

## Geocoding the Locations of Schools for an Obesity Monitoring Programme

### **Scenario:**

As part of a campaign to monitor obesity levels in children, the weights and heights of every child entering the state school system is being recorded and then measured again five years later. To better understand the effect of the environment on obesity in children (e.g. access to green spaces such as parks, the nature of neighbouring retail outlets, sports facility, deprivation, and so on), the zip / postal codes of participating schools are being geocoded. This will enable the school to be located on the map and linked across to the characteristics of surrounding neighbourhoods.

(**Note:** In reality, it would be far more likely that we would be geocoding zip codes for patient addresses. We use school zip / post codes for the exercise here because they are in the public domain. For a 'real world' example of such a programme, see the links to the National Child Measurement Programme)

### **Data:**

**Pennstate\_school\_zipcodes.txt** – a text file, containing zip codes / addresses for some public schools in Pennsylvania

**UK\_school\_postcodes.txt** – a text file, containing post codes for a selection of participating schools in Hampshire, UK

Both files contain a mixture of formats for the post codes and some have more detail than others. The US file, has addresses, 5-digit zip codes (representing postal zones within cities, say), and '+4' zip codes (more detailed, representing blocks of addresses within postal zones). Similarly, the UK file has post code districts (SO10 would be an example), post code sectors (SO10 1 would be an example – roughly equivalent to a 5-digit zip code), and unit post codes (SO17 1BJ – roughly equivalent to a '+4 zip code').

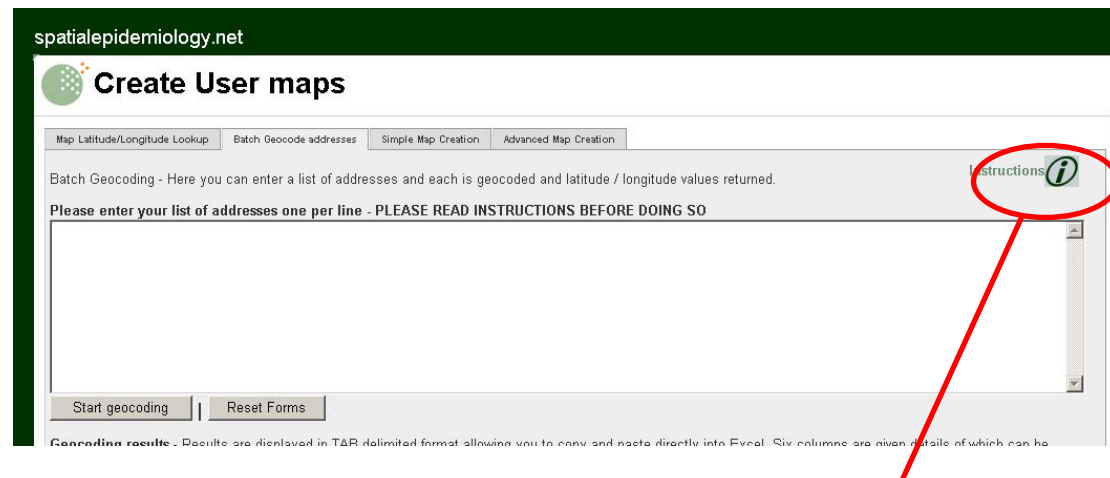
### **Instructions:**

There are now many ways of geocoding address or zip code data. In the UK, for example, post codes are maintained by a consortium known as GridLink®, which includes the national mapping agency, the Ordnance Survey, and the state-run postal service, Royal Mail. These organisations produce various products for geocoding addresses and zip codes, including Code-Point, Address-Point, and a National Post Code Directory, updated on a quarterly basis.

There are also various free resources that can be used to geocode data. Examples include OpenStreetMap and the Application Programming Interfaces (APIs) for software such as Google Maps.

In this exercise, we will use a web site that makes use of the Google Maps API to geocode addresses and zip/post codes (this is because the geocoding resources above are not in the public domain and therefore not available for this exercise):

- follow this link to an online geocoder  
[http://www.spatialepidemiology.net/user\\_maps/](http://www.spatialepidemiology.net/user_maps/)
- click on the tab marked 'batch geocoder' at the top of the screen



- Read the instructions on the top right-hand side by clicking here
- Make a note of the accuracy codes and the meaning of the 'number of addresses' output that will be created by this site
- Open up the text files for the schools in the US and UK (**Pennstate\_school\_zipcodes.txt**, **UK\_school\_postcodes.txt**). You can do this from within Word by choosing *file*, then *open* and then *files of type...all files*.
- Copy and paste all the zip codes and addresses for US schools into the top window in your web browser and press 'Start Geocoding' (note that you do not need to do anything else until a message appears on screen confirming that the geocoding operation has finished).
- Copy and paste the results into an Excel spreadsheet and delete the last two columns. Then copy the remaining first four columns.
- Back in your web browser, click on the tab marked *simple map creation*. In the upper box for the map title, and the lower box for the data to be used, delete the existing entries about African capital cities. Paste the data you just copied from Excel into the lower box and type in a meaningful title into the upper box, then press *Submit*.

spatialepidemiology.net

## Create User maps


Map Latitude/Longitude Lookup   Batch Geocode addresses   Simple Map Creation   Advanced Map Creation

### Simple Map creation

Here you can enter data in **TAB delimited format** and have the points displayed on a permanent map. Your data **MUST** be in the correct format (see below) and we recommend using the excel template below

Give a name to your map (this will be displayed as the title).

School-based Child Measurement Program

 Click here for an excel template for your data (you can copy and paste directly from excel).

Paste your own data here in the **TAB DELIMITED FORMAT - name latitude longitude description**

e.g. mypoint,51.5174, -0.1736, hospital1

19335	40.0210291	-75.7255192	5		
15459	39.7572014	-79.4591378	5		
17837	41.0014539	-76.9873477	5		
17844	40.9325133	-77.0794705	5		
15436	39.8019326	-79.7287703	5		
18411	41.482404	-75.737149	5		
105 Weikert Road, Laurelton	40.68145	-77.200704	8		
19001	40.1198332	-75.1253492	5		
15234	40.3698621	-80.0143656	5		
.....	.....	.....	-		

- Follow the link to the map of your data to view the result
- Alternatively, you can generate a set of points from these data within ArcGIS by:
  - Inserting some column headers in row 1 of your Excel spreadsheet (e.g. **zipcode, ycoord, xcoord, accuracy**)
  - Saving the resultant spreadsheet
  - Opening the new spreadsheet up in ArcGIS
  - Using *tools...add xy data* to specify that the *x field* is **xcoord** and the *y field* is **ycoord**
  - Specify the format of the co-ordinates by clicking on the *edit* button, then choosing *select, geographic coordinate systems*, then *WGS 1984* (the most widely used latitude and longitude format)
- Repeat this for the UK school post codes. Note that you may need to press the *Reset Forms* button to clear the data on the US schools.

### Questions:

1. Did all your data geocode successfully?
2. To what level of accuracy are the post / zip codes geocoded in these two examples?
3. More generally, what limitations do you think there might be in using geocoding to generate locations for disease cases?

Turn the page for suggested answers to these questions.

### ***Suggested Answers:***

1. You are likely to have found problems with at least one UK address, which was incorrectly formatted (and any 'blank' rows at the end of your data too). Where people supply their own addresses and post codes, mistyping of such information is common.
2. None of the results from this particular geocoder are very precise. The UK post codes are geocoded at the post code sector level (i.e. SO40 1), and sometimes the geocoder returns several different co-ordinates for the same post code. With this particular geocoding software, the geocoder does not capture the unit post code detail (i.e. SO40 1BX can only be geocoded as SO40 1). Similarly, the US zip codes are geocoded at the 5-digit postal zone level only, so the '+4 detail' is lost.
3. More generally, some potential limitations of geocoding include:
  - At least some post codes are likely to be incorrect or incomplete and so only a subset of cases or deaths may be represented spatially. Often, the cases or deaths that can be represented spatially may be different from those that are not (e.g. long-term residents will remember their post / zip codes, whilst transitory residents will not).
  - Mortality is usually geo-referenced using residence at time of death, but this may not reflect the location where an illness was contracted. Similarly, residence at time of diagnosis may not reflect residence at time of infection or exposure to a hazard.
  - Depending on the data source used for address or zip/post code locations, the locations may not be that precise. The exercise here using the Google Maps API geocoder shows this very clearly. As another example, in the UK, the most detailed post codes are known as unit post codes. These represent small groups of individual property addresses (usually less than 50). The co-ordinates for the unit post-code are for the central property within the group, but in rural areas, the other properties may be scattered over several square kilometres around this central property.