3tc Software
Command. Control. Inform.

Premise Gazetteers and Location Matching for the Modern Control Room

WHITE PAPER - GEODE

Author: Steve Taylor. BA, MA, MSc

GEODE

Table of Contents

Copyright

3tc Software Ltd
The Coach House
Desford Hall,
Leicester Lane
Desford, Leicestershire
LE9 9JJ

www.3tcsoftware.com
Call 0845 688 9950

Disclaimer

Introduction

One of the key activities in any emergency service Control Room is identifying locations relevant to the call. For many years technology has been used to assist in this process with the principal tool being, computerised searching of a gazetteer and more recently supplemented by the use of GIS and EISEC/ALSEC data. This has become well established technology but it is now facing new challenges and new opportunities.

This white paper is a discussion of those challenges and opportunities and an examination of where, in an ideal world we as an emergency service community may like to go next.

There are a range of challenges and opportunities arising from changes over the last decade

- The move from street based gazetteers to premise based gazetteers leads to significant data expansion
- Increasing co-operation between control rooms is leading to a need to hold gazetteers for larger areas
- Search techniques have become stalled
- Improved computer technology should enable the use of more advanced ideas
- Users have come to expect a better solution as a result of experience of other computerised search tools.

Data Volumes

Over the last few decades emergency service Control Room solutions have moved from parish gazetteers, through to street gazetteers, to the modern ideal of a full premise gazetteer. There are now a variety of data sources for this information that are being brought together under the PSMA (Public Sector Mapping Agreement) to offer a good solid basis for full premise mobilising as the norm rather than the exception. However, the move to premise gazetteers has led to a major scale up in data volumes which have consequences for the efficiency of location matching processes.

In addition with the need for greater co-operation between controls there is a need to hold gazetteer information for other service areas so that overflow calls and fallback services can be provided. This leads to a further increase in data volumes illustrated by some example gazetteer sizes.

Typical Data Volumes	
County Street Gazetteer	30,000 entries
County Premise Gazetteer	1,000,000 entries
Regional Gazetteer	3,000,000 entries
Multi-Regional Gazetteer	10,000,000 entries
UK-wide Gazetteer	30,000,000 entries

These increases in data volumes have a number of consequences.

Increased data means that traditional computerised search techniques are slowed down and given the scale of the data expansion this remains a challenge despite major improvements in computer technology.

Interacting with the caller now requires Control Room staff to confirm more information. They now need to try and pin down the full premise address potentially extending the time taken to process a call. Whilst this issue could have a technical solution it is about process and therefore requires a more sophisticated approach to make sure that Control Room solutions assist in expediting this process.

Last, but not least, there is now a much more significant data maintenance task. Existing street gazetteers can take significant resources to keep up to date and to customise to deal with local difficulties. It has become the norm that services modify the raw data they receive to make sure that the Control Room processes achieve a location match during call taking. If the scale up in data volumes leads to tenfold increase in gazetteer management effort this represents a cost that few, if any, emergency services can afford.

Implications of Co-operation

Co-operative working between different Control Rooms raises the further issue of consistency. This has traditionally only been of concern to those actually creating gazetteers and this means that many emergency services have not directly experienced the issue. However, it has been clear that it has been a major issue for NLPG data but it doesn't become obvious until you merge data from multiple custodians. There are a number of data consistency issues but all have their impact.

If generic names are entered differently in different areas it can have implications for searching and matching. An easily accessible example is the variations of "Sainsburys" and "J Sainsburys". All

custodians need to use the same form or there are potential issues for Control Rooms trying to find addresses and "Sainsburys" is a particularly good example as a couple of years ago the different versions existed in NLPG data but it is in the process of being corrected to a consistent standard. However, whilst data consistency is an issue to be corrected by the gazetteer custodians it is the emergency Control Room staff who remain stuck with the consequences.

Gazetteers have built into them the ideas of different address fields such as property name, street name and locality name; typical location matching products expect you to know what type of data you are entering and put it in the right field. This is usually quite simple but sometimes it can be very difficult to reliably determine the correct field to use. When is a block of flats a street? When is a street a locality?

A more subtle consistency issue is the meaning of data not being in the Gazetteer. It is important that if one custodian puts data into the gazetteer then all do. We can reasonably assume that the data available under the PSMA obeys this rule, but if you are going to share data with another Control Room or partner organisation, it is important that the data you share has been handled the same way. If you import telephone kiosks into your Control Room gazetteer then you need kiosks to be in all the gazetteers that you use. This immediately raises an issue if the individual Services are modifying their Gazetteers or importing different datasets.

Co-operative working also has a direct impact on the location matching process since if we have a twelve county gazetteer any search method is likely to find twelve times as many possible matches. This is because the single county search is using an implied 'in this county' search criteria in every search and when we lose that, the search process is inevitably affected.

Location Matching

Location matching techniques have a long history starting with systems deployed in the late 1970's with initial search techniques simply being a full exact match on the data entered. Location matching has developed over the decades with a variety of improved techniques available. The timeline shown below is, of necessity, based on our experience in working on location matching and products in service will typically lag behind this progression.

1980s

Exact Match

A successful exact match would require the user to enter details exactly as found in the Gazetteer. This is relatively easy to implement with the computers from this era but required the caller to give the correct address information. Once gazetteers had multiple fields (no longer parish mobilising) the algorithms extended to only requiring an exact match

on the fields entered.

Partial Match

For partial match the computer would only check the first few characters of what was entered giving a list of possible matches that an operator can choose from. Any inaccuracy in the first few characters would prevent a match and sometimes there could be a very long list of potential matches to be checked.

Phonetic Match

In this period phonetic matching was implemented using the Soundex algorithm. This represented a significant improvement but the algorithm has a range of known weaknesses and often failed to find similar sounding addresses whilst presenting others that were radically difference.

Geographic Relationships

Even in these early days we were using the geographic relationships between different parts of the address. The search algorithm knew the adjacent locality to the locality entered and would consider addresses with those localities as well as the one entered.

STD Code Areas

The exchange codes from the caller's number could be used to add locality information based on a list of localities that the exchange covers. This could be quite effective when calls were universally on landlines and STD codes were locked to localities by the telephone technology of the day. However, this approach is losing its effectiveness as exchange codes lose their geographic relationship. Plus, many calls now come from mobile phones.

Modifying the Gazetteer

A key solution to the inability to find things was to add alternative names based on common errors to the Gazetteer. This would neatly sidestep some of the issues with location matching but required detailed knowledge of problem addresses and significant data maintenance effort.

The range of techniques developed in the 1980s offered a fairly sophisticated search environment that very often found the relevant address but using the full range of techniques was often slow and required the operator to request the different types of search as they attempted to find the address. Typically an operator was left browsing through a list of addresses trying to find the correct one in a potentially long list. This could easily lead to questioning the caller about spelling being the most efficient approach to finding the address.

1990s

Scoring and Sorting

This was a key innovation massively improving the ability of Control staff to find the correct match more quickly. By sorting the possible matches based on how well they matched the search data, the correct match, if present, was often at the top of the list.

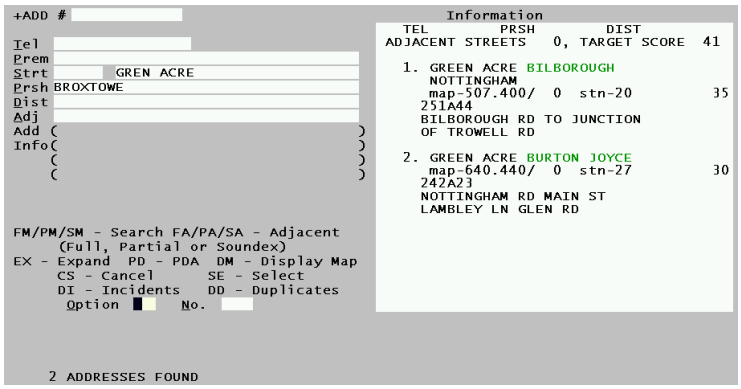
Automatic Searching

Automatic searching is letting the computer automatically carry out the alternative search techniques without operator intervention until the search found some good matches to show to the operator. The decision making is linked to the scoring process mentioned above so that the automated process can terminate at an appropriate point.

Improved Phonetic Algorithms

In the late 1990s we did some work on improved phonetic algorithms – Phoneticus and Phoneticus 2. These are roughly equivalent to the more modern Metaphone and double Metaphone algorithms and whilst they did offer some benefit, it was limited and the project was abandoned before completion as we had begun to investigate an alternative approach.

Metaphone was published in this period and has come into use as a substitute for Soundex.



Typical 1990s style address search
GREN phonetically matched to GREEN and BROXTOWE adjacent to BILBOROUGH

2000s

Improved Phonetic Algorithms

The Metaphone algorithm has seen several improvements over the years and the use of

double Metaphone and now Metaphone 3 takes phonetic algorithms about as far as they can go. This retains some weaknesses as it is based on United States pronunciation and will fail significantly when dealing with Welsh or Gaelic place names. There are also issues relating to large numbers of UK place names having unexpected pronunciation.

Word Indexing

This is a key opportunity for improvement. Most database products now support word indexing of text which allows an address search to look for addresses containing particular words rather than beginning with specific text. This could give a significant improvement in searching but still has significant weaknesses around correct spelling.

Free Text Search

Database performance is now adequate to support a complete free text search of a gazetteer within reasonable timescales. This means that a search can simply look for addresses containing specified text without making assumptions about position or being a complete word. This remains a significant performance challenge especially for larger Gazetteers.

EISEC / ALSEC

The introduction of EISEC now makes a massive difference to large numbers of calls to the emergency services by providing a direct path to the caller's address. This bypasses many of the location matching issues but it is only properly useful when the caller is reporting an incident at their address and they are calling from a landline.

EISEC data has a secondary role and that is to replace the implied localisation of 'in our county' that used to come from taking an emergency call along with replacing the STD code localisation. EISEC data for both landlines and mobile phones can be used to replace that input.

What Makes a Good Search Experience?

In all the discussion of the technology it is easy to lose sight of the fact that location matching is carried out by Control Room staff and the process we are trying to optimise is their process of establishing the incident address. If we keep this in mind then there is a very serious question of exactly what makes a good search experience?

The answer has to be a search experience that;

- Works quickly, speed is a feature
- Shortens how long the process takes
- Makes it less stressful
- Reduces the risk of incorrect matching

To pursue this issue we really need an example of a good search experience and probably the best example is to look at how Google web searches work. The current Google search offers a number of hints at what makes a good search.

- It is fast, internet users are notoriously intolerant of delays
- It searches whilst you type without waiting for you to ask it to search
- It tries to allow for spelling mistakes
- It offers you possible full matches to your search criteria as you type

All these techniques can be seen to be an attempt to cut short the search process and yet only one of them is about the actual performance of the service. This gives a strong clue that improving the search experience has a lot more options than just doing things quickly.

There are a number of conclusions we can draw from looking at Google and published research around user's reactions to how well it works.

Performance is a Feature

It is quite clear that how quickly something happens changes the user perception of things. Using some work we have done on gauging Control Room reactions to performance we come up with the following figures;

Response Time	Perception
Less than 0.5 seconds	Perceived as an instant response
0.5 – 2 seconds	Seen as fast
More than 2 seconds	I'm waiting
More than 5 seconds	I'm starting to wonder if its broken

These figures are actually quite interesting as they are actually rather shorter than typical computer users and represent the intolerance for slowness in an emergency service situation.

Users like to get the answer before they ask / before they finish typing

This is a key improvement Google made when they introduced Google Instant (for more details search for Google Instant). This is a significant improvement for Google users and ought to be even more effective for location matching. Google claim 2-5 seconds saving on total search time and we would expect rather more than this. This is more time than can be saved by speeding up a 2 second search.

Make sure it's useful

This sounds like a slightly odd thing to say but one of the most important things that Google does day in and day out is to work on making sure that their results are what you are looking for. If you use Google a lot then you notice when somebody starts gaming the system and you begin to get results you don't want, useless web pages that are just trying to con you to go and look at them. Google works ceaselessly to make sure the results you see are the ones you want and this remains the key to a good search experience.

3tc Software's Solution

At the end of the last millennium we were in the position where we had decided that location matching needed a fundamental rethink as the old techniques and incremental improvements on them were not going to solve the problem. At the 2000 Fire Exhibition we presented a demonstrator of some new ideas to solicit opinion. At this point the focus was on providing a better search experience with an 'as-you-type' search process.

"This is the only thing I've seen today that I want"

Over the next 5 years we developed that demonstration into a new product targeted at solving the key problems of location matching. Although execution was quite complex the actual objectives of the project were quite simple;

- Eliminate the need for human intervention in processing source gazetteer data into something suitable for location matching
- Create a search process that finds correct matches for all those difficult addresses that otherwise require expert local knowledge or detailed caller interrogation
- Examine the ergonomics of the location matching processes and streamline to maximise call taker performance – maximising speed, minimising errors

We believe we have succeeded

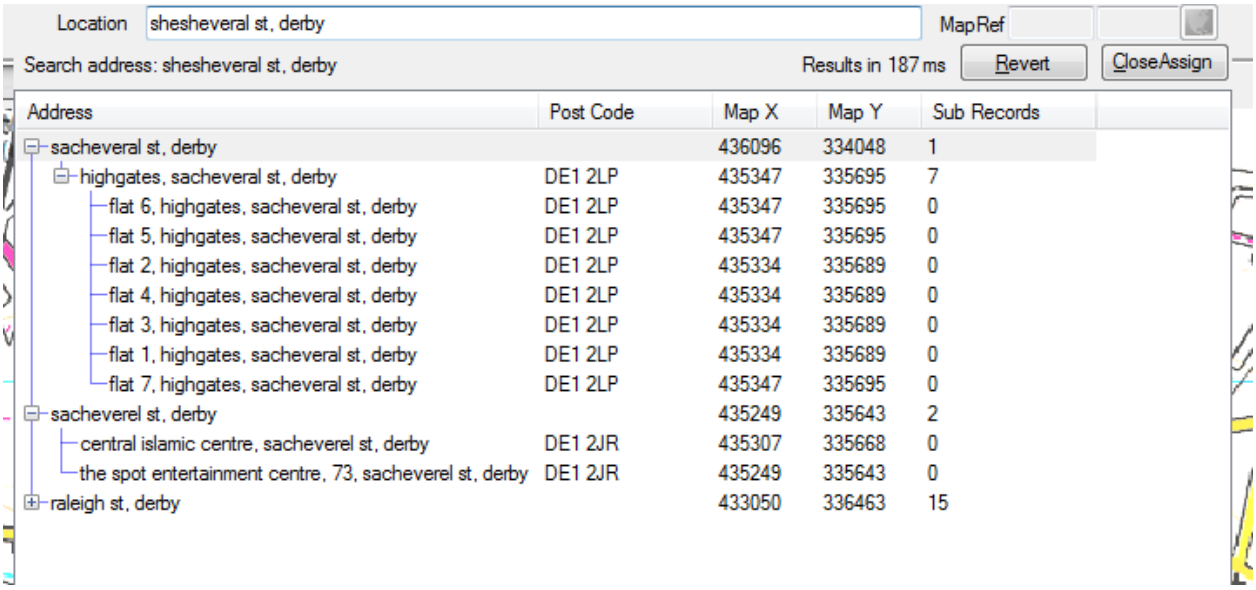
That revolutionary new product is called **Geode** and not only does it address all of the challenges of moving to premise gazetteers and collaborative data scale up but it also significantly improves the underlying location matching process to yield gains in speed, efficiency, risk reduction and general user experience. It is based on a completely new approach to location matching that overcomes all the weaknesses of previous search techniques.

- It no longer relies on the operator getting particular bits of the address correct
- It no longer relies on the operator telling it what to try
- It no longer relies on the operator telling it when to search
- It no longer cares which parts of the address you get wrong
- It no longer needs address elements to go in the right fields
- It no longer makes the operator wait
- It no longer requires anyone to prepare the gazetteer
- It no longer requires anyone to modify the gazetteer to help find things

"It simply searches as you type and finds the right match"

Geode searches as you type providing a continuously updating list of possible matches allowing the operator to stop when they have a match rather than having to continue and type everything in. The search algorithms depend only on finding some content that matches and it doesn't matter where errors are. This creates a highly efficient and successful search process.

"One of the biggest challenges of adopting Geode was teaching staff not to go back and correct typos"



Example Geode Search

The most important difference between **Geode** as a search engine and other products is that **Geode** will find an address based on how much of what has been entered is correct. It doesn't care which part of the address is correct and it doesn't really matter how correct. As an example:

Typical NLPG data from multiple custodians will contain a number of Sainsbury's supermarkets. The style of entry varies with some being entered as "J Sainsbury's" and some as "Sainsbury's". Most search products rely on the first letter being sort of correct and the Sainsbury's problem will leave them stymied. **Geode** has no such issue as all it cares about is that the "Sainsbury's" bit is in both and they are obviously a match.

Now, whilst the above example is a generic issue that could be addressed by the data custodians, or special provision could be made for it, it is just an example of a general problem perhaps better illustrated by considering whether a call taker would enter the silent "t" at the start of "Tsimbi Institute" or many other addresses with similar issues.

Another good example, illustrated above, is a search for 'Shesheveral St' (actual spelling 'Sacheveral St').The example illustrated above, is a search for 'Shesheveral St' (actual spelling 'Sacheveral St'). This is taken from an actual call some years ago where the operator had difficulty making a match. **Geode** simply offers the correct address as the first option as you type leading to rapid termination of the location matching process. Other systems may find this match as it is subject to phonetic matching but you need to ask for a phonetic search and even then it may not be the top offering. Overall the effect of **Geode** is to speed up and remove the stress and risk from the call taking process

Geode has a secondary benefit, in the Sacheveral example we can see in the results a match on Sacheverel. This is a simple spelling error but it is a mistake made by the gazetteer compiler not by the user and **Geode** reliably offers you not only the correct address but also the next most similar addresses providing the comfort that there isn't another similar address that you might have considered. For **Geode** there is little point in correcting the gazetteer as it simply does not make any difference, **Geode** finds it anyway.

Creating a Gazetteer

Geode has features built in that eliminate the need for any work to be done on the gazetteer data other than running an automatic import from the source data. On older systems all the search techniques based on geographic relationship required manual addition of data to the gazetteer. The **Geode** import process carries out an automatic geographic analysis of the data and works it all out for itself.

Geode has been used operationally to take over 1,000,000 emergency calls with automatically generated gazetteer data with no user modification.

How does the location matching process work?

Location matching is now quite simple. You type in information and **Geode** constantly searches on the data so far and shows an updating list of possible matches. At any time you can go down to the list and start exploring the address tree – check the premises on a street, look at building sub-divisions – anything you like. Once you are happy choose an address.

Searching offers a list in order of how well they match and will show all addresses similar to the search even if it has an exact match. This gives the operator confidence about what alternatives there might be.

Benefits in the search process are;

- Search as you type supports cutting short data entry
- Sorted on match quality gives results in a natural viewing order
- Listing of close mismatches gives confidence on match selection
- Ability to explore the gazetteer around a match helps discuss details with the caller
- Search speed means never waiting for the system

What factors does a Geode search consider?

The actual search techniques are proprietary and rather complex but we can list of all the information that is considered by a search.

- How well does the text entered match the gazetteer
- If part of what is entered mismatches is it in fact somewhere nearby
- Premise numbering (multiple levels)
- Various types of positional data can be input, mobile phone cell ellipse, caller provided coordinates, near to caller's EISEC address, approximate grid reference
- Post code analysis

So how fast is a single Search?

The speed challenge is a bit of a non-issue with **Geode** offering a full search - all options - in less than half a second for regional scale gazetteers. This performance is achieved despite carrying out all of its search options on every search to offer the best possible speed. Other solutions often require the user to request special searches such as phonetic matching; **Geode** always uses all options and is fast enough that this does not cause problems.

Scale	Typical search times
National Gazetteer (England, Scotland and Wales)	Full search every 0.7 seconds
10,000,000 Premises	Full search every 0.25 seconds
Regional Gazetteer on a cheap computer	Full search every 0.3 seconds
County Gazetteer	Full search every 0.1 seconds

Implementation

Geode is available integrated into 3tc Software products or as a component for third party integration.

Available Products;

- Integrated into other 3tc Software products
- Stand-alone address searcher
- ActiveX for local embedding
- Web Service for server based searching

- Socket Interface for customised server based searching

ActiveX and server versions come with a 3rd API for integration into other products

Benefits

Enhanced Control Room performance is significantly enhanced by **Geode** because it cuts short the number of interactions between Control staff and the caller and between Control staff and the system. In a standard deployment **Geode** executes a full search on every key-press offering a possible match list as the operator is typing allowing instant feedback on the information found so far. All available search techniques are used on every search and **Geode** significantly shortcuts the required interaction with the caller creating a much more dynamic and interactive process than traditional location matching products leading to a major reductions in call taking time.

Geode offers significant gains around gazetteer management as it is able to function effectively from raw national data and requires no data cleansing. Rather than the customer having to modify the data for the search engine **Geode** simply works on the data 'as is' and finds the matches anyway offering a substantial cost saving and risk reduction. **It is hard to exaggerate the significance of this as it leads to the potential for massive cost savings and major risk reductions in call taking**.

Reduced Risk

The risk of course, is that an address isn't updated, correctly formatted or found accurately – and resources may be sent to the wrong location. The potential for damage to property or loss of life is real and preventable. The risk to the organisation concerned is unnecessary when there is a simple and intelligent solution such as **Geode**.

Location matching failures have a secondary risk which results from the stress experienced by the call takers. Being unable to find an address in a potentially life threatening situation can be very stressful and this represents a health and safety issue for the emergency service involved.

The Economic Argument

It is likely that a typical premise gazetteer user will have to expend many person months of effort up front to get their gazetteer in order for a traditional search product and then additional on-going effort to maintain their gazetteer in working order over the following years. With current national projects it is likely that premise gazetteer updates will, in a very few years' time, be issued on a daily basis and will require prompt update of systems to reduce the risk of not having an address in your system that has been made available.

All this is very costly. Geode eliminates that cost

Geode could easily reduce the effort required for gazetteer management by well over the labour equivalent of £10,000 - £50,000 per year and with the typical Control Room system being retained for 10 years this implies that the long-term saving from having something like **Geode** is at least £100,000 - £500,000. Cost savings can easily be much higher than this when one factors in how skilled staff can be better utilised in front-line duties.

3rd party integration: **Geode** is available as an ActiveX (OCX file) to embed locally within an application or as a web service (or socket interface) running on a server.

Summary

The introduction of premise gazetteers along with increased cooperation between geographically different services is massively increasing gazetteer data volumes. This escalates a number of pre-existing problems around location matching in emergency service control rooms.

Location Matching Performance

The technology is often slow and its weaknesses puts increased load on call takers resulting in the need for increased questioning of callers and delays in resource dispatch.

Data management costs

Creating and managing a gazetteer is already an expensive proposition and the changes taking place are escalating those costs to a point where they may be a significant drain on emergency service resources.

Data Consistency Issues

Sharing data is going to make data consistency issues worse and these can lead to serious risk issues for location matching processes.

Traditional location matching techniques don't solve these problems

3tc software has developed a revolutionary new product called **Geode** that solves these problems.

- Significantly reduces location matching time
- Significantly reduces location matching risk of error
- Significantly reduces call taker stress
- Eliminates data management costs

This can lead to significant cost savings for an emergency service Control Room by eliminating data management costs and significantly increasing Control Room capacity. Please contact 3tc Software for more information or a **Geode** demonstration.