

W3.4

The 'simple' NLPG creation methodology

John A Couzens, GIS Manager & NLPG Custodian, Reigate & Banstead Borough Council

1 Introduction

What is the fundamental point of the NLPG? Probably discussed at length in many local authorities and elsewhere, and probably never truly answered. If it is to be a way of uniquely identifying all land and property, then certain rules and conventions must apply. Thankfully, BS7666 has arrived in a timely manner. Discussed at a greater or lesser extent within most local authorities, many scenarios have been acted out and some solutions found. This is an outline of a simple solution to produce the NLPG compliant LLPG ready for the "1st cut"

2 Background

At Reigate and Banstead Borough Council, the generation of a Local Land & Property Gazetteer (LLPG) had been a key deliverable for their Corporate GIS project.. With much spatial data already collected, only the LLPG was an outstanding and needed to be delivered. This coincided with the various e-Government initiatives for implementation of service delivery. The various "N" projects, (NLIS, NLPG, NSG, NLUD etc.) added credibility to the need for a corporate involvement to the creation of the LLPG but at the same time, strategic planning within local government tends to produce results later rather than sooner. The GIS viewpoint was that there was nothing fundamentally in the way of delivering an early version of our LLPG. With all of this in mind, last October (2000) a target of getting a LLPG for Reigate & Banstead into the "1st cut" of the NLPG was set.

3 Fundamentals

In order to understand the scope of the project, a number of key pieces of information were needed. A copy of BS7666 (version 2) was obtained, and resources allocated for to the creation of the LLPG were identified. Strategic planning helps to provide a budget and staff resources in due course but does not allow for a rapid development of a project as in this case. As no budget had been made available, the proposed work had to be accommodated within the existing constraints of resources. Staff resources were therefore identified as being the GIS manager who would undertake the work alongside his normal duties.

In order to build a LLPG, a base-line data set was required. Like many local authorities, investigation revealed a multitude of gazetteers and address based data sets. These included the Electoral register, Council Tax, Non-Domestic Rates, Housing Revenue account, Housing benefits, Building Control, Planning, and of course Address Point from the Ordnance Survey. Many of them would be identified by their respective "owners" as being complete or up to date. Close scrutiny of samples proved otherwise.. Additional factors to be considered included the fact that most of these database systems in use were due to be replaced or upgraded over the next 2 to 3 years.

The choice of the base-line data set eventually fell on the Building and Development Services database which had very recently been enhanced and updated to run within the Uniform 2000 software from CAPS Solutions. Extensive work had been undertaken by the suppliers as well as in-house to ensure that as much of the legacy data was incorporated into a new Address point based data set.. The legacy data included some records going back to 1947.

4 Building the LLPG

4.1 Preliminary data gathering

Having identified a suitable dataset to become the base-line, the next stage was to understand the structure of the records and how close or far it was from the BS7666 standard. A number of sequences of checking were identified (Table 1) which gave confidence that the data was inherently sound and also useable for the purpose.

It was quickly apparent that given the short time frame since the implementation of this new system (Uniform 2000), the data was still in good order and whilst several thousand new records had been added, they were well structured.

One benefit that was available from this dataset was that the Eastings and Northings were available as most of the records had been matched the postal address within AddressPoint on conversion, and the opportunity was taken at the time to incorporate the OSAPR record into a spare filed within the Uniform 2000 database. At the time that this was specified, it was not known whether it would be of use but it could easily be deleted if not required at a later date. (this piece of forward planning saved many hours of work)

A download from the Uniform system was taken to include all the address elements (including the OSAPR). This data was then imported into a number of standard desktop systems for further work to build the LLPG. It was fortunate that the initial record counts indicated that there would be no more than about 65000 property records within the area would be identified for the first version of the LLPG. This meant that MS Excel could handle the entire dataset without being truncated by the maximum record numbers or have to be split over multiple sheets.

The first step was to load the exported records from Uniform 2000 into MS Excel. A simple comma separated values (*.csv) file was used for this. Once in MS Excel, the steps as indicated in Table 1 were completed. It was also important to identify any leading spaces in the various fields and remove them, along with punctuation to obtain consistency. Spelling was also corrected where possible. At this point I should be noted that as we are correcting an existing dataset, we ensure that we are able to pass these corrections back. The unique reference number from Uniform 2000 was retained and can be utilised as the common key field to post the data back into Uniform 2000 at a later date.

The next stage was to identify the unique roads within the data set. Again, using a filter within MS Excel, it is possible to extract the roads as a unique list. This was to be used to match to the Unique Street Reference Number (USRN) from the National Street Gazetteer (NSG) maintained by Surrey County Council (SCC) as the local highway authority. A copy of the NSG for the area was obtained and investigated. A number of possible problems were identified but as it conforms to BS 7666 those problems were not deemed to be significant (at the time). Using a match within MS Access, the USRN was applied from the NSG to the LLPG records for the streets. On investigation, it appeared that there were some 100+ streets identified within the Borough with properties that did not match any record within the NSG. A few were identified with different spellings, but there was still a significant number of roads missing from the NSG. The additional start and end points for the streets were also appended.

Table 1

| | Task | Method |
|---|---|--|
| 1 | Check for data consistency | Import into MS Excel and check that filed contents are consistent throughout entire dataset. |
| 2 | Check for spelling problems | Use MS Excel to 'sort' and check for consistent spelling |
| 3 | Check for abbreviations and punctuation | Use MS Excel to 'find' and 'replace' |
| 4 | Check condition of road/street names | Use MS Excel to filter for 'Unique' records |

4.2 Getting started

Having obtained a copy of the Data Transfer Format (V5.3) for the NLPG, it became clear that much of the LLPG build could take place in MS Excel on a single worksheet using multiple columns. Having sorted the data by Town Name and then by road within the towns, the block of Unique Property Reference Numbers (UPRN) were assigned.

The latest AddressPoint data was also 'trimmed' to the outline of the Borough and matched to the existing OSAPR's within the putative LLPG to check for additions and changes. Eastings and Northings from this latest version of AddressPoint were added. The USRN and coordinates from the match to the NSG was also added..

The MS Excel worksheet was set out to incorporate all the major fields to be found in Tables 11, 21, 22 and 24 of the Data Transfer Format. Those fields that were marked as 'optional' were then identified and completed as far as possible.

Careful use of the 'sort' function was then undertaken to identify any 'mandatory' fields that were still blank. Some required a value such as a 'change type' or 'match status' indicator, but others required a more definitive value and these mostly revolved about the currency of the NSG information.

4.3 Those 'missing' streets

The NSG for the Borough was created in June 1998 and despite regular submissions by the various boroughs within the county, updates did not appear. This probably caused the high level of 'missing streets'. In order to partially rectify this, the GIS (ArcView) was used to look at the Ordnance Survey OSCAR data and extract the street names from it. This also included the coordinates for the start and end point of each elementary street unit. This was then matched to the list of 'missing' streets from the LLPG and, whilst no USRN was available, street start and end coordinates were obtained. And added back into the LLPG

4.4 Submission

The final act of preparing the data ready for submission to the NLPG was to extract from the MS Excel worksheet the data in a form similar to the various tables in the Data Transfer Format document.. At this stage, careful attention was placed on getting the correct data into the correct fields and in the correct format. All data was re-formatted to ensure that it was in a "Text" format and the two Land and Property Identifier (LPI) fields for Primary and Secondary Addressable Object Names (PAON & SAON) were also very carefully investigated to ensure that the full field structure was adhered to. This is essential to ensure that the correct elements of the descriptions are placed at the correct starting position within the field. This is the most difficult part to understand at first but can soon be created within MS Excel and the use of the "merge" function.

After being satisfied that all the elements for each table of the Data transfer Format was completed, each table was exported to a text files (*.csv) and then examined within a simple text editor (Wordpad). This was then compared with the examples found in the Data Transfer Format documentation. Finally, all the tables had a "Header" and "Trailer" record added in this final text stage before being compiled into a single compressed file to be e-mailed of to the NLPG custodian.

5 Identified problems

Basic Land and Property Units (BLPU) existed in the GIS and could be appended to any AddressPoint records within them using a 'point in polygon' search. This provided added validity to the AddressPoint and also rejected those AddressPoints which fell outside the building footprints. This helps to increase the confidence of the AddressPoint data

The currency of the NSG is under review with Surrey County Council. It is proposed that the Districts and Boroughs (as the street naming and numbering authority) issue the USRN immediately a street is named. This gives the added benefit of being able to update and append a validated entry into the LLPG for the street and the properties of a new development without having to wait for a 6 – 12 month update cycle of the NSG. The updated USRN's held within the NLPG can then be accessed by the NSG custodians asking for "Change Type" and "Entry Date" to pick up these new records on a regular basis.

The creation of the entries for the Primary and Secondary Addressable Objects (PAON & SAON) was not at first fully understood, but eventually the methodology was devised and can now be relied upon.

6 Conclusion

The entire process and understanding of the processes of “data cleaning” and the building of a LLPG was facilitated by the reading (and understanding) of the BS7666 and the NLPG Data Transfer Format documents. When read in conjunction with a sample of the data held within a simple format (MS Excel), the relationships and formats of the various tables became clear.

Reigate and Banstead Borough Council now have a Local Land & Property Gazetteer which is part of the NLPG. The benefits of this can now be realised in that over the next 2 to 3 years, the replacement systems for Council Tax, Non-domestic Rates, Housing Benefits and Electoral Registration can all be seeded or cleaned using this initial version of the LLPG. All these systems have been specified to incorporate the ability to link, join, import or export addresses in a BS7666 format. The management and workflow of how this will happen has yet to be defined, but the existence of the LLPG has now become the “driver” in these systems for compatibility.

The whole process from start to finish took 4 months working part time by one person. The confidence level on this data is very high and the estimates for the completeness of this validated data is in excess of 80%. The 20% remainder and the matching of data from other systems is now programmed over the next 3 years as the various systems are replaced.

John A Couzens
GIS Manager & NLPG Custodian
Reigate & Banstead Borough Council
Town Hall
Reigate
Surrey
RH2 0SH

T: 01737 276000
E: john.couzens@reigate-banstead.gov.uk