



Creation of a Local Land and Property Gazetteer

Cotswold District Council

December 2003



Contents

List of files supplied	Page 3
1.0 Initial Processing	Page 4
1.01 Data Supply	
1.02 Data Cleaning	
1.03 Street Gazetteer	
1.04 Unmatched Streets	
1.05 Unidentifiable Streets	
1.06 Deleted Records	
2.0 Merging of Datasets	Page 6
2.01 Conversion to BS7666	
2.02 Creation of LLPG	
2.03 Matching and Merging	
2.04 Merging to NLPG	
2.05 Inserted BLPUs	
3.0 Grid Co-ordinates	Page 10
4.0 Data Returned	Page 11
4.01 Anomalies Files	
4.02 The LLPG	
4.03 Potential Duplicates	
5.0 The Next Step	Page 13
5.01 Loading the LLPG	
5.02 Checking the LLPG	
5.03 LLPG and NLPG Maintenance	
6.0 Intelligent Addressing Services	Page 14
7.0 Important information about your Data Transfer File	Page 16

List of Files Supplied

File Name	Contents	Status
COT IA report.doc	This document	
Street Gazetteer	Initial street gazetteer with additions and corrections supplied by Cotswold District Council and IA	Supplied as part of X1610_03.TXT
Unmatched_streets.xls	Unique street / locality / town combinations that could not be matched to the National Street Gazetteer	Supplied
Unmatched_raw.xls	Raw data with street names that cannot be matched to the street gazetteer	Supplied
Nostreet.xls	Raw data that does not appear to contain street names	Supplied
nopaon.xls	Raw data that does not appear to contain sufficient information to determine a PAON	Supplied
Deleted.xls	Raw data that Cotswold District Council have indicated should be deleted from the raw data files	Not supplied
Non_BS7666.xls	Raw data unable to be converted to BS7666	Not supplied
X1610_03.TXT	Draft LLPG in NLPG Data Transfer Format v6.3	Previously sent by Email
Pos_duplicates.txt	Possible BLPU and LPI duplications	Supplied

1.0 Initial Data Processing

1.01 Data Supply

Data Sets provided by Cotswold District Council

- Street Gazetteer
- Uniform
- Council Tax
- Electoral Registration

1.02 Data Cleaning

Intelligent Addressing have processed the data for Cotswold District Council through a number of stages (as outlined in the NLPG briefing pack):

1. Load and Initial conversion of data
2. Street data extraction and initial street clean-up
3. Street matching
4. Conversion to BS7666, data Merging and LLPG creation
5. Matching to NLPG and adoption of national UPRNs
6. Identification of possible anomalies in LLPG and candidate BLPUs in NLPG

Table One below indicates the level of matching of your data against an extract of the National Street Gazetteer for the Cotswold District Council area, as at the end of stage 3.

Table One
Initial Data Supplied and Matched to NSG

	Initial Load	USRN matches	Not Matched	Unique Streets	Deleted	No Street name
Street Gazetteer	1,864					
Uniform	45,914	27,868	9,751	1,025	0	8,295
Council Tax	41,157	28,477	4,519	858	0	8,161
Electoral Registration	36,767	26,283	3,820	557	0	6,664
TOTALS	123,838	82,628	18,090	1,879	0	23,120

1.03 Street Gazetteer

The initial data clean up process matched some 60.0% of the total number of addresses sent to Intelligent Addressing with the extract of the National Street Gazetteer for Cotswold District Council. This figure has been improved to effectively 82.0% after the removal of records with no streets, records marked for deletion and the inclusion of the work carried out by Cotswold District Council in resolving identified anomalies.

The initial extract from the National Street Gazetteer for Cotswold District Council included 1,864 streets with Unique Street Reference Numbers (USRNs) in the range 9400001 to 9401870. The resolution of street anomalies by IA and Cotswold District Council in Report Two

have resulted in an additional 293 streets being added to the National Street Gazetteer. These streets have been allocated USRNs in the range 9405000 to 9405292.

The complete street gazetteer has been included in the final output file that accompanies this analysis. It should be noted that the additional streets have been given the start date of 18th November 2003 and that these records have no start or end co-ordinate pairs.

1.04 Unmatched Streets

There are still 18,090 records from the source data that remain to be matched to the street gazetteer. These records may have been flagged as “unknown” on Report Two as further information was required about each property to decide which street the property was to be associated with.

A file called **unmatched_streets.xls** has been prepared containing details of the address information that we were unable to match against the extract from the National Street Gazetteer for Cotswold District Council. This file is similar to the street anomalies files previously sent to Cotswold District Council and provides a list of unique combinations of street, locality and town with the information grouped together by street, locality and town.

The associated file **unmatched_raw.xls** containing the original data that cannot be matched to the street gazetteer, including the original source reference and all of the original address elements in a single field separated by the “|” symbol has been prepared.

1.05 No Identifiable Streets

Within the raw data supplied there are a total of 23,120 records that do not seem to contain a reference to a street. These records often include rural properties where only a building and locality name are provided or, in the case of commercial property data, organization only records. In order to incorporate these records into the gazetteer, the appropriate street names and USRNs need to be added. These records are set out in the attached file **nostreet.xls**, containing the original data source reference and all of the original address elements in a single field separated by the “|” symbol.

Please refer to section 6.0 for information on how Intelligent Addressing can help you to incorporate these records into your LLPG.

1.06 Deleted Records

In some cases Intelligent Addressing are obliged to exclude a number of the records supplied from the final analysis on the basis that they have been marked for deletion.

These records may have been selected for deletion for the following reasons:

- 1) They were marked for deletion on Report Two
- 2) The records contained the text “deleted”, “demolished”, “do not use” or words to that effect

In the case of Cotswold District Council there were no such records and the file **deleted.xls** has not been supplied.

2.0 Merging of data sets

2.01 Conversion to BS7666

After all of the changes to the street gazetteer requested by Cotswold District Council have been made, each data set has been converted to comply with BS7666 and the individual data sets merged to form a draft Local Land and Property Gazetteer, LLPG, for Cotswold District Council.

Excluded from this creation process are any records with insufficient address elements to identify a Primary Addressable Object, such as records with street only addresses or records with only floor or flat related data. There are a total of 802 such records as shown in Table Two below. These records are set out in the attached file **nopaon.xls**, with the original data source reference and all of the original address elements in a single field separated by the “|” symbol

Table Two
Conversion to BS7666

Data Set	Initial Data Load	USRN Match	No PAON	To BS7666	Not to BS7666
Uniform	45,914	27,868	325	27,543	0
Council Tax	41,157	28,477	419	28,058	0
Electoral Registration	36,767	26,283	58	26,225	0
Total No Records	123,838	82,628	802	81,826	0

As can be seen from column six of Table Two, all of the records from the source data sets that have been matched to the street gazetteer have at least an identifiable PAON and have been successfully converted to BS7666. The file **non_BS7666.xls** containing details of the records that could not be converted to a BS7666 format would normally be supplied in the same format as the other anomalies files, containing the original data source reference and all of the original address elements in a single field separated by the “|” symbol.

To increase the number of records that have grid co-ordinates in the final LLPG all datasets without co-ordinates are matched to Code-Point on the basis of the postcodes contained within the records. This allows general grid co-ordinates to be added to the creation process.

2.02 Creation of LLPG

Table Three is a summary of the merging process that starts with the local gazetteer as the only data set that contains both residential and commercial property information, and detailed grid co-ordinate data.

Table Three
Creation of Local Land and Property Gazetteer

Data Set	Input Data	Full Match	Partial Match	Inserted BLPU	Inserted in LLPG
Uniform	27,543	262	426		26,855
Council Tax	28,058	18,452	2,060		5,713
Electoral Registration	26,225	18,742	6,866		2,450
Total of original LA data	81,826	37,456			35,018
Additional BLPU				1,105	
TOTAL					35,123

For each insert into the LLPG a new Basic Land and Property Unit (BLPU) record is created containing the UPRN and grid co-ordinate, together with an appropriate Land and Property Identifier (LPI) containing the address data. The UPRNs used are temporary numbers to ensure uniqueness of the data when it is submitted to the NLPG. The final UPRNs allocated are a combination of the UPRNs already in the NLPG and those allocated by the NLPG custodian to any BLPUs inserted from the Cotswold District Council LLPG.

The additional BLPUs in the penultimate row are PAON only records, added to ensure BS7666 compliance. They occur where there are LPIs with SAONs and PAONs without the equivalent PAON only record. For example there may be BLPUs with LPIs for Flat 1, 23 High Street and Flat 2, 23 High Street but no record for just 23 High Street.

These PAON only records are important, as there may be application data that refers to a building as a whole rather than to specific parts. They can be identified from the list of BLPUs as they will not have any corresponding entry in the cross-reference table.

2.03 Matching and Merging

The basis for the matching and merging of the data sets is an exact match between all elements of the address, Primary Addressable Object Name, Secondary Addressable Object Name and Unique Street Reference Number. However, a level of "fuzzy" matching has been added to reflect the fact that there are many different variations on the way a particular address may be held within the raw data files.

There are currently five levels of fuzzy matching that we apply during the creation process: -

- Identical SAON, PAON and USRN combinations with differing postcodes are assumed to refer to the same BLPU with all variations of the postcode stored as alternative LPIs;
- Flat 1 Alpha Court, High Road matched to 1 Alpha Court, High Road and again one BLPU and two LPIs would be created, one LPI marked as approved and one as alternative;
- Alpha House, 23 High Road matched to 23 High Road and one BLPU and two LPIs created;
- 24 to 26 High Road matched to 24 High Road and one BLPU and two LPIs created;
- The White Hart or similar matched to White Hart Public House and White Hart PH.

The postcode level of matching was introduced because, although postcodes are optional under BS7666, many authorities find them useful attributes held against a particular LPI. This is particularly true if mailing lists are to be generated from the LLPG.

2.04 Merging to NLPG

Once created, the LLPG was submitted to the NLPG custodian on behalf of Cotswold District Council. The NLPG custodian subjected the file to a series of checks for compliance to BS7666 and the various NLPG conventions and data transfer formats.

Once satisfied that the LLPG was of appropriate standard in terms of format, the NLPG custodian merged the LLPG with the NLPG. Given that Cotswold District Council are the official custodians for land and property data for their area, the LLPG has taken priority during this merge phase. This means that:-

- Where there is a record for a particular BLPU / LPI in both the NLPG and LLPG, the UPRN is passed back to the LLPG and any additional data, such as the grid co-ordinates, start date etc is updated in the NLPG.
- Where the LLPG has a BLPU/LPI that does not appear in the NLPG this information is added to the NLPG and a new UPRN allocated.
- Where the NLPG has a BLPU/LPI that does not appear in the LLPG, the status of this record is downgraded to candidate.

The candidate BLPU/LPI combinations are left in the NLPG pending further research by Cotswold District Council as local custodian and are not distributed as part of the NLPG.

During this NLPG/LLPG merge phase, application of the same "fuzzy match" rules outlined in 2.03 will result in some of the LLPG records being matched together.

The merge to the NLPG has resulted in a final NLPG/LLPG extract of 42,703 BLPUs, including 7,134 candidates and a total of 54,962 LPIs including 7,374 candidates and 12,018 alternatives.

2.05 Inserted BLPUs

Some authorities have commented on the apparently large number of additional BLPUs created in the draft LLPG, particularly when comparing the total of the combined Council Tax and Non Domestic Rates data sets or a Local Gazetteer / Address Point against the final total. In the case of Cotswold District Council There is already an apparent discrepancy between Council Tax (41,000), Electoral Registration (36,700) and Uniform (45,900) and a final NLPG/LLPG extract of some 42,000.

There are four possible reasons for these apparent discrepancies.

- There are a large number of records from all data sets that have been excluded from the analysis because they do not contain any street data or have yet to be matched to the street gazetteer;
- Both the Council Tax and Non Domestic Rates data sets are based on the collection points for revenue and therefore, in the case of sub divided buildings, will refer only to those sub divisions. However, planning data and crucially postal data, as used in Address-Point may refer to the building as a whole. Thus a number of “whole building” entries must be added to the total of Council Tax and Non Domestic Rates entries;
- In those cases where sub division data is in fact held in the supplied data sets, there is a strong possibility that differing descriptions between the data sets will cause apparent mismatches and therefore additional insertions. These differing descriptions are highlighted in the file of possible duplications detailed below and need to be viewed and resolved by Cotswold District Council;
- Finally, it is possible that there are entities within the other data sets that simply do not appear in the Council Tax or Non Domestic Rates data. Planning, Environmental Health and Housing data are prime sources of this type of information, with new BLPUs relating to groups of properties or larger areas or new combinations of sub-divisions.

3.0 Grid Co-ordinates

As far as possible, each BLPU created in the LLPG has been given a grid co-ordinate. The accuracy of the co-ordinate is indicated by the Relative Positional Accuracy field which contains a value from 1-5

Table Four
Analysis of accuracy of grid co-ordinates supplied

Relative Positional Accuracy		Number of BLPUs
1	Visual centre of BLPU	24,822
2	General point within BLPU	3,951
3	Nearest 100m grid square	1,246
4	Start of street	2,025
5	Postcode	10,271
9	Centroid of local authority area	388
Total		42,703

4.0 Data returned

4.01 Anomalies Files

The various files created during the initial processing and conversion to BS7666, Sections 1.0 and 2.01 are contained on the CD-Rom that accompanies this report. A summary list of the files supplied is set out on Page 2.

4.02 The LLPG

The local land and property gazetteer for Cotswold District Council has been returned to you as a comma separated file, X1610_03.txt. This file is in the current NLPG data transfer format, details of which can be downloaded from the NLPG web site, www.nlpg.org.uk.

All records within the file are marked with the DTLR code for Cotswold District Council, 1610.

Within the file are five record type for BLPUs, LPIs, Cross References, Provenance or Source, and Streets together with a single header and single footer record. All of the cross reference records have a type 23 code shown as the first field in each records. The type 23 records contain all of the reference numbers from the source data. The source data sets are identified in part by the reference numbers but also by the source field that is the third field from the end of each type 23 records and contains the following values:

Local cross-references

CTWO	Uniform
CTWC	Council Tax
CTWE	Electoral Registration

National cross-references

N13A/NEWA/ ADDA/MASA	Address-Point
N13C	Council Tax
N13N	Non Domestic Rates

Other cross-references

WOKS/U99C

These cross-references indicate that another data provider or user is holding data against properties in the Cotswold District Council area. These are mainly social services records, where another local authority is holding details of previous addresses etc. Cotswold District Council need not take any action regarding these cross-references.

4.03 Potential Duplicates

In addition to the draft LLPG, there is a file that contains a list of possible BLPU or LPI duplications, **pos_duplicates.txt**. The LPI duplications are primarily concerned with alternative or missing postcodes and have been included to enable you to quickly decide which, if any, postcode you would like to add to a particular LPI. Once a decision has been made, the LPI logical status field can then be used to indicate which LPI displays the approved address for the BLPU and which contain alternative address information.

The BLPU duplications are a more problematic area. During the LLPG creation process we are able to match apparently “different” addresses together to create single BLPUs with more than one LPI using the “fuzzy” matching rules. For example 1 Alpha House 23 High Street would be matched to Flat 1 Alpha House 23 High Street. However, there are some assumptions that cannot be made, for example Flat A, 23 High Street may or may not be the same as 23A High Street. Although these cannot be safely matched electronically and will be added to the draft LLPG as different BLPUs, they can be displayed together once processing has been completed for an individual assessment of the possible match to be made. These are the BLPU entries that are contained in **pos_duplicates.txt**

It is possible that some of these anomalies could have been resolved automatically during the merging process by an extension of the fuzzy matching rules outlined above. However, experience with a large number of local authority address data sets has shown that there is always the possibility of exceptions to any matching rules and it is felt that an element of user input is required to resolve the identified anomalies and preserve the quality of the finished local land and property gazetteer.

Many of the anomalies can be resolved by inspection of the data and will not require further research or on-site inspection. Your standard contract with Intelligent Addressing does not include the resolution of these anomalies. However should you require Intelligent Addressing to assist your local authority with the cleaning of these records we would be more than willing to offer advice on how this can be achieved.

Please bear in mind that if you undertake this work internally you must ensure that any merging, deletion or renumbering of individual BLPUs or LPIs is reflected in the cross-reference table.

5.0 The Next Step

5.01 Loading the LLPG

As a first step you must ensure that you can read the file X1610_03.txt and that you understand its contents. This is most easily achieved by actually loading the data into BS7666 compliant gazetteer software.

At a general level we will discuss this file with you to confirm that you are satisfied with the level of BS7666 compliance and the matching rates for your data. In the light of these discussions we may need to reprocess the data to incorporate any observations you may make.

5.02 Checking the LLPG

Once the general quality and format of the data returned to you has been accepted, you will have to begin the process of checking the data prior to the commencement of the LLPG-NLPG maintenance cycle.

There are a number of broad groups of checks to be done:-

- Reviewing the contents of `pos_duplicates.txt` (4.03).
- Adding or amending the grid co-ordinates for those BLPUs with a value in the relative positional accuracy field other than 1 (3.0).
- Reviewing any candidate BLPUs and LPIs.
- Reviewing the files created during the initial processing and conversion to BS7666 (1.0 and 2.01 - please refer to section 6.0 to see how Intelligent Addressing can help you)

Ideally the work will also include checking the returned data against a map base to highlight any properties that may have been missed during the creation process

5.03 LLPG and NLPG Maintenance

In all cases this work will involve inserts to, deletions from or changes to the LLPG which must be passed back to the NLPG custodian in the appropriate format, details of which are set out on the NLPG web site (Data Transfer Format v5.3 or v6.3). A useful document is the National Land and Property Gazetteer Guidance Document 1 (Developing and implementing a LLPG from and extract from the NLPG) which can also be downloaded from the NLPG web site.

Once you have started on this checking and maintenance work, please contact the NLPG National Custodian, Simon Barlow (0207 747 3503) to discuss a timetable and medium for the supply of the updates from your LLPG to the NLPG.

6.0 Intelligent Addressing Services

Intelligent Addressing (IA) can help local authorities improve their LLPG's by incorporating rejected records once they have been corrected, and by matching additional datasets to the LLPG to allow more departments within each local authority to link to the LLPG.

Records without street names

A spreadsheet of all records without street names, **nostreet.xls**, will now have been sent to you. In order to incorporate these records into the LLPG, a street name and its corresponding entry in the National Street Gazetteer (NSG) will have to be identified and keyed onto the spreadsheet.

These records may then be manually entered onto the LLPG using your gazetteer maintenance software, however this is likely to be time consuming and costly in terms of data entry and associated staffing costs, particularly if your authority is rural and has thousands of properties to key on manually.

IA is now offering to take back the **nostreet.xls** spreadsheet, complete with an additional column containing the USRN for each property, and match this data to your LLPG and supply back a Data Transfer File which will automatically insert the new properties into your LLPG - eliminating the need to manually key these properties into your LLPG. The only technical requirement for this service is that your gazetteer maintenance software can interpret status change/insert records - please check with your software provider.

Possible Duplicates

A text file of possible duplicates, **pos_duplicates.txt**, will now have been sent to you. This report shows records that are already in the LLPG, but where duplicates may have been detected.

In order to remove duplicates from the LLPG, the normal process is to close the duplicate BLPU and then swap the attributes (i.e. LPI, Provenance, cross-reference) from the now-closed BLPU to the BLPU that you wish to retain. This can be time consuming and depends largely on the functionality of your gazetteer maintenance software.

IA can help you remove duplicates, simply by providing a list of "Old UPRN" and "New UPRN" along with the LPI in question. IA can then generate a Data Transfer File, which will automatically close redundant BLPUs and re-assign the LPIs, provenances and cross-references to the correct BLPU. As with the "nostreet" records, the only technical requirement is that your gazetteer maintenance software can interpret status change/insert records.

Other LLPG Updates

IA can also issue Data Transfer Files to make bulk updates to your LLPG where required. For example, you may wish to standardize the use of the "posttown" field on the LPI record, without actually going into each LPI and checking/editing it manually. Costs for this service vary, depending on the complexity of the changes that you would like to make.

Additional Datasets

IA can match additional datasets to your LLPG and supply back a Data Transfer File which will insert new properties, cross references etc to enable you to link more systems to the LLPG. With many software vendors upgrading their software to the BS7666 standard, such as social services and housing, you may choose to use IA to link these other systems to your LLPG.

The process is similar to the initial processes for the creation of the LLPG. Your additional data will be matched against the NSG and any anomalies will be passed back for resolution. However, bearing in mind that your Report Two has already been completed, we will use this first to ensure that any street mismatches passed back to you have not been resolved already on Report Two.

The additional data is then matched against the LLPG and any new properties, changes, new cross-references etc. will be compiled into a Data Transfer File to enable automatic updating of your LLPG.

What to do next

For further information or to discuss your requirements contact Simon Barlow, NLPG Custodian at Intelligent Addressing.

Telephone: 020 7747 3500 extn 103
Fax: 020 7747 3501
Email: sbarlow@nlpg.org.uk

7.0 Contents of Data Transfer File

When loading your Data Transfer File into your BS7666 accredited gazetteer maintenance software, please check that you select the correct DTF version number.

Your transfer file is in Data Transfer Format version 6.3.

After loading your gazetteer, you should check to make sure the correct number of records have been loaded in.

Your Data Transfer File contains:

Record Identifier	Record Type	Number of records
10	Header record	1
99	Trailer record	1
11	Street record	2,184
21	BLPU record	42,703
22	Provenance record	90,889
23	Cross-reference record	126,079
24	LPI record	54,962
12	Street cross-reference record	0
13	ESU record	0
14	ESU co-ordinaterecord	0

If your software has not loaded the above number of records, check error log files etc. to identify any problems.