

Exploiting Volunteered Geographic Information to describe Place

Ross S. Purves and Alistair J. Edwardes

¹Department of Geography, University of Zurich, Switzerland
 Tel. (+ 41 44 635 6531) Fax.(+ 41 44 635 6848)
 ross.purves|alistair.edwardes@geo.uzh.ch, www.geo.uzh.ch/~rsp

KEYWORDS: Volunteered Geographic Information, data mining, place

1. Introduction

Traditional geographic information focuses on well-structured spatial data, generally based on either crisp objects or continuous fields, which usually reflect an institutional viewpoint and the purpose for which a dataset was collected. The attributes of such data are typically single valued at a given location or for a given object and thus, by definition, cannot reflect multiple viewpoints. Such representations are of course well suited to many tasks where we apply geographic information (for example querying the rateable value of individual houses or the elevation of a particular point in space). Equally, though it is clear that if we asked ten GISRUK participants about Manchester, we would hear ten differing perspectives on the *place* that is Manchester and that these perspectives would reflect the experiences and backgrounds of those that we asked. To date, relatively little research in GIScience has focussed on these multiple perspectives, in contrast to work in human geography where the notion of *place* is seen as central to the discipline itself (Cresswell, 2004). Fisher and Unwin (2005) recognised this gap within GIScience research:

“GI theory articulates the idea of absolute Euclidean spaces quite well, but the socially-produced and continuously changing notion of place has to date proved elusive to digital description except, perhaps, through photography and film.” (p. 6).

Equally, GIScientists have recently seized upon the potential of opportunities provided by Volunteered Geographic Information (VGI). VGI is loosely defined by Goodchild (2007) as “a special case of user-generated content” whereby geographic information is created outside of a formal, official framework. Thus, for example OpenStreetMap aims to create a user-generated, editable map of the world based on data collected by volunteers (www.openstreetmap.org/). However, in this paper we are interested not in the use of VGI as a means to replace or replicate traditional structured forms of geographic information, but rather as a way of starting to describe *place* in GIScience. We contend that place lies at the opposite end of a continuum of geographic perspectives from space (Figure 1) and that VGI provides us with a real opportunity to describe place.

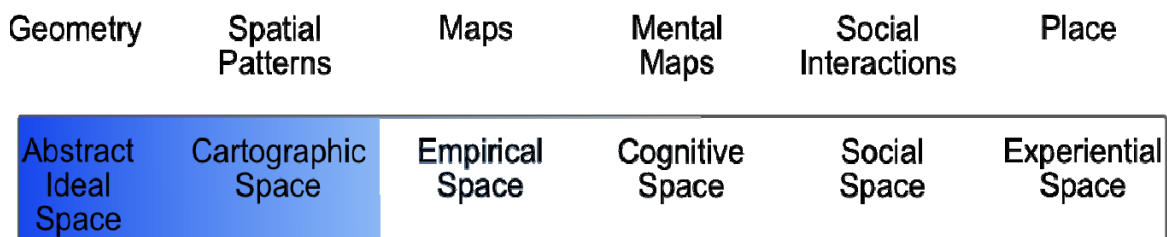


Figure 1. The space–place continuum (from Edwardes, 2007)

We set out to illustrate our contention through a case study. As part of a European project which aims to automatically add indexing terms to geo-located digital photographs we have been exploring Geograph, a classic example of VGI. The Geograph project has the aim of collecting “geographically representative photographs and information for every square kilometre of the UK and the Republic of

Ireland.” Contributors submit photographs representing individual 1km grid squares for moderation and images are uploaded together with descriptive captions to a searchable web site. As of September 2007 around 5000 users had contributed more than half a million photographs. In this paper we set out to explore two questions pertaining to the use of Geograph in exploring notions of place.

- 1) Can volunteered sources of geographic information provide new perspectives on describing place?
- 2) What limitations must we be aware of when working with such data?

In the rest of this paper we first briefly describe how place might be described, before setting out the methodology we applied in our experiments with Geograph. Finally, we describe the results of these experiments from the perspective of the questions posed above.

2. Describing place

In this paper we are primarily interested in methods for eliciting terms to describe place. Previously, much research in this domain has focussed on identifying so-called basic levels, whereby a basic level is one which is both informative and summative – thus for example *littoral zone* provides a detailed geomorphological description of a coastal feature, whereas coastline provides a very general description. Beach is both informative (in that it suggests a set of particular qualities and activities that are not offered by, for example, coastline) and summative (in that it encompasses a range of possible subclasses). Previous work to both identify basic levels and associated descriptive terms has been largely based on human subject testing. A key difficulty here is that such experiments are complex to organise, difficult to repeat and generally have relatively small numbers of subjects. We thus wished to explore the extent to which we could exploit Geograph in both replicating past research on place and forming new perspectives.

We discuss here a set of three experiments, some of which were previously described in more detail in Edwardes and Purves (2007). The first experiment set out to compare the frequencies of terms suggested as basic levels in previous empirical research with the frequency of occurrence of the same terms in the Geograph collection. In the second experiment, we explored the co-occurrence of descriptive terms with a selection of basic levels. Finally, in the third set of experiments we explored the relationships between a set of 1381 nouns identified manually within the Geograph dataset which occurred more than 100 times. To explore these relationships we analysed clusters of significant groupings amongst the nouns, using cosine similarity and hierarchical clustering techniques (Salton et al., 1975).

3. New perspectives on place?

In the first experiment, where we compared the term frequencies in Geograph with previous participant research we found that the terms identified were in many cases very similar (Edwardes and Purves, 2007). Thus, we showed that the rankings of the following terms identified by Battig and Montague (1969) as being category norms (broadly equivalent to basic levels) were significantly correlated: Mountain; Hill; Valley; River; Rock; Lake; Canyon; Cliff; Ocean; Cave. This first result suggests that Geograph can be used as a proxy for participant experiments.

In our second experiment (Table 1) we explored how people described two of the basic levels identified above (mountain and hill) and two further basic levels (beach and village) identified in previous research and in the Geograph data (Edwardes and Purves, 2007).

Table 1. Co-occurring descriptive terms with basic levels beach, village, hill and mountain

Beach n=2824			Village n=12707		
Activities	Elements	Qualities	Activities	Elements	Qualities
Surfing	Shingle	Sandy	Conservation	Pub	Deserted
Bathing	Sand	Deserted	Reading	Shop	Pretty
Defence	Cliff	Eroded	Fishing	Inn	Green
Swimming	Headland	Soft	Playground	Church	Quiet
Tourism	Bay	Rocky	Defence	Housing	Lovely
Wading	Sea	Warm	Bowling	Edge	Pleasant
Protection	Rock	Glacial	Tourism	Cottage	Beautiful
Sport	Coast	Low	Football	Main Road	Remote
Shipping	Shore	Beautiful	Entertainment	Village green	Unusual
Golf	Island	Lovely	Sitting	Stone	Large
Hill n=16232			Mountain n=1256		
Activities	Elements	Qualities	Activities	Elements	Qualities
Climbing	Fort	Steep	Biking	Peak	Distant
Skiing	Top	Distant	Kayaking	Summit	Black
Holidays	Summit	Wooded	Outing	Ridge	Remote
Observation	Horizon	Black	Mountaineering	Moorland	Rocky
Sitting	Ridge	Rough	Escape	Quarry	Grassy
Walking	Sheep	Grassy	Walks	Stream	Steep
Running	Valley	Round	Fun	Sheep	Natural
Cycling	Side	Big	Racing	Forest	Dark
Preservation	Trees	White	Climbing	Top	Broad
Escape	Track	Broad	Cycling	Path	Running

In the third experiment we looked at the associations between nouns through clustering techniques and used these to build sets of related terms. Table 2 shows illustrative examples of some of the associations derived using these techniques.

Table 2. Clusters and illustrative associated terms

Cluster	Associated terms
Road network	road, roundabout, junction
Hills and valleys	hill, valley, bridleway, walkers
River systems	river, bank, waterfalls, floodplain, water, streams, valleys, levels, aqueduct, sewage, anglers, salmon, otter
Buildings	home, wall, manor, doorway, gable, mansion, roof, glass, house, grounds, architect, foundation, columns, castle
Woodland	Forest, woodland, plantation, oak, beech, pony, heathland, commoner
Arable agriculture	Fields, pasture, crop, wheat, farmer, harvest, barley
Mountain landforms	Glen, beinn, allt, meall, loch, corrie, sgurr, garbh
Coastal features	Sea, beach, bay, peninsula, headland, sands, islands, creek, cave, foreshore, mud
Rural landscapes	Pond, fence, orchard, wire, birds, bushes, flowers, hedgerow, scenery, flock

We contend, that these experiments illustrate how we can use VGI to start to address the description of *place*, which as we have illustrated, is a more descriptive and less geometric form of geographic information. Thus for example, we identified in Table 1 typical descriptors for mountains, beaches, hills and villages. In turn, such descriptors could be extracted from more traditional sources of spatial data (e.g. identifying village greens in topographic data) to identify and locate typical and atypical examples of villages as characterised by Geograph users.

Equally, the Geograph dataset reveals notions of place with respect to the British Isles. For example, the term ‘Sea’ is used in preference to ‘Ocean’, and geographic features termed ‘Hill’ and ‘Loch’ are far more commonly encountered than those such as ‘Mountain’ and ‘Lake’. Our work on identifying descriptive terms also points to potentially stereotypical viewpoints. For instance the terms found most commonly with ‘Village’ were somewhat bucolic – “green, quiet, inn”. Whilst this is probably not due to a conscious bias by the contributors, it may relate to unconscious avoidance of taking pictures that are not in some way aesthetically pleasing.

In this paper we have ignored the spatial distribution of the terms which we discuss – that is to say, we have not considered if there are variations in the way that place is described within the British Isles – this is the subject of parallel work described in a paper by Dykes et al. at this conference.

6. Acknowledgements

This research reported in this paper is part of the project *TRIPOD* supported by the European Commission under contract 045335. We would also like to gratefully acknowledge contributors to Geograph British Isles, see <http://www.geograph.org.uk/credits/2007-02-24>, whose work is made available under the following Creative Commons Attribution-ShareAlike 2.5 Licence (<http://creativecommons.org/licenses/by-sa/2.5/>).

References

- Battig W and Montague W (1969) Category norms for verbal items in 56 categories: a replication and extension of the Connecticut Norms *Journal of Experimental Psychology* **80**(2) pp1-46
- Cresswell T (2004). *Place: A Short Introduction*. Blackwell, Oxford.
- Edwardes AJ and Purves RS (2007) *A theoretical grounding for semantic descriptions of place* LNCS: Proceedings of 7th Intl Workshop on Web and Wireless GIS W2GIS 2007 Ware M and Taylor G (eds) Berlin: Springer pp106-120
- Fisher P and Unwin D (2005) *Re-presenting Geographical Information Systems in Peter Fisher P and Unwin DJ (eds), Re-presenting GIS, John Wiley & Sons, London. pp 1-17*
- Goodchild MF (2007) Citizens as sensors: the world of volunteered geography *Geojournal* **69** pp211-221
- Salton G Wong A and Yang C S (1975) *A Vector Space Model for Automatic Indexing* Communications of the ACM **18**(11) pp613–620

Biography

Ross Purves is a lecturer at the Department of Geography of the University of Zurich, where Alistair Edwardes is a post-doctoral research assistant.