

Industrial-strength Geocoding

Nine requirements for cost-effective business decisions

Overview

Every business entity and government agency deals with address information. A prospective customer, for example, may request service at Primăverii 7, Sector 1, Bucureşti. But what, exactly, can you tell from an address? Can you provide service at this location? If so, does your prospective customer qualify for a specific offer or special rate?

Addresses help Poșta Română deliver mail; they also help other carriers, for example to deliver packages and parcels. They do not, however, indicate important geographic-related attributes and characteristics that can impact risk, customer satisfaction and profits. For example, organisations need an accurate, automatic way to:

- Assign customers to territories for sales, service or pricing
- Determine the distance from an office or service junction
- Identify the census tract & parcel number for regulatory compliance
- Calculate the proximity to a coastline, hurricane path or nearest fire station

Today, geographic-based data drives social policy, network planning and market analysis. In business, these insights are essential for risk management, regulatory compliance, pricing and strategic planning. That's why so many organisations employ geocoding. Geocodes translate common reference points, such as customer addresses, into latitude and longitude coordinates that makes it easier to analyse data.

Although some people think of geocoding only in the context of maps, geocoding is also the enabling technology for spatial analysis—and makes it possible for organisations to determine the relationship between two or more locations.

With accurate latitude and longitude coordinates, organisations can calculate the distance between two points, compute the distance to a specific

boundary and determine if an address is situated within a zone or territory.

Unfortunately, most databases do not contain geocodes as part of the associated address data. When they do, there may be a high degree of uncertainty in the quality of those geocodes, which can lead to faulty business decisions and costly mistakes. On aggregate, these mistakes can lead to lawsuits, attrition and significant losses.

Components of Industrial-strength Geocoding

Geocoding is very complex. There are several low-cost consumer-oriented applications that offer geocoding solutions—but these tools were designed to support general consumer needs.

Industrial-strength geocoding, on the other hand, is meant for problem solving. When organisations take a close look at the risks associated with a poor or inconsistent approach to geocoding, they soon find that accuracy, flexibility, functionality and experience can make all the difference.

Most business-oriented geocoding solutions consist of three primary components: address data standardisation and validation, geocode determination and data enrichment.

- Just as Poșta Română cannot deliver an incorrectly addressed envelope to its rightful destination, valid, well-defined addresses are important components of geocode accuracy. There are international conventions and guidelines governing punctuation and abbreviations, format, and address components for address standardisation. The first step in augmenting an address with a geocode is to standardise and validate that address.
- When it comes to geocode determination, there will be times when it is not possible to deliver a geocode centred on a specific address. Industrial-strength tools will recognise this and apply consistent rules, automatically cascading to the next most-specific point of reference. In most cases, no geocode is better than a wrong geocode.

- Lastly, you will need an ability to append additional data sets, including demographics, purchasing preferences and lifestyle data. Plus, an ability to append spatial data derived from your analysis – such as territory assignment or flood zone determination. The best industrial-strength tools will integrate all three of these components into a single solution, providing the most accurate results possible. In many cases, organisations can take advantage of on demand and cloud-based solutions that provide “pay-as-you-go” access with no overhead expenses.

Challenges with Geocoding

Short of placing a global positioning system (GPS) at an address to determine its true longitude and latitude, there are several approaches to geocoding. These range from less precise methods based on postcodes, all the way to parcel centroid level geocoding. Organisations interested in doing more with location-based data need to understand the challenges associated with geocoding.

Postcode Limitations

Postcodes are not necessarily stable over time. While a location’s latitude and longitude will never change, Poșta Română will often modify a location’s postcode. Approximately 600 of the 37,000 postcodes in Romania have changed recently. On top of that, many homes and businesses use addresses that are not recognised by Poșta Română for mail delivery: typical examples would be Piața Iancului Metrou or Complex Nord lângă Simigerie.

Among the potential problems associated with postcode geocoding, is that a postcode may span other administrative boundaries. Also, some areas, such as Păltineni from orașul Nehoiu in Buzău only have mail delivery to the local post office but not to individual homes because there are no street names or numbers. There are also individual post-boxes (căsuțe poștale) for both people and organisations (e.g. Biaroxmin SRL) that just have a post-box address but no street address. A further complication is to know that, for example, P.O. Box 98 corresponds to Str. Ecaterina Varga Nr.23 in Brașov. A more complex situation is exemplified by Serviciul de Informații Externe located at Șos. București - Ploiești, nr. 280-282, sector 1, București, but with two items of PO information: Oficiul poștal 41, and căsuța poștală 59, București.

At the other extreme are large buildings in metropolitan areas. For example, along Bulevard Republicii in Resita, there are buildings which have

separate postcodes assigned to different entrances (scări) in individual blocks.

However, a robust address matching system will take account of these various special cases.

Varying Degrees of Accuracy

Not all geocodes are the same. Parcel (parcele) centroid geocoding places the coordinates at the centre of the parcel associated with the physical street address (these are often referred to as ‘point-level’ geocodes’). In Romania, the next level up is the so-called address-level geocode (ADD) which may approximate to the entrance of a building: in the US, these are called ‘roof-top’ geocodes. Next comes Street Segment (SSC) geocoding which is based on knowing the geocodes at nearby intersections and then estimating the geocode of that address by interpolation. For example, if a street segment runs from Strada Principală nr. 201 to 581 in satul Moieciu de Jos, then nr. 330 would be approximately halfway between the two end points. A slightly lower level of accuracy would be street centroids (SC) where there are no intersections along the street.

Since postcodes were designed to quickly sort and deliver mail, they are not nearly as accurate as Address-level (ADD) or Street Segment (SSC) geocodes, or even parcel centroid geocodes. Postcode-based geocodes assign the latitude and longitude associated with the centre point of a postcode and only approximate the location of any particular address. For example, geocodes based on Romania’s six digit postcodes will span an area much greater than those derived from nine-digit postcodes in other countries. Even six digit geocodes vary in their level of accuracy, for example: in București, a postcode may represent approximately 60 households on the same side of a street (block-level accuracy), while in a rural area, a postcode may have up to 400 households.

The situation is further complicated in Romania where postcodes don’t always cover a contiguous area: for example, in Bacău there are two areas of the city separated by approximately 1 km but with the same postcode: this is an extreme case of a very common problem in e.g. București where there are many examples of a single postcode being assigned to two locations separated by intermediate postcodes.

The situation can be even more complicated at higher levels of geography: for example Municipiul Sibiu includes Păltiniș which is located

approximately 30 km remote from the actual city of Sibiu.

Multiple Data Sets

As can be seen from the examples cited, Geocoding engines that rely on a single data source cannot possibly deliver accurate geocodes. In addition to the Poșta Română data, one or more spatial data sets are needed to determine latitude and longitude coordinates as the spatial data sets also contain address information for non-Poșta Română recognised addresses such as railway stations and metro stations, and also for the many streets that are not listed by Poșta Română.

Several data vendors provide the required spatial street network data (Navteq, Tele Atlas, Geo Strategies, and others). Geocoding tools must be able to parse, standardise and match the input address against all address dictionaries, score all potential match candidates, and only then determine the best match.

Sequential Processing

Some geocoders do not integrate Poșta Română data with third-party spatial data sets. They simply apply different data sources and address dictionaries sequentially, and stop when a match is found. This type of sequential processing leads to errors, as the tool may identify a match through one data source while a more accurate match is available elsewhere. This is known as a “false positive”. Geocoding tools should be sophisticated enough to analyse multiple data sets as a single address dictionary and then compare an address against all potential geocode candidates in a single pass.

Out-dated Data

Given the pace at which postcodes change, new street names are assigned and new addresses created, the data driving any geocoding solutions needs to be updated frequently. It’s imperative that your geocoding solution can support regular updates to provide the highest number of matches and the most accurate results.

The number of physical addresses continues to grow as a result of new developments and the conversion from Rural Route and PO Boxes to house number style addresses.

However, due to the development schedules and priorities of the data providers, spatial street vector and address range coverage may be incomplete. In cases where multiple matches cannot be resolved (for example, București, Șoseaua București 10 may be Șoseaua București-

Ploiesti from București, or it may be Șoseaua București-Nord from Voluntari, Ilfov: Șos. București -Nord sometimes occurs in databases for București and, furthermore, it isn’t recorded by Poșta Română in the current version). Most geocoding software reverts to postcode geocoding - if a postcode is present- when address-level geocoding is not possible: however, in the event of no postcode being provided, an address such as illustrated above would have to be reported as an ambiguity.

While these challenges exist, however, there are easy, effective ways to avoid potential problems. Given the value that an accurate geocode provides, organisations would be well served to search for geocoding solutions that can overcome these obstacles.

Geocoding solutions turn addresses into latitude and longitude coordinates, making it easy for organisations to understand the relationship between two or more points. With an accurate geocode, organisations can gain insights, perform calculations and automate processes. These relationships- uncovered through spatial analysis- are generally classified into three categories: distance between two locations (e.g. nearest store), distance to a boundary (e.g. coastline), and determining if an address is situated within a zone (e.g. tax district). The last category is also referred to as a “point-in-polygon” application. The polygon is defined by the latitude and longitude coordinates of a series of line segments that form the polygon.

The specific value that geocoding provides an organisation varies from industry to industry. For example:

All Industries

Geocoding helps analyse data in the context of location so organisations can make smarter decisions regarding targeting, networks and risk.

- Marketing – targeting, segmentation and modelling
- Field service routing – installation, maintenance & repair services
- Sales territory assignment
- Tax jurisdiction assignment

Property & Contents Insurance

Risk varies greatly based on location, and even tens of metres can mean the difference between a profitable policy and a costly claim.

- Risk management – proximity to landslides,

earthquakes fault lines and other hazards:

- * likelihood of terrorism, fires, wind storms, etc.
- * located in or near flood plains or areas prone to flooding
- * distance to nearest fire station and fire hydrant
- * overall risk accumulation/aggregation
- Territory and rating determination
 - * Pricing
 - * Driving distances
 - * Straight-through-underwriting
 - * Group policies

Municipalities & Public Sector

Geocoding and mapping tools have applications across government agencies.

- Constituent services
 - * Eligibility for services and programs
 - * Voting precincts and polling locations
 - * Nearest office, government resource
 - * Crime mapping
- Tracking trend data to anticipate needs
- Economic development and urban planning

Telecommunications/Utilities

With networks that cross traditional geographic boundaries, location-based insights add value to nearly every functional group.

- Service eligibility
 - * Accurate location of reported bad coverage
 - * Service availability/coverage locator
- Network management
 - * Asset management
 - * Network design and maintenance
 - * Call before you dig

Financial Services

Geo-based market segmentation plays a critical role as proximity to branch and competitive branch locations can affect both sales and profits.

- Branch expansion
- Performance goals by branch and product
- Sales potential of new products for each market and channel
- Fraud detection and risk scoring
- Mortgage protection

Healthcare

Healthcare must be accessible, so companies need to demonstrate that they can adequately cover members and provide quality care.

- Mapping provider networks against member addresses

- Gap analysis
 - * availability of specialists
- Fraud detection
- Health monitoring—including epidemics across geographies

Shipping Logistics

Geocoding provides a clearer, smarter picture of just how to deliver goods to customers, with an ability to append weather, traffic and construction updates and proximity to important locations such as distribution centres or stores.

- Determining the true territory or zone an address is in
- Plotting how addresses are located relative to each other
- Determining ways to optimise all aspects of delivery from point of origin, to loading of goods, to routing deliveries, to tracking of items in transit

Retail/Real Estate

Location impacts most every strategic decision facing organisations in the retail and real estate industries.

- Site selection
- Market expansion
- Managing store performance
- Trade area determination

Top nine requirements

The nine ‘must-haves’ for optimal geocoding

Given the wide-spread impact that geocoding can have on business decisions, organisations need to verify that their geocode technologies measure up to industry best practices.

Nine factors in particular can determine how cost-efficient, effective and consistent your solution performs.

1. Integrated Address Quality

Accurate addresses are a prerequisite for accurate geocoding. Leading-edge solutions will offer the ability to cleanse data, standardise addresses and validate that source addresses are correct before applying geocodes.

Benefit of SMARTaddress Romania

While some geocoding vendors have partnerships with third-party name and address data quality vendors, it is more effective when these capabilities are integrated in one platform. This way you can standardise and correct address information in a single pass using multiple parsing and matching algorithms giving you the ability to potentially resolve and standardise addresses a stand-alone algorithm might have rejected.

Having to conduct two or more separate passes (against multiple data sets) will also slow down data processing and potentially introduce erroneous matches and geocodes.

2. Beyond Poșta Română Data Sets

Poșta Română data by itself, however, is just the start. It is vital that your geocoding solution can standardise and validate addresses against data sets that combine the most current Poșta Română (PR) and non-PR address data in a single, standardised address dictionary.

There are many households and organisations that receive mail at a Post Office Box (căsuța poștală); so the Ionescu family may receive mail at PO Box 123 even though their actual physical location is at Șoseaua București-Nord, 31, Voluntari, Ilfov. For example, in the current official lists from Poșta Română there is no record of Șos București-Nord even though it exists in street segment databases, such as from Navteq or Tele Atlas.

By combining information from Poșta Română with a street segment database, one benefits from a

more complete universe of address data that provides more accurate matches. This process is known as conflation. Since any address verification and/or geocoding technology requires “fuzzy” matching – realising the input addresses are not always perfect – this can lead to “false positives.” For example, if Șos București 10 is entered and the Poșta Română file only shows Șoseaua București-Ploiești, the software will assume this is correct. By conflating Navteq data with Poșta Română data, the more complete universe will show that there is both a Șoseaua București-Ploiești and a Șoseaua București-Nord (which is actually in Ilfov, but often misrepresented as being in București). In this case, no match would occur because the match is ambiguous, which is better than making the wrong match.

3. Validated Geocode Results

Analysing geocode results based on positional accuracy doesn’t provide the full insight needed to make critical business decisions. Even when source addresses are fully validated, the geocoding process needs to ensure that the address is located at the right spot. Some solutions return geo-coordinates without providing any details regarding the accuracy level—even when no close match is found.

Geocoding tools should return detailed match codes that indicate the portions of the address that match the source data, as well as detailed location codes that indicate the accuracy of the assigned geocode. These codes can then be incorporated into rules-based processes and automated decision making. For example, a commercial insurance company may have a rule that says when the value of an insured risk such as a property exceeds €1M, there must be an exact match to the street name (including directionals) and the geocode must be accurate at the street level or better—otherwise it will be rejected for exception processing.

4. Multiple Levels of Accuracy

There will be times when it is not possible to deliver a geocode centered on a specific address or parcel. Based on the available data, the solution should be able to return latitude and longitude coordinates at successive degrees of precision, including parcel centroids, building-level interpolation (ADD), street address interpolation (SSC), street centroid (SC),

postcode centroid (PC), locality centroid (LC), as well as geographic centroids such as city and county. The tools you use should recognise this and apply consistent rules, automatically cascading to the next most-specific point of reference. This approach ensures the best available information.

5. Data Flexibility

One should expect the flexibility to handle different types of databases and input addresses, including structured/unstructured, residence/business, etc. Most importantly, organisations need the flexibility to utilise and switch among multiple data sources. In addition to the conflated address dictionaries highlighted above, organisations may also have proprietary data sources. For example, a company may own street data or point of interest data in one town, or, may have access to data from a local municipality. These custom User Dictionaries should work in tandem with Poșta Română and street segment data sources as the goal should always be to provide for the best decision possible.

6. Data Enhancements

The ability to support data enrichment is also important. Addresses and geocodes can be the keys to files containing additional data that could be added to customer records, including census bureau demographics, income levels, purchasing preferences and lifestyle data. Several third-party sources provide demographic and lifestyle data such as Mosaic (Experian segmentation) that can be appended to individual consumer or household records. The lowest level about which the National Office for Statistics (INS) publishes demographics is the Census Tract level. At the previous census, there were nearly 15,000 Census Tracts, which typically encompass a city block in urban areas, but could be several square kilometres in rural regions. Census Tract data includes number of residents, number of housing units, median age and median age by sex, number of residents in various age brackets, number of households, average household size, number of households with children, median household income, family income, along with several other fields. Marketing organisations can incorporate this data into their spatial analysis in order to determine where to open new stores, conduct targeted mailing campaigns, or advertise products and services.

7. Geo-confidence

As each geocoding technique is subject to varying degrees of precision, best-in-class geocoding

solutions not only return the latitude and longitude, they also measure the degree of confidence in the results and the maximum possible error.

When using street segments, for example, the maximum possible error would be the physical length of the street segment on which the address is located. This might be 250 metres in București to several kilometres in some rural areas. A desirable feature of any geocoding product is the ability to generate a “geo-confidence surface” or buffer area that can be used to define, and perhaps with the help of mapping software display, the smallest possible region in which the address must exist. These buffers represent more than a simple circle with the most probable location at its centre, rather they are custom polygons based on the level of accuracy that can be derived from the data.

When used in conjunction with point-in-polygon analysis, organisations can assess the probability of whether an address falls in one area or another. If a buffer area spanned multiple floodplain boundaries, for example, the tool should indicate the percentage lying within each floodplain and outside of it.

Some geocoding systems have the ability to use confidence surfaces to indicate the maximum and minimum possible distances between two locations—such as distance to a central office, or the closest and furthest possible distance to a coastline.

Products that provide these capabilities have a distinct competitive advantage over those that do not. Suppose for example that you have a low-quality geocode (e.g. one based on a six-digit postcode alone). If you calculated that entire ‘geo-confidence area’ fell within your target zone, you could still be 100% confident in the end results.

8. Multiple Deployment Options

While some business situations can be satisfied by assigning geocodes to a group of addresses in a batch environment, other situations require this determination be made to an individual address in real-time. A product that can be deployed in both batch and real-time modes serves a broader range of business applications—including future needs. With SaaS (Software as a Service), typical integration efforts may take hours instead of days. With no investment in servers, no on-going maintenance and no software updates to install, organisations can save time and money month after month. Benefits include:

- Lower total cost of ownership
- Quicker time to value
- Flexibility to rapidly scale based on market opportunities
- Automatic access to the most current—and most accurate—information
- More efficient collaboration

9. Experienced, One-stop Service

While simple in concept, geocoding applications can have subtleties that are not obvious to those implementing them for the first time. Vendor expertise, including a track record of successful implementations across a variety of business cases, can prove invaluable both during the planning and implementation stages. Experienced vendors can often add value by suggesting the appropriate data augmentation databases and providing suggestions as to how other clients with similar issues have utilised geocoding technology. A vendor's financial stability should also be examined in determining future support and product stability and reliability.

Overall, solutions need to be simple to use and flexible enough to meet different business requirements. A single technology platform that matches up with your overall corporate objectives can help ensure that a consistent standard will be applied in every market. Likewise, maintaining one platform reduces cost of ownership and can speed up system integration. A single interface also simplifies training and education, and makes it easier for your company to gain the skills and capabilities needed to achieve a competitive advantage.

Even if two vendors are using the same underlying geocoding database, match rates will vary due to differences in parsing and matching techniques. Conducting a benchmark designed to reflect your business requirements will provide valuable information in helping you select one technology over another.

Conclusion

More than 70% of all business records include a location component, so it is not surprising that location-based analysis is in such high demand nowadays. While a physical address is of paramount importance for postal delivery purposes, it does not tell you where that address is located relative to other locations, such as a sales territory, flood plain or specific tax district—or distance to the nearest office.

Spatial analysis serves as a basis for business decisions in marketing, operations and a wide

variety of industries such as insurance, telecommunications and the public sector. But before you can analyse, extrapolate or profit from location data, you need to associate each record with an accurate latitude and longitude coordinate—a process known as geocoding.

Geocoding can be complex, and mistakes in geocode assignment can lead to poor business decisions that impact risk, relationships and profits. Organisations need to consider several factors when choosing a geocoding solution, including accuracy, flexibility, functionality and experience. Today's leading enterprise geocoding tools combine address quality, multi-level geocoding and spatial analysis in a single solution that can apply the same standardised rules in every market— so you can act with confidence.

About Geo Strategies

Geo Strategies was founded in 1993 and has become the leading supplier of geo-spatial information and tools for Romania.

Their core expertise is built around geo-spatial data products, analytical and modelling tools, consumer segmentation, bespoke data services, consultancy, training, and project management.

This expertise has been developed according to internationally recognised best practices and methodologies, to which significant innovation / value is being added to make it relevant and actionable in the local context.

Geo Strategies partner with Pitney Bowes, Experian and Navteq (part of the Nokia group) to provide the best-in-class products for data quality management (data cleansing, manipulation and integration), socio-demographic profiling and targeting and world-standard mapping for GIS and other spatial applications.

From services in data cleaning and enhancement, data integration to Mosaic consumer segmentation, Micromarketer area analysis and profiling, to the application of customer driven insight and targeting - we can assist.

Acknowledgements

Geo Strategies has been a partner with Poșta Română since 2005 and with Pitney Bowes since 1994.

This paper incorporates best-practice information from Pitney Bowes Business Insight in addition to Romania-specific examples from Geo Strategies.

Contact us at info@geo-strategies.com or call us (+44 1223 205080 or +40 722 940) to discuss the steps for your geocoding requirements.

To find out more about what Geo Strategies could do for your business in Romania go to www.geo-address.com

Geo Strategies