

SMARTaddress
for Romanian address
cleansing, de-duplicating
& geocoding

Geo Strategies



SMARTaddress

Welcome to the next generation of address cleansing and geocoding solutions for Romania



What is it?

Just as organisations have adopted CRM and Data warehousing technologies, they need to put in place a data quality strategy, too. Business in general, and marketing in particular, needs to become increasingly data-driven. If there is no care to govern, in particular, the address data quality, very soon there will be gaps in key activities.

A data quality strategy helps to identify the responsibilities of each function and level of management to ensure that legislation and codes of practice are followed; there are real dangers for organisations if their compliance is poor. On top of this, using inaccurate, out-of-date or incomplete address data – or just not having the right address data at all – can undermine all the effort put in to establishing positive customer and prospect relationships.

Geo Strategies has developed and operates [SMART-address](#) a comprehensive address parsing, cleansing and geocoding system &/or service for Romania.

This is fed by the most comprehensive location knowledge-base compiled and maintained by the company since they started activities in Romania in 1993. This includes locality and place names, over 700,000 business locations, landmarks and Points of Interest for Romania, together with all recognised street names – all with the correct diacritics. More importantly, the tool operates a parser and rules-based algorithms tailored to Romanian phonetics and is supported by a secondary database of variants – abbreviations, local names (unofficial names which are used in the local area) and, indeed, old address names.

SMART-address can process databases of up to 5M records and the success rate varies between 75% and 99% depending on the cleanliness of the input data. Typical success rates for Romanian databases (which are all-too-frequently less than ideal) are in the order of 85%; however, Geo Strategies is happy to work with customers to correct gross errors in their address data.

Address cleansing

Identifying the individual address components (county, locality, street type, street name, numbers and postcode).

Address standardisation

Correcting any mis-spellings, adding diacritics, ensuring conformance to standard abbreviations, taking account of updates, name changes, etc.

Postcode attribution

Assigning the correct 6-digit postcode from Poșta Română.

Geocoding

Assigning coordinates to the address (lat/long WGS-84).

What does it do?

The cost of unclean data takes its toll at multiple levels. There is an industry rule-of-thumb that estimates that it costs one dollar (\$1) to fix a bad address record at time of entry; \$10 to fix it at the back end; and \$100s if that bad address record is allowed to fester uncorrected over time. If we multiply the 1-10-100 dollar ratio by hundreds-of-thousands or even millions of addresses, it becomes clear how fast these costs add up.

In Geo Strategies we have witnessed direct evidence of this problem.

Address standardisation

SMARTaddress is based on a semantic parser which recognises and extracts individual address components (counties, localities, street names, numbers and postcodes - if included). Once extracted into separate fields, the various names are compared with an internal reference database and corrected &/or updated if necessary: this includes replacing abbreviations, updating names where they have been changed and adding diacritics if required.

The resulting address can be output as capitals, proper case (first letter of names capitalised), with locality types (Oraş, Satul, etc.), with the correct street types (Bulevard, Strada, etc.) and with the numbers (Nr., Bl., Sc., Et., Ap.) either in separate fields or concatenated.

Address enhancement

Once an address has been identified and validated against the national database, it can be enhanced by the addition of the correct 6-digit postcode. Due to changes of PCs, it is sometimes necessary to correct the postcode which was supplied in the original database: if a change is made, then all such changes are identified in the output data.

De-duplication

Many organisations are now moving to an SCV (a Single Customer View): to achieve this all duplicates must be identified and removed.

This is easily achieved in SMARTaddress that can recognise, for example, that Str Ateneului ,Nr.3 and Str Episcopiei, Nr.3 in Bucharest are identical (the name was changed in 1991).

Geocoding

In all cases, coordinates are assigned with the highest level of accuracy possible. Due to the lack of official reference information for certain areas of Romania there are four different levels of accuracy:

ADD Address (or individual building number)

For streets with address ranging, coordinates are interpolated from the coordinates of the nearest street intersections.

SSC Street Centre Centroids

For those streets which are included in the mapping and which extend beyond a single postcode, the separate street segments are assigned to the correct postcode and with their own coordinates: for extra accuracy, the centroid is snapped to the street centreline.

SC Street Centroids

Similar to SSC, but this applies to streets wholly contained within a single postcode.

LC Locality Centroids

Locality centroids are supplied when only the locality name is supplied as the address or when the named street does not appear on published mapping: this is particularly prevalent in rural localities.

Clean Addresses = Better Business

Who is it for?

Who has the address pain?

IT departments are quite often overwhelmed by the task of bringing together financial and operational databases from across the organisation. The demands from Customer Insight and Marketing will be down on their list; rather, the IT manager's problem is to reduce the IT burden and promote efficiencies.

Is there value to IT from SMARTaddress?

Definitely! SMARTaddress can be rolled out in various ways:

- Based on a (local) thin client i.e. an Excel Add-in working with a remote server.

There is no requirement for expensive in-house servers and infrastructure as costs are incremental i.e. per record processed. This is recommended for small batches e.g. for daily address maintenance and geocoding.

- SMARTaddress also supports flexible working as SaaS (Software as a Service) via the Cloud: it is particularly suitable for larger databases (between 5,000 and 500,000 records).
- An API is available for integrators who may wish to establish a direct connection from within their CRM or data warehouse.

Geo Strategies also offers an address cleansing, standardisation and geocoding service which is particularly suitable for organisations wishing to audit and overhaul their contacts' database. This service includes full QA and reporting of errors plus recommendations for improving data management procedures.

Similar benefits apply to the Integrators who very often face end-client dissatisfaction on new technology implementations due to the quality of the data rather than the technology brought in by the integrator.

The CRM Manager's challenge is to evolve the CRM from 'data storage' to proper, effective customer management. The CRM manager needs to add value to corporate decision-making and this can be enabled through geocoded, clean address data.

This is particularly true as senior decision-makers (the Board and, especially, Marketing Managers) begin to realise the benefits of knowing precisely where their customers are located and how far they have to travel to obtain goods or services.

McKinsey, Gartner and others all agree that over 85% of all data contains a location element and SMARTaddress can provide that information. Location is king!

What is SMARTaddress' value to CRM?

Clean and geocoded data enables internal processes such as corporate mapping through to Call Centres via location search. It helps present a single geographic representation of the business 'as is' and also market opportunity data. It can integrate the CRM with the Internet to promote self-service. The analytical benefit is seamless: for example, a critical retail optimisation application is based on calculating distances from any organisation office to a POS, competitor or to a customer location.

The Marketing Manager is quite frequently inhibited by technology. However he/she needs easy access to a single view of customers (SCV) to generate insight, develop and match propositions and for communication.

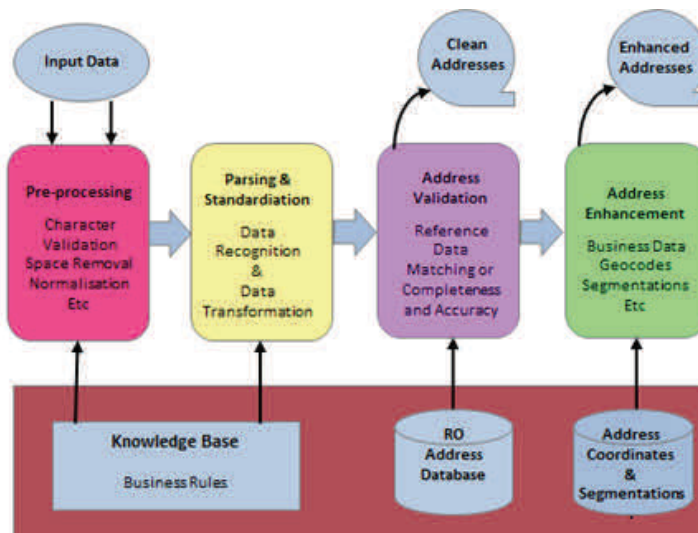
Knowing where customers live (being able to see them as dots on a map) provides huge insight as *"birds of a feather flock together"* i.e. similar people tend to cluster in e.g. Pantelimon, Baneasa or Snagov. SMARTaddress underpins this ability.

What does SMARTaddress do for the Marketing Manager?

It makes life easier by operationalising mountains of data which the IT doesn't own. Once correctly addressed and geocoded, customer data can be segmented, mapped and enhanced with additional information linked by address. The power of the address (location) is that it is relative to all other data from internal and external sources. Overall, it strengthens Marketing's resilience for critical analysis and decision-making; and it helps Marketing challenge the IT for bespoke slicing and dicing of bigger data.

The Data Analyst needs location information to provide the information increasingly required by Boards. GIS has become relatively commonplace in larger organisations but all the maps in the world are useless without real business information. Again, SMARTaddress can provide the geocodes to unlock the wealth of information and, potentially, business insight that resides in an address database.

How does it work?



Romanian addresses

Romanian addresses are morphologically rich - the correct term for what, in English, is normally termed 'flowery'. This, in turn, means that they are relatively complex e.g. Subcomis. de Pol. Ion C. Pompilian.

This complexity would not be a problem for automatic computer recognition and processing if such addresses were transcribed into databases correctly but the fact is that most databases contain many (almost random) abbreviations. Furthermore, there is a prevalent legacy of poor database design and implementation, compounded by poor data entry disciplines.

The overall result is many address databases with random spellings, concatenated data (or mixed data in pre-defined fields), missing address components (e.g. no county) and a varied approach to numbers (Nr., Bl., Sc., Et., Ap.).

SMARTaddress

SMARTaddress has been built on a comprehensive database of Romanian locality and street information, together with a rules-based engine which defines the structure and content of all addresses for Romania. This is all supplemented by a 6-digit postcode database and coordinates for individual addresses, road or street segments, individual streets and all localities in Romania.

Pre-processing

All address data is passed through a pre-processor to remove artefacts e.g. blank spaces, rogue characters (often arising from file or format conversion), inappro-

prate content (e.g. telephone numbers incorporated into the address), etc.

Parsing

Once the individual address records have been tidied-up, then they are submitted to the Parser. This system 'recognises' the individual address components (county, street name, locality, street type, etc.) and places them in the correct field. This is a very complex process based on pattern recognition and natural language processing similar to that required for automatic language translation.

Validation, standardisation & enhancement

The separate address components are then checked against a comprehensive Reference Database of localities and street names. This database contains not only the correct address components but also recognised variants such as Bucharest for Bucureşti, DTS, Rm. Valcea, etc. Where necessary, the address components are changed to match national standards e.g. Tâncăbeşti instead of Tin-cabesti.

Postcoding & Geocoding

If required, 6-digit postcodes &/or geocodes are then identified and attached to the addresses.

Reporting

A report is generated and appended to every record to identify any errors which were detected and changes which have been made.

How can I use it?

Address services are available in one of four ways:

Via an Excel Add-in

A Tab (or Ribbon) in Excel provides direct access to a remoter server. This is ideal for small batches e.g. up to 5,000 records.

Via the Cloud

The remote server is accessed directly over the Internet. This is ideal for large databases.

Via an API

SMARTaddress can be accessed over the Internet via a SOAP compliant API e.g. for direct interface to a CRM.

As a managed service

Address services (including de-duplication) are available from Geo Strategies for those that require a turnkey result. Customers are provided with a report to help them improve their internal processes for the future.

Cleansed and standardised addresses can be provided in a variety of formats, with our without diacritics.

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 and marketing services, consultancy, training, and project management.

This expertise has been developed according to internationally recognised best practices and methodologies, to which significant innovation and 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

'Address data issues impact company performance in many ways : the costs of bad or useless address records are high due to leading to duplicate communications, poor segmentation, missing opportunities based on mis-calculating the unique customer value, poor service, poor decisions. SMARTaddress is a ready-to-use, powerful tool to turn bad address data into good and valuable information for every organisation which stores and manages Romanian addresses.'

Daniela Florea, CEO of Geo Strategies Ltd

Contact us at www.geo-address.com or call us (+44 1223 205080, +40 269 210832 or +40 722 940) for more information.

To find out what Geo Strategies could do for your business in Romania, go to - www.mosaic.geo-strategies.com

SMARTaddress

Geo Strategies Ltd
St John's Innovation Centre
Cowley Road
Cambridge CB4 0WS
United Kingdom

T. +44 (0)1223 205080

M. +40 (0)722 244940

W. www.geo-address.com

W. www.mosaic.geo-strategies.com

Geo Strategies SRL
Bvd. G-ral V. Milea 10A
Sibiu
550325
Romania

T. +40 (0)269 210832

M. +40 (0)722 396034

W. www.geo-address.com