

# **‘Freelocation’: combining open-source data and methods for accessible location analysis**

Robin Lovelace<sup>1</sup>,

<sup>1</sup> Department of Geography, The University of Sheffield, SHEFFIELD, S10 2TN United Kingdom

(0044) 01142 227 973

robin.lovelace@shef.ac.uk

KEYWORDS: open-source, location analysis, store location, participatory GIS

## **1. Introduction**

Location analysis is a field of geographical research set in the positivist paradigm of spatial science (Johnston 2000). Location analysis has a close relationship with GIS software (Birkin 1997) and can be framed mathematically, as a constrained optimisation problem (ReVelle & Eiselt 2005). Our argument is that these features make location analysis inaccessible, to all but those with specialist skills or large bank accounts to pay for expert knowledge.

Due its commercial relevance, the academic field has a large overlap with industrial applications. Perhaps due to the commercial advantages conferred to businesses which can harness the art of location analysis, much of the data and methods are ‘located’ in the private sphere (see Hernández et al. 2005). Public dissemination of findings is rare. Up-to-date datasets on supermarket locations, ownership and floor area, for example, are generally available only to large companies able to hire consultants such as. However, this information may be of use to people on lower budgets too, such as charities, small businesses, and concerned citizens.

Based on this research problem, the purpose of this paper is to outline free data and methods for store location analysis, an approach we have labelled ‘freelocation’ for brevity. Following the open-source software movement, we use the word ‘free’ here to refer not only to financial costs – zero. The paper will explain how the ‘freelocation’ approach is accessible to potential users – through an accompanying tutorial – and modifiable – with reference to the constant evolution of the collaborative Open Street Map (OSM) and open source GIS software.

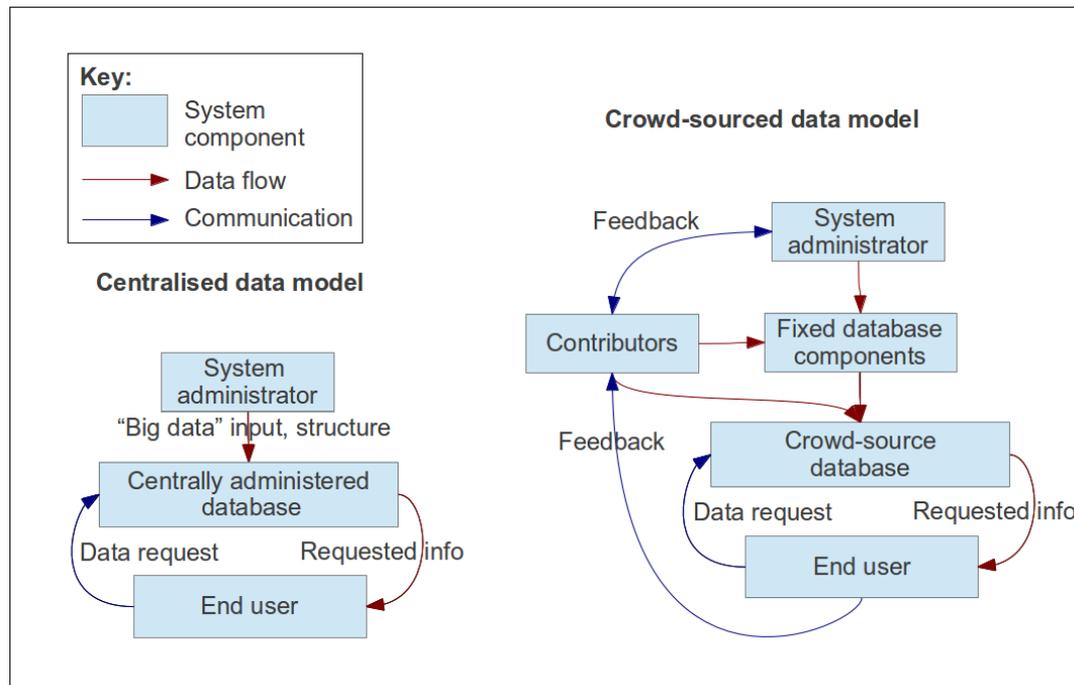
## **2. Framing: crowd-sourced vs centralised data model**

The paper is predominantly practical. It will explain:

- the pros and cons of crowd-sourced datasets for location analysis,
- how to access and process them to deal with problems of consistency, reliability (ground-truthing) and duplicity and
- options for disseminating the information including Voronoi polygons of supermarket territories and the creation of an interactive online map that can be used

by the public.

Nevertheless, this relatively narrow case study of store location analysis (in two cities, one with a strong OSM community, one without) will be framed in theoretical terms, as a test of the potential for the crowd-sourced data to overcome some of the limitations of the traditional centralised model. Figure 1 illustrates this theoretical framework.



**Figure 1:** Schematic diagram of centralised and crowd-sourced data models.

### 3. Methods

Based on the open-source ethic mentioned in the introduction, the paper will showcase only methods that have been conducted on open-source software. The supplementary tutorial (which is based on a workshop session for an MSc module in GIS and data visualisation) demonstrates the methods in a didactic, step-by-step manner. The paper will focus on explaining the methods, why they were chosen in the context of other options, and their theoretical underpinnings.

The central stages of the method are as follows:

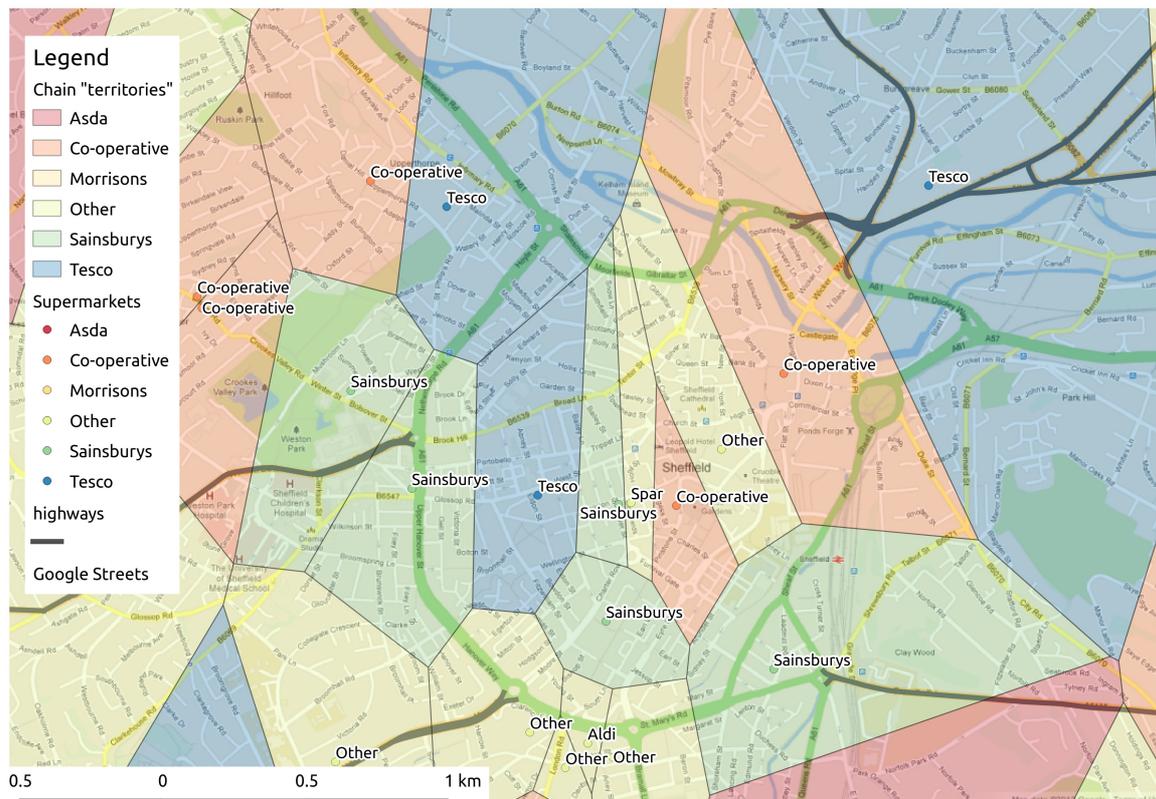
- Downloading the OSM data – this can be done in three main ways with increasing difficulty and geographical coverage – up to the 6 Gb “planet.osm” file.
- Filtering the data to access only the features we are most interested in: supermarket and other food shop points and polygons using command-line and GUI tools
- Reading the OSM data into a PostGIS database or a desktop GIS
- “Cleaning” the messy raw OSM data for consistency, manageable numbers of store

categories, duplicity and error.

- Analysing the datasets: this section will showcase methods for the investigation of “food deserts”, “spatial monopolies”, and accessibility.
- Ground-truthing, including distance-based and fieldwork-based options.

## 4. Initial findings

The central aim of the paper, to investigate the potential for a 'freelocation' approach to location analysis, has already been met in the context of supermarket store location analysis. Data for Sheffield (which has a strong OSM community) were downloaded from a wider South Yorkshire OSM dataset and manipulated using QGIS to extract supermarket stores. It was found that information on the supermarket chain (Tesco etc), location, and in some cases size were available to anyone using open-source GIS and data analysis tools. However, issues of data consistency, duplicity, accuracy and coverage were identified. Methods for overcoming these issues were implemented, leading to the following “territory map” using Voronoi polygons (Figure 2). This map was uploaded as an interactive layer using OpenGeo's Geoserver software to a virtual machine running on a free trial of Amazon's Elastic Computing 2 (EC2) service. This has the potential to help identify which supermarket chains operate spatial monopolies where they live and, based on more OSM data, to allow the identification of alternative independent food shops.



**Figure 2:** Voronoi polygons illustrating the competition for space and market share between supermarkets surrounding central Sheffield, with Google Streets online map in background.

More broadly, a preliminary finding is that the 'freelocation' approach outlined in the introduction may be of great use to educators, small businesses, charities and concerned

citizens. The crowd-sourced data from OSM may not be of the same high quality as the proprietary data that geolocation consultancies have access. However, open source data may in some cases be richer and more up-to-date. The main advantage of the free methods and data outlined in this paper seem to be that they are available to anyone. The case study of supermarkets flags issues about potential inaccuracies in the information, the need to improve and 'clean up' OSM data, and the importance of validation (as illustrated by the [osmgb.org.uk/](http://osmgb.org.uk/) project). Provided these issues are taken into account and acted upon, the 'freelocation' approach outlined in this paper has the potential to make the 'dark art' of location analysis accessible to a much wider audience.

## 5. Acknowledgements

This work is funded by the EPSRC, through the E-Futures Doctoral Training centre. Many have provided feedback on the early stages of the research including the local food initiative Grow Sheffield.

## References

- Birkin, M., 1997. Retail location modelling in GIS. In P. Longley & M. Batty, eds. *Spatial analysis: modelling in a GIS environment*. London: John Wiley & Sons, Inc.
- Hernández, T. et al., 2005. International Journal of Retail & Distribution Management Emerald Article : The art and science of retail location decisions The art and science of retail location decisions. , (2000).
- Johnston, R., 2000. Locational Analysis. In R. J. Johnston et al., eds. *The dictionary of human geography*. London: Blackwell Publishing Ltd, p. 958.
- ReVelle, C.S. & Eiselt, H. a., 2005. Location analysis: A synthesis and survey. *European Journal of Operational Research*, 165(1), pp.1–19.

## Biography

*I am a final year PhD student. Oil depletion and environmental concerns inform my research, which uses spatial microsimulation and commuter flow data to investigate the energy costs of transport to work. My desire to use geography for the greater good has driven me to use open source software throughout.*