

## Geographical references in the Virtual Microdata Laboratory and Secure Data Service

Original geomarkers and postcodes are valuable fields for statistical use. However, they are directly recognisable identifiers and so are not allowed in the VML or SDS. Within the VML and SDS they have been replaced with alternatives which maintain the exact statistical characteristics of the original variables. These variables are internally consistent within the VML and SDS. They are however, no longer consistent with external geo-identifiers. This note explains the structure of the revised variables.

### 1. Variable characteristics

#### 1.1. Postcodes

Postcodes have two functions (1) to reference other geographies (2) as a source of information in their own right about geographical clustering. The VML postcode retains these characteristics.

##### *Looking up other geographies*

As a reference, the VML postcode is simply a one-to-one mapping which allows one to lookup a postcode in one of the index fields. The true postcode has been replaced consistently across all lookup tables in the VML and SDS. Users wishing to use the postcode to access higher-level geographies in the Postcode Directory, for example, can use the same merge procedures as under the pre-SDS (real postcode) system.

##### *Using the nested structure of the postcode*

Postcodes in the VML have a fixed hierarchical structure in a nine-character string variable:

Element:	postcode area	postcode district	space	postcode sector	postcode walk
Characters:	1-2	3-4	5	6-7	8-9

Each of the components is two alphanumeric characters long. Thus a researcher wanting to group observations into postcode districts would combine observations with the same first four letters; grouping into postcode sectors means ensuring that the first nine characters are the same.

Note that real postcodes are of variable length. Therefore code written to parse real postcodes in the pre-SDS era may need to be adjusted to reflect this fixed structure.

#### 1.2. Geo-markers

Eastings and northings are available in the VML/SDS files, and map to VML postcodes consistently. The distances between points are the true distances. The actual points do not reflect publicly available mappings and are not usable outside the VML/SDS.

#### 1.3 Map files

Map files (shape files) are in the process of being consistently replaced so that maps can be drawn within the VML/SDS. Again, VML/SDS shape files are not usable outside the VML/SDS.

### 2. Substantive changes to variables

As part of the conversion, it was noted that many of the ASHE home postcodes were poorly coded. These have been clarified. As a result of this, extra steps were taken to clean all postcode variables under a consistent framework. Although we believe these steps only have an affect on the ASHE home postcodes.

Annex 1 notes the cleaning steps that were carried out on postcodes.

### **3. Use**

Use of postcodes to lookup other geographies should require no change to coding.

Use of postcodes which relies upon them being a maximum of eight characters long will fail.

Parsing of postcodes into sub-sections is no longer necessary. All VML postcodes follow the same 2-2-1-2-2 structure.

Geo-markers have the same relative references; therefore for distance estimates no additional coding should be required. Where absolute references are required (eg for map drawing), shape files are consistent.

### **4. Abuse**

The VML/SDS geographical variables are internally consistent and should have all the desired statistical characteristics. There is no valid reason for trying to reverse-engineer the variables. Any users trying to do so, whether successful or not, will be banned from the VML and SDS, will be liable to loss of ESRC funding at a personal and institutional level, and may be guilty of a criminal offence under the 2007 Statistics and Registration Services Act, s39. Note that all VML/SDS sessions are recorded.

Researchers who suspect that some statistical activity cannot be carried out on the VML geographies are advised to contact the VML team in the first instance, rather than trying to develop their own solutions.

### **5. Contact Details**

For more information on the VML postcodes and the replacement process please contact the Microdata Analysis and User Support team on:

Email: [maus@ons.gsi.gov.uk](mailto:maus@ons.gsi.gov.uk)

Phone: +44(0) 1633 45 5844

## Annex 1

A cleaning process is applied to all postcodes prior to being replaced by the VML postcode. Although all postcodes are put through this process, we believe the ASHE home postcode to be the only variable affected by this cleaning process. The steps in this process are:

1. Remove all spaces from the postcode.
2. Remove any trailing zeros.
3. Replace “l” with “1” where they have been incorrectly entered.
4. Replace “O” with “0” where they have been incorrectly entered.