

Ordnance Survey Position paper: Social Networks as a Future Geographical Data Source¹

Ian Holt and Jennifer Green,
Ordnance Survey of Great Britain
{Ian.Holt, Jenny.Green}@ordnancesurvey.co.uk

As the national mapping agency of Great Britain, Ordnance Survey's position at this workshop is that of a data vendor. Ordnance Survey has an active research department, covering many different areas, allowing us to stay ahead and move with the growth of the geospatial information industry.

We manage one of the largest vector spatial databases in the world, describing half a billion 'real world objects' such as buildings, road segments, fields and post boxes. Our surveyors and photogrammetrists submit around 5000 changes to the database every day, recording accuracies of around 10 cm precision on the ground. Integral to using this large volume of data are the names given to the objects, such as place names, and improving this data will be the use case in this paper.

In Ordnance Survey data, place names are represented as text within graphical products, based on paper maps, and as descriptive fields on digital products. Gazetteers are provided to facilitate searching for geographical locations and usually indicate a place's location using coordinates of a point that is within or near it. This can suffice for providing a visual result, where the purpose is to centre a displayed map on the feature. However this kind of data gives no indication of a place's size, extent or relationship to other features. It is up to the user to know that the name they have searched for a city and to infer its size and relations from the displayed information.

These current solutions lack both accurate location and name information and does not deal with the vague extents that people use to define areas nor the possibility of multiple names for a place (see Figure 1). Investigation into solutions to these problems has led us to look at ways to improve our names data sets and explore the collection of a vernacular place names data set, the geography reflected in ordinary people's language and perceptions. Examples include "the city centre" or "the Midlands". In the most extreme cases these can be areas which have no meaning except to a very local group.

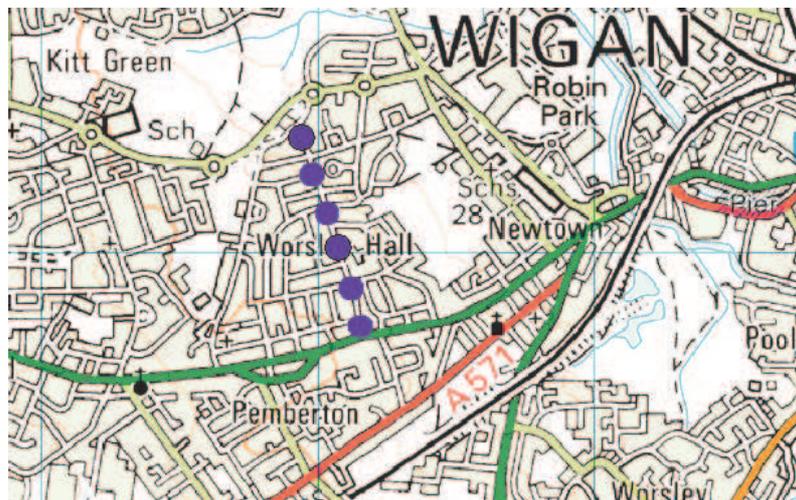


Figure 1 - Is the street marked by dots in Pemberton? Gazetteers and addressing data said not (in 2005), yet locals on the street used the name when calling emergency services.

Although investigation into the capture of vernacular place names has been going on for some time, current trends are moving towards looking at the web as the main source from which to extract data. Some have used existing online gazetteers and some use data mining and web scraping techniques to gather place names based on spatial terms used within the page.

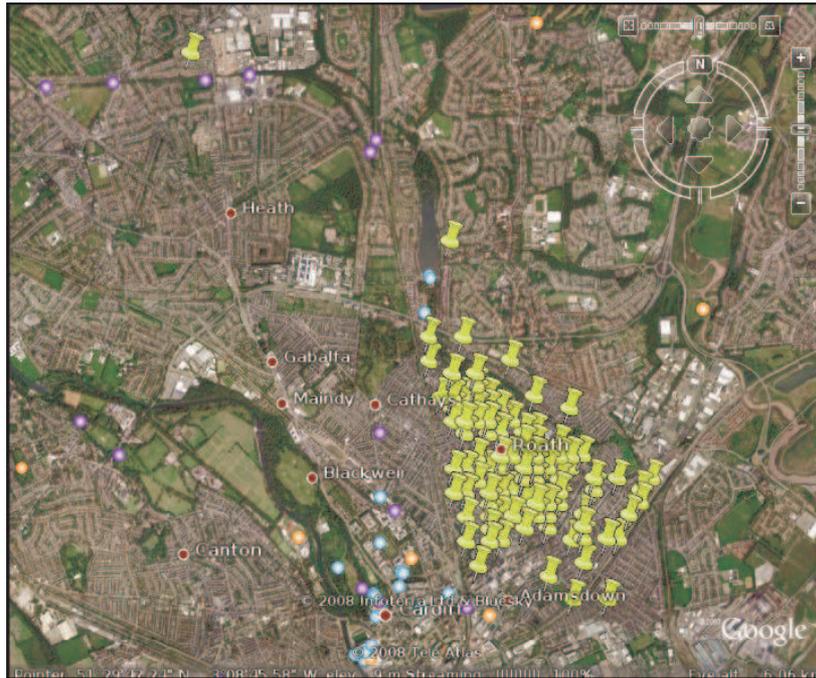


Figure 2 – Points labelled ‘Roath’, mined from Gumtree in Twaroch et al (April 2008)

The main problem with these techniques is identification of suitable sources. There is much research within this area that make use of sources such as Flickr, Geograph, Gumtree and Geonames but using individuals websites is proving more problematic due to their lack of formal structure.

Social networks offer unique aspects for place name data collection. Firstly, as opposed to individuals' web sites, much of the information tends to be very structured – users are encouraged to describe themselves and their activities, making it easier for automated data analysis. Secondly, much of the content is very dynamic, often using descriptions and language in common usage not necessarily authoritative in the case of place names, thus giving a more ‘vernacular’ snapshot.

Using current data capture techniques we have successfully demonstrated, with our research partners, that extracting vernacular place names from social networks is more than possible. Using text mining and natural language processing techniques Twaroch et al (April 2008) demonstrated the extraction of vernacular place names from Gumtree, as well as using fuzzy and uncertainty modelling to produce a representation of the names geographic extent, see Figure 2. Whilst we continue to refine these we also see future work in combining data from different social networks and refining the quality of that information to provide an even more comprehensive dataset.

Whilst we have a specific research requirement for investigating the content within social networks it is clear that there is value for all in the data itself. As more studies are carried out in this area, such as Ackland (2008) which are investigating the use of social networks as a data source for social science research, users will gain invaluable access to information and the consumer will see their service improved from the analysis of the data they provide.

Some of the questions we believe our use case and similar studies have raised and would be valuable for this workshop to discuss are:

- Is the data created in a social network a marketable commodity? Social networks represent a semi-structured, often highly dynamic, data source which has huge potential as a commodity for data and trend analysis.
- What are the possible legal and intellectual property issues with regards to this information?

- Should there be a common shared set of standards across social networks for data sharing, possibly along the lines of the Data Portability Project (<http://www.dataportability.org/>)
- How is user data security and anonymity dealt with? No individual either directly or indirectly should be identifiable.

References

Ackland, R. (2008), "Using Facebook as a Data Source and Platform for e-Researching Social Networks," refereed paper presented at the Fourth International Conference on e-Social Science, 18-20 June 2008, University of Manchester.

Twaroch, F. A., Clough, P. D., Pasley, R., Jones, C. B., Mansbridge, L., & Davies, C. 2008, "Social Web Sources of Vernacular Geography", in Submitted to: Geographic Information Science Conference GI Science '08, Utah, USA, September 2008.

Twaroch, F. A., Jones, C. B., & Abdelmoty, A. I. "Acquisition of a Vernacular Gazetteer from Web Sources", in LocWeb 2008: First International Workshop on Location and the Web, 22 April 2008, Beijing, (in press).