White paper

pitney bowes

Location Intelligence

Understanding recent innovations in geocoding

Master Location Data and pbKey™: Twelve Q&As for data architects

Organizations require location-specific data. However, the available data is a real challenge: It comes in different proprietary formats and data structures. Each requires different inputs and handling. Combining multiple sources and varieties of data effectively has proven costly, time-consuming and error prone.

Master Location Data (MLD) and pbKey™ change the whole dynamic.

- MLD is the most complete and precise universe of US address data Pitney Bowes has ever assembled. It's available as a locked binary encoded version for use inside our geocoder or an unlocked open version that can be inserted into a database.

- pbKey is a unique identifier assigned to each physical addressable location within the MLD which allows users to more easily manage their address data and unlock a wealth of information linked to it.

Together these tools provide improved data coverage, precision and agility so organizations spend less time assembling data, and more identifying insight.

This white paper outlines how MLD and pbKey work to bring new precision. It provides answers to the questions that data architects might be called on to answer.

Gain new precision in marketing, risk analysis, planning and engagement.

There is an incredible volume of spatial data available: Your organization needs better ways to harness this mountain of data for marketing, planning and evaluation of opportunities and risk. To help them do this in the most efficient and cost-effective way, Pitney Bowes developed Master Location Data (MLD) and the pbKey™.

MLD is a comprehensive US address dataset that can be used in conjunction with other spatial data to support pricing, modeling and analytics. Each US address can be instantly coded and enriched with attributes such as demographics, proximity to hazards, availability of services and other property information. All these attributes and many more can be accessed instantly using the pbKey assigned to each location.

MLD is able to provide our highest quality geocoding because it contains:

- Virtually all deliverable US postal addresses
- Non-postal street address in smaller and/or gated communities
- Multi-unit dwellings
- Rural-route addresses
- "Hidden addresses" such as basement apartments
- Units in co-op buildings
- Non-USPS addresses
- Addressable parcels with no buildings

MLD, geocoding and precision

01. What is geocoding?

Geocoding takes location information, such as a street address and transforms it into latitude/longitude coordinates. Then, it can be used for operational processes, analytical applications and mapping purposes. Geocoding is done via a coordinated blend of software and data. It can be performed on a single location description at a time or across many addresses at once.

Not all geocodes are the same. Some are based on ZIP codes. Others are derived from street-level data. There are geocodes that indicate the precise coordinates of the center of a property. The most precise "point level" geocodes mark the center of a building (rooftop) or even its front door.

02. What is Master Location Data (MLD)?

MLD is a US geocoding dataset. It's multi-sourced: We identify and select the best, most precise location available for each address.

03. What makes MLD our largest, most complete US geocoding dataset?

MLD includes unit numbers for commercial buildings and residential apartments. It also includes millions of non-postal addresses not served by the USPS or phone services. In total, this amounts to approximately 180 million+ unique property locations. Records will include tens of millions of alias addresses, created, for example, by vanity addresses, expanded ZIP codes and changed street names (e.g., 123 Main also becomes known as 123 Martin Luther King Blvd.)

Innovations in Pitney Bowes geocoding enhance quality and value.

Faster

Process hundreds of millions of records per hour.

More precise

Achieve a higher percentage of point-level geocodes.

More accurate

Improve match rates by up to ten percent or more.

Richer

Gain insight to specific details on properties and people.

Accessible

Embed Location Intelligence into every business process.

04. How does geocoding use Master Location Data (MLD) to efficiently deliver more precise results?

Before MLD was created, organizations requiring the highest match rates and the most accurate coordinates were forced to run multiple US geocoding data products in parallel. Different input datasets varied in completeness and precision, so organizations had to license numerous geocoding datasets and invest a great deal of time and effort in defining and applying their own logic to choose what appeared to be the best results. This allowed greater room for error, and slowed access to answers.

MLD does the work.

With MLD, Pitney Bowes has taken on responsibility for building the rules that select the most complete coverage and best locations available. MLD utilizes a massive, multi-sourced pool of data to deliver our most precise geocoding results. By culling and combining many address and location datasets, utilizing our existing geocoding assets, developing our own intellectual property and utilizing the rich assets from Pitney Bowes' mailing business, we can deliver better results with less effort. This means organizations can trade steep implementation costs for the small premium we charge for MLD.

05. How is the quality of MLD quantified?

The quality of MLD is best explained in terms of two factors: match rate and geographic precision.

Match rate. MLD delivers a better match rate because it exceeds the inputs of other data products on the market:

- More address points
- More complete geographic coverage
- An innovative data-management approach that makes it easier to identify and isolate location information within both structured and unstructured data

Geographic precision. MLD delivers better geographic precision because:

- The 'golden point' or most precise coordinate available for every address is selected from the "best of the best".
- Any coordinates that are 'interpolated' or calculated are the result of advanced, proprietary Pitney Bowes algorithms tested and retested for quality.

A full 88 percent of the records in MLD are at 'parcel-centroid or better', with 'better' signifying structure centroid or front door.

Understanding the match/precision challenge

Here's an example of why both match rate and precision are critical.

Let's start with an input address:
3001 Summer St, Stamford, CT 06926

A street geocoder will find Summer St in Stamford, CT 06926, then use the address range information embedded in the street data to interpolate (make an educated guess) where the building 3001 is likely to be. This approach has two shortcomings:

- **An interpolated coordinate may not be as precise as necessary for the business problem it is meant to solve.** Financial Services companies, for example, need precise information for underwriting, yet an interpolated coordinate can easily be misplaced if the street-range information is wrong or complex streets shapes such as cul-de-sacs are involved.

- **The geocoder does not validate whether the street number actually exists.** If *3001* had been mistyped as *3002*, an address that does not actually exist, the geocoder would still generate a result. It might be very close to *3001*, but it might also be somewhere entirely different along the street. We call this the "False Positive" problem.

An "address point" approach overcomes these issues. In it, all of the input address elements are matched against a list of known addresses. The returned address is the one that matched the best, and the coordinate comes directly from that list. With this approach, the chance of false positives is dramatically reduced and the returned location is more precise.

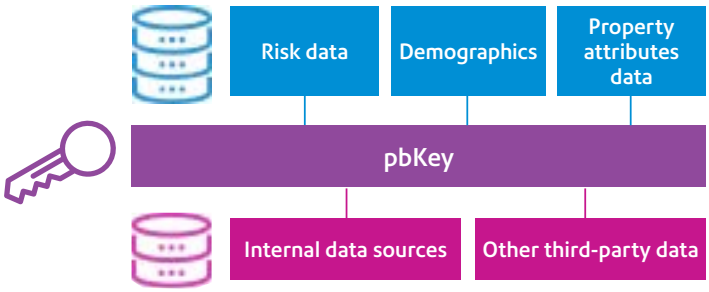
In our geocoding, the address point approach is always the first attempted, and the vast majority of geocodes are assigned this way. When an address isn't found in the data, then a "fallback" process finds the next best level of geocode, usually an interpolated result from the "Street Geocoding Dataset".

The MLD dataset is the product of this rigorous approach. It is continuously being refined. As a result, if an address isn't found within MLD, there is a high probability that the address is very poorly formed, it's a brand-new build, it has recently changed or it doesn't really exist.

The power of the pbKey™

06. What is the pbKey and why is it attached to every record in MLD?

The pbKey is the unique identifier attached to every address record in MLD. It will not change (it will persist) from version to version, so it can be used to reference the address without storing the whole address string.

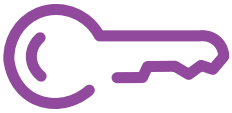


pbKey enables companies to use data better: It enables them to easily tap into internal data as well as data from Pitney Bowes and third-party sources.

Like a personal social security number or unique company identification number, the pbKey is definitive. With it, users can:

- Enrich data with additional attributes such as property information, risk data and demographics.
- Simplify address management across multiple systems.
- Streamline information sharing between companies.
- Connect data across multiple datasets for predictive analytics.
- Determine relationships between addresses.
- Perform change analysis on the impact of any new information related to the address.

Like a personal social security number or unique company identification number, the pbKey™ is definitive.



Get data coded for confidence.
Each geocoded address in MLD is assigned a location code. This location code indicates the level of precision (building centroid, parcel centroid, etc.) we're able to achieve for the address. It helps to more precisely describe the logic and confidence behind the resulting geocode.

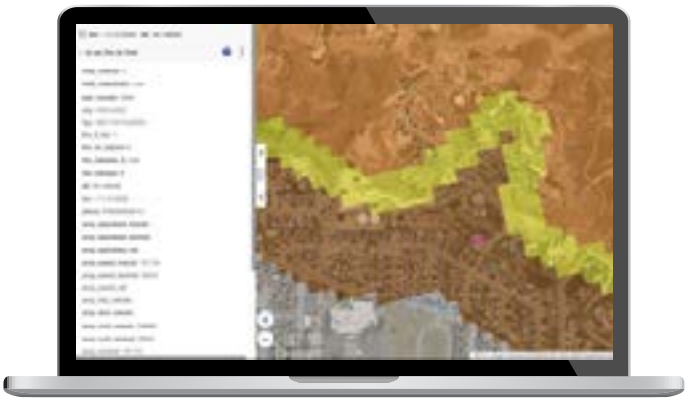
The algorithm and geocoding process are continually reviewed and augmented as necessary to improve quality and confidence.

07. What data is attributed to each PbKey?

pbKey unlocks a wealth of rich, attributable information about each location. For example:

- A communications service provider could quickly assess coverage for a new customer or a new market, use findings to optimize tower placement, and more.
- An insurer can gain rapid access to key underwriting data, such as detailed property characteristics as well as individual and cumulative views of risk.

There is virtually no limit to what can be added as new information becomes available.



pbKey unlocks a wealth of rich, attributable information. Example: MLD provides input for mapping of wildfire risk.

Example: MLD informs mapping of drivetime distances, showing the properties within a 10-minute drive from a local fire station.

GeoEnrichment

08. What is GeoEnrichment?

GeoEnrichment is the process of augmenting address data records with additional information. It uses the geocode coordinate to relate the address to one or more spatial data files such as demographics, property detail, proximity to hazards, availability of services, and more.

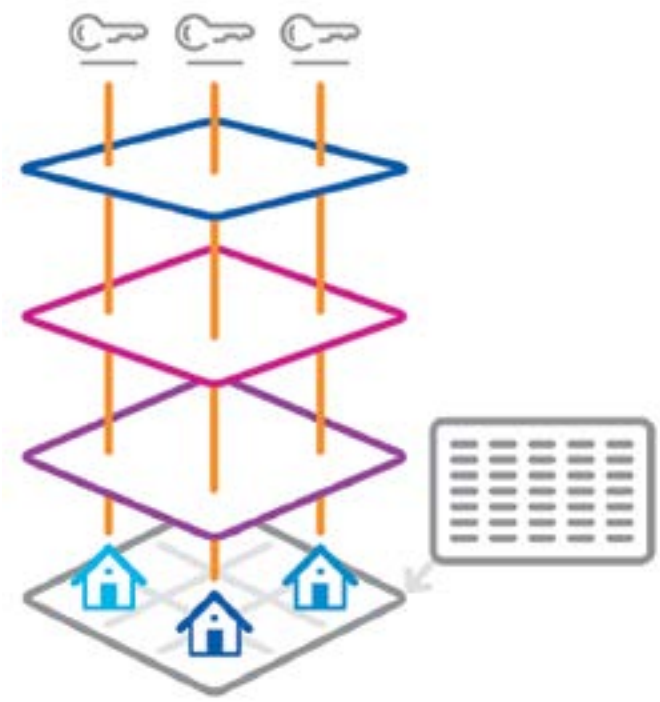
In the past, this has been a laborious and expensive process, requiring extensive use of computing power, tools and reference data. Now, the pbKey™ can instantly link to additional datasets and attributes.

09. What is the GeoEnrichment Module and how does it work with MLD geocoding?

Spatial processing can be slow. The datasets can be very large to handle and expensive to acquire. The GeoEnrichment Module makes this processing affordable, fast and easy:

- All spatial analysis is performed in advance.
- The pbKey provides access to a pre-processed database of attributes linked to each address.

GeoEnrichment is the process of augmenting address data records with additional information. It uses the geocode coordinate to relate the address to one or more spatial data files such as demographics, property detail, proximity to hazards, availability of services, and more.



Master Location Data and GeoEnrichment improve decision making by delivering insight-ready data instantly.

MLD and the MLD Address Fabric

10. Why are there two forms of MLD, and how do they differ?

Pitney Bowes offers MLD and MLD Address Fabric, two solutions based on the same powerful technology. These two different formats exist to address two different types of use cases:

MLD is a richly multi-sourced dataset. It delivers industry-leading match rates and has the highest precision coordinates of any of our geocoding products.

The MLD dataset is useful when you have a specific address or set of addresses you want to understand better. Like all other address point datasets, it's delivered as an encoded and encrypted binary file that is only accessible by our geocoder. You can only get the benefit of MLD if you already know the address or its coordinates.

The MLD Address Fabric is the "open" version of Master Location Data. It gives users access to 180 million+ geocoded unique addressable locations which can be used for a wide range of analytical and marketing purposes.

MLD Address Fabric is comprised of the same data as MLD, but it's delivered as a file that can be opened in MapInfo Pro™ software, loaded into a database or used in an analytics environment.

11. Why would the MLD Address Fabric be bought?

There are a number of ways that the MLD Address Fabric could be used, including building your own geoenriched dataset: Many organizations already have their own address-linked or spatial datasets and need to use those to build their own geoenriched dataset. With the MLD Address Fabric, they can apply their own reference data to augment each record. By pre-scoring every US property, augmentation and calculation becomes much simpler.

MLD is a rich, multi-sourced dataset that delivers industry-leading match rates.

MLD Address Fabric is the "open" version of Master Location Data giving users access to 180 million+ address locations.

12. How can MLD and the MLD Address Fabric help in a big data environment?

Big data capabilities add speed, volume and affordability to the processing of large volumes of data.

The pbKey provides a simple way to access data for processing in a big data environment.

- It's persistent, so it's a reliable way to relate data when required, without the need for hierarchical organization.
- It makes it easy to GeoEnrich data with additional attributes without reformatting the data.
- It can be used to keep track of an address over time and to create an audit trail of changes made to the address.

With the use of MLD and the pbKey™, organizations can overcome traditional challenges:

- The data may change, but the persistent nature of the pbKey makes it easy to stay connected to the best version.
- Data formats may vary; however, it's easy to link the data using the pbKey as the common denominator.
- Providing speed and scale, MLD and the pbKey work easily with spatial data lakes and big data solutions:
  - Apache HBase with Apache Phoenix secondary indexes
  - Amazon Aurora
  - Hadoop databases such as Hive
  - Other NoSQL databases

Locate success with Pitney Bowes.

Gain precision in planning, marketing and risk analysis. From comprehensive data quality to robust Location Intelligence solutions, Pitney Bowes can help you achieve a more comprehensive, actionable view.

United States
3001 Summer Street
Stamford, CT 06926-0700
800 327 8627
pbsoftware.sales@pb.com

Europe/United Kingdom
The Smith Centre
The Fairmile
Henley-on-Thames
Oxfordshire RG9 6AB
0800 840 0001
pbsoftware.emea@pb.com

Canada
5500 Explorer Drive
Mississauga, ON L4W5C7
800 268 3282
pbsoftware.canada.sales@pb.com

Australia/Asia Pacific
Level 1, 68 Waterloo Road
Macquarie Park NSW 2113
+61 2 9475 3500
pb.apac@pb.com

For more information,
visit us online: pitneybowes.com

17SWS03532