may exist in the original source map or when determining the exact position of an address during the geocoding process. The confidence area takes into account source map and geocoding uncertainty by providing other hazard values that are within the confidence area. Again, this type of uncertainty does not exist with Sanborn data.

Benefits of Building-level Geocoding

Sanborn building-level geocoding allows companies to look at their exposures in a new dimension. Traditionally, high-resolution geocoding meant that locations were geocoded to the street address level. If a company is concerned with urban exposure, where accumulations within small areas and perils such as terrorism with relatively small damage footprints can be the norm, street address geocoding has some shortcomings that can have significant impact on analysis results.

Street address geocoding uses interpolation to identify a specific latitude/longitude coordinate, and while this method is recommended for many types of analyses, it may not place an address in an appropriate location for analyses that involve smaller footprints or in areas where there is a high variability of hazard information in a relatively small area. For example, if one reviews a map of

locations geocoded to the street address level, it is not uncommon to discover that actual locations are located in the center of streets, placed on the wrong side of the street, or located outside any building footprint. For example, Figure 1 is a 6-block by 6-block area in downtown Chicago. The RED dots (♦) represent the building centroid of the Sanborn data. The YELLOW dots (◆) represent the valid addresses for each building from the Sanborn data. The GREEN dots (◆) represent the locations of all these valid addresses using street-level geocoding (the Sagent geocoding engine using GDT data). Notice how street address coordinates (GREEN dots ◆) are not always clearly aligned with specific buildings. For analyses within urban areas, these types of differences are meaningful.



Figure 1: Sample addresses in Chicago

To further consider the differences between Sanborn building-level geocoding and street address geocoding, RMS geocoded all Sanborn addresses using GDT data to compare the positional differences between building-level geocoding and street-level geocoding. The median difference between a building-level geocoded coordinate and its corresponding street address geocoded coordinate is 98 feet (30 meters).

Based on all Sanborn addresses that geocoded at both the building and street-level:

- 60,101 locations (18.47%) geocoded more than 50 meters apart (164 feet)
- 9,472 locations (2.91%) geocoded more than 100 meters apart (328 feet)
- 1,265 locations (0.39%) geocoded more than 250 meters apart (820 feet)
- 246 locations (0.08%) geocoded more than 500 meters apart (1,640 feet)

ENHANCED EXPOSURE DATA

In addition to the benefits of Sanborn building-level data for more accurate location positioning and accumulation management, the data includes information on the attributes of individual buildings that can be used to validate and enhance the exposure data that is reported for individual accounts. Specifically, data on construction class, occupancy type, year built, number of stories, and square footage enables insurers to validate data submitted on location schedules and to enhance missing or erroneous data. For example, if an insurer receives a location schedule that includes an office building at 303 Second Street, San Francisco, CA, the Sanborn attribute data can be used to verify that the structure was built in 1988, is 12 stories, and has a construction class of Reinforced Concrete Shear Wall. RMS' RiskBrowser® enterprise underwriting system provides integrated functionality to automatically compare all reported building attribute data to the Sanborn data to identify differences and to update the reported data if desired.

Sanborn collects data using multiple methods. First, aerial imagery for the coverage areas is created and processed. The data collected using aerial imagery includes building footprint and building shape information. Building height is also determined from the aerial imagery, which Sanborn uses to derive the number of stories (assuming 12 feet/3.6 meters per floor). In addition, Sanborn derives the estimated square footage by using the building height, footprint, and shape. These attributes exist for all Sanborn buildings. Other data elements included in the Sanborn data are construction, occupancy, and year built.

While Sanborn collects information on all buildings in the coverage area, the priority is on complete and accurate attribute information for buildings over five stories. Thus, the percentage of known attribute values is greater for buildings over five stories. The following table shows the percentage of buildings that have known attribute information, comparing all buildings to buildings over 5 stories.

	Known Construction (%)	Known Occupancy (%)	Known Year Built (%)
All Buildings	10%	94%	9%*
Buildings Over 5 Stories	58%	98%	53%

**54% of buildings have either year built or year range available*

Complete and accurate building attribute data is another important analysis component as missing values can significantly impact analysis results. For example, terrorism vulnerability in the RMS models is a function of construction and building height. Having both of these values for each location will reduce the variability in the modeled damage ratio. Having only one of the values, though not as good, can still greatly reduce the variability. The RMS damage ratios in the terrorism model are based on blast pressure. At a given pressure value, the variability around the mean damage ratio when both construction and height are unknown can be plus 40% or minus 20%. However, simply knowing the height, especially for buildings over 7 stories, can significantly reduce this variability. Similarly, earthquake damage ratios vary by construction, height, and age. Knowing at least one of these attributes can significantly reduce the range of possible damage ratios.

IMPROVED RISK ANALYSIS

Geocoding Differences

Positional differences can impact results, particularly for events in dense urban environments. Geocoding accuracy is a critical component for assessment of risk in city centers for analyses ranging from terrorism attacks to accumulation and natural catastrophe analyses.

While accurate geocoding is critical to accurate analyses, consistency is also an issue. Consistency within the exposure data will help ensure accuracy between locations in order to properly understand adjacency issues. In addition, for accumulation and terrorism analyses, geocoding impacts not only the exposure data analyzed, but also the points selected as the centroid of the analysis. This is unique to analyses in urban areas because of the relatively small accumulation and terrorism event footprints combined with the high density of exposure. For example, if exposure is geocoded at the building-level using Sanborn data, but the analysis is centered on the street-level geocoded coordinate of the primary address, the results may be inaccurate. The centroid of the analysis may be at the corner of the building while the exposure data is placed at the building centroid.

Hazard Value Differences

Positional accuracy is important to correctly assign hazard values, such as distance to targets for terrorism and flood zones and soil types for natural catastrophes such as floods and earthquakes. This is particularly true in areas where hazard information is highly variable within a relatively small area. For example, flood zones can be site-specific. The determination of being in or out of a 100-year flood zone can be a matter of feet.

In addition, soil type in San Francisco has a great deal of variability. Figure 2 shows a soil map for a portion of downtown San Francisco with Sanborn building footprints, building centroids, and valid addresses to highlight the variability of soil type within a small area, even varying by specific buildings.



Figure 2: Soil types in San Francisco

In California, RMS soil types range from 1.0 (rock) to 4.0 (soft soil). The high-resolution data provides a continuous range of values between 1.0 and 4.0, with a lower value corresponding to harder soil. In comparing the soil types assigned when using building-level geocoding to the soil types assigned when using street-level geocoding, RMS found that 10% of the Sanborn buildings in San Francisco that geocoded at GDT street-level had soil differences equal to or greater than 1.0. For an earthquake analysis, these differences can translate into significantly different loss estimates.

For example, The Marriott Hotel at Fisherman’s Wharf in San Francisco is a 5-floor structure built in 1983. With street-level geocoding, the soil type returned is rock (value of 1.5), and with building-level geocoding, the soil type returned is soft soil (value of 3.5). Figure 3 shows the soil types in this area, overlaid with Sanborn building footprints. The addresses are street-level geocoded locations of the valid addresses for the hotel structure.

Landslide and liquefaction values also change as a result of different geocoding placement. Because these are key components in determining the vulnerability, modeled losses also change significantly. RMS modeled the historical 1906 Great San Francisco Earthquake against this location for both geocoding cases. The ground-up loss increased more than 54% with building-level geocoding due to the more accurate placement of the building and more accurate assignment of hazard values.



Figure 3: Soil types at the Marriott Hotel, Fisherman’s Wharf, San Francisco

HIGHER RESOLUTION ACCUMULATION MANAGEMENT

Sanborn CitySets data provides the unique capability to tie exposure to a specific structure instead of just an address or a geocoded point. Through building-level geocoding, companies can evaluate multi-line accumulations at the precise location level, which is a cornerstone of current catastrophe risk management best practices.

To understand the value of building-level geocoding, consider that most buildings in urban centers have multiple addresses or are often identified by building name. If the desired analysis being performed is building specific, it can be difficult to ensure that the correct exposure data is analyzed. For example, the Empire State Building has ten valid addresses. The primary address is 350 5th Avenue, but there are other addresses on 5th Avenue, as well as addresses on 33rd and 34th Streets. A traditional analysis to estimate the exposure in a building may look at a 100-foot (30-meter) radius around the street-address coordinate that corresponds to this primary address, based on an assumption that 100 feet (30 meters) is a typical building size. However, doing so results in two problems:

- Valid Empire State Building addresses get missed because they are located more than 100 feet (30 meters) from the 350 5th Avenue coordinate
- Addresses for adjacent buildings are included in the analysis because they are within 100 feet (30 meters) of the 350 5th Avenue coordinate

This is demonstrated in Figure 4. The RED dots (•) are the valid addresses for the Empire State Building. The BLUE dots (•) are associated with other buildings surrounding the Empire State Building, illustrating the issue with using a radius analysis to determine the exposure within the Empire State Building.



Figure 4: Empire State Building example

The building-level accumulation functionality made possible with Sanborn data allows companies to establish and monitor guidelines for multi-line risk accumulation by specific buildings. For example, companies can establish building-level guidelines for property and workers compensation and monitor actual accumulations against these guidelines on a day-to-day basis.

Insured Building	Location	Property Guideline	Property Actual	Workers Compensation Guideline	Workers Compensation Actual
ABC Bank	New York, NY	\$75 million	\$84 million	100 employees	52 employees
XYZ Tower	New York, NY	\$75 million	\$35 million	100 employees	76 employees
123 Plaza	Chicago, IL	\$100 million	\$78 million	150 employees	0 employees
456 Plaza	New York, NY	\$75 million	\$45 million	100 employees	112 employees
ZZZ Hotel	Los Angeles, CA	\$100 million	\$50 million	150 employees	58 employees

IMPROVED EVENT RESPONSE

Immediately following the September 11, 2001 terrorist attacks, few insurers and reinsurers were able to determine their exposures in the specific buildings that were primarily affected by this event: 1–7 World Trade Center, 3 World Financial Center, 140 West Street (Barclay-Vesey Building/New York Telephone), and 130 Liberty (One Bankers Trust Plaza/Deutsche Bank). Even for insurers and reinsurers writing property lines of business, where typically the highest quality location-level exposure information is found, the geocoding resolution was not high enough to determine the exposure for each of the buildings affected. Quantifying exposure for lines of business other than property, such as workers compensation, was even more problematic because location level employee data was, for the most part, non-existent. Thus, any attempt to quickly identify multiple lines or multiple account building-level accumulations was impossible. Improved data collection, in conjunction with building-level geocoding, can significantly improve how companies respond to events.

Hypothetical Fire at the Key Tower in Seattle

Sanborn data improves the efficiency of generating analyses such as accumulations at a building level. In February 1991, One Meridian Plaza in Philadelphia suffered a severe fire. The fire started on the 22nd floor, and raged for more than 19 hours, gutting eight floors and causing an estimated \$100 million in direct property loss. It was later described by Philadelphia officials as “the most significant fire in this century.”

If a similar fire were to occur in another skyscraper, how could Sanborn data facilitate the process of identifying a company’s total exposure?

The Key Tower in Seattle is a 62-story building built in 1990. Its primary address is 700 5th Avenue, but Sanborn data also identifies four other valid addresses:

1. 523 Columbia Street
2. 525 Columbia Street
3. 717 6th Avenue
4. 715 6th Avenue

In the absence of Sanborn data, the best assessment of exposure in this building might consider only those locations coded at the primary address. Obviously, this could understate the true exposure. The RMS recommended approach would be as follows:

- Begin geocoding exposure information using both Sanborn CitySets[®] and GDT data to ensure high-resolution geocoding in urban areas. Exposure captured with any valid address or building name will be stamped with a unique building ID during the geocoding process, which will enable building-level analysis capabilities.
- Post-event, use RiskLink[®] software to perform a building-level accumulation analysis for the Key Tower to identify all locations (and associated accounts) that have been geocoded to the building-level. This is particularly important because an insurer’s exposure in a large office building such as the Key Tower can come from multiple lines of business (e.g., property and workers compensation), multiple coverages (e.g., building, contents, business interruption) and multiple tenants (e.g., financial institution, retail, government entity). The unique Sanborn building ID pulls all of these pieces together in order to generate a complete accumulation analysis.
- Post-event, use RiskLink software to perform specific area accumulation analyses using 100, 250, and 500 feet (30, 76, 152 meter) radii around the Key Tower to identify any adjacent building exposures. This may also indicate additional building-level analyses that should be performed.

APPENDIX: SANBORN CITYSETS® FACT SHEET

Coverage

Sanborn CitySets® data is available for 39 cities, with over 189,000 buildings, and a combined coverage area of approx. 161 square miles.

Sanborn CitySets® Cities	Square Miles	Buildings	Primary Addresses	Secondary Addresses*	Total Addresses
Arlington, Virginia	2.3	3,591	3,591	2,058	5,649
Arlington (PEN), Virginia	1.7	1,230	1,230	342	1,572
Atlanta, Georgia	3.4	2,090	2,090	2,606	4,696
Baltimore, Maryland	2.4	2,558	2,558	4,264	6,822
Boston, Massachusetts	2.9	3,694	3,694	8,032	11,726
Buffalo, New York	3.5	5,475	5,475	505	5,980
Chicago, Illinois	15.3	19,018	19,018	12,479	31,497
Cincinnati, Ohio	3.8	1,882	1,882	892	2,774
Cleveland, Ohio	4	1,618	1,618	145	1,763
Dallas, Texas	3.8	2,106	2,106	2,752	4,858
Denver, Colorado	2.2	1,874	1,874	2,314	4,188
Detroit, Michigan	2.9	1,365	1,365	2,649	4,014
Ft. Lauderdale, Florida	3.5	5,248	5,248	2,208	7,456
Honolulu, Hawaii	4	3,537	3,537	189	3,726
Houston, Texas	3.3	1,648	1,648	1,643	3,291
Jersey City, New Jersey	3.4	5,183	5,183	4,849	10,032
Kansas City, Kansas	3.7	1,544	1,544	396	1,940
Las Vegas, Nevada	6.5	3,291	3,291	1,305	4,596
Long Beach, California	3.2	3,868	3,868	1,347	5,215
Los Angeles, California	10.7	27,743	27,743	14,456	42,199
Memphis, Tennessee	3.7	2,471	2,471	208	2,679
Miami, Florida	3.9	2,755	2,755	2,328	5,083
Minneapolis, Minnesota	2	864	864	2,844	3,708
New Orleans, Louisiana	3.2	4,409	4,409	297	4,706
New York City, New York	12.3	21,651	21,651	33,318	54,969
Newark, New Jersey	3.6	7,044	7,044	3,993	11,037
Philadelphia, Pennsylvania	5.4	15,935	15,935	5,976	21,911
Phoenix, Arizona	3.4	3,882	3,882	198	4,080
Pittsburgh, Pennsylvania	3.8	4,086	4,086	647	4,733
Portland, Oregon	1.3	988	988	3,511	4,499
Sacramento, California	3.3	3,575	3,575	526	4,101
San Diego, California	2.1	1,337	1,337	2,533	3,870
San Francisco, California	5.8	8,603	8,603	22,863	31,466
Seattle, Washington	3.6	2,462	2,462	4,462	6,924
St. Louis, Missouri	4.9	2,164	2,164	4,162	6,326
St. Paul, Minnesota	1.1	489	489	1,347	1,836
St. Petersburg, Florida	2.5	1,716	1,716	1,695	3,411
Tampa, Florida	0.8	256	256	535	791
Washington, D.C.	7.5	6,484	6,484	11,211	17,695
Total	160.70	189,734	189,734	168,085	357,819

* Secondary addresses are addresses associated with buildings that have multiple addresses.

In general, the coverage area includes 4–5 square miles (10–13 square km) around the downtown area of each of the 39 cities. However, Chicago’s coverage area is 15 square miles (39 square kilometers) and Los Angeles, New York City, and Washington, D.C. have multiple coverage areas to capture non-contiguous regions with exposure concentrations outside of the traditional downtown areas.

Attributes

City centers often have buildings that have multiple addresses and/or are commonly referred to by a building name. Sanborn data captures all valid addresses for each building in the coverage area as well as any building name. The geocoding engine used in RMS products applies this address information to enable geocoding at the building level.

Sanborn collects data using aerial imagery, existing Sanborn maps, and field surveys. The data collected using aerial imagery includes building footprint and building shape information. Building height information is also determined from the aerial imagery, which Sanborn uses to derive the number of stories (assuming 12 feet/3.6 meters per floor). In addition, Sanborn derives the estimated square footage by using the building height, footprint, and shape. These attributes exist for all Sanborn buildings. Other data elements included in the Sanborn data are construction, occupancy, and year built.

Construction

Sanborn collects this data by surveying the building’s exterior. It is often difficult to gather more detailed construction information, such as ATC construction class, via this type of survey. Thus, the construction types Sanborn uses are masonry, non-masonry, steel frame, and unknown. They also collect fire protection class from Sanborn maps and more detailed exterior construction materials identified during the field survey. Using the additional attribute information, RMS developed a methodology to derive construction types. For example, if Sanborn classified a building as commercial occupancy, steel frame construction, non-combustible fire protection class, built before 1940, and 31 stories or greater, the RMS-derived construction would be braced steel frame (ATC construction class 9). For many of the buildings, particularly the smaller ones, RMS assigned an unknown construction type to reduce the likelihood of inaccurate data. 98% of the buildings have a known construction assignment, and 49% of the buildings five or more stories tall have a known construction class.

Occupancy

The Sanborn data includes ATC occupancy codes for 97% of the buildings.

Year Built

Year built was captured from the Sanborn Fire Insurance Maps. In addition, year built and year range information was collected during the field surveys (if the latter, the ranges used were pre-1940, 1940–1970, post-1970, pre-1970 or post-1940). Year built is a key component of RMS vulnerability models. In addition, the year ranges are also useful in determining derived construction. Note that while only 10% of the buildings have a known year built, 93% of the buildings over 5 stories have a known year built. Also 63% of buildings have either year-built or year-range information available.