

Article

Ensuring VGI Credibility in Urban-Community Data Generation: A Methodological Research Design

Jamie O'Brien ^{1,*}, Miguel Serra ², Andrew Hudson-Smith ¹, Sophia Psarra ², Anthony Hunter ³
and Martin Zaltz Austwick ¹

¹The Bartlett Centre for Advanced Spatial Analysis, University College London, London, W1T 4TJ, UK;
E-Mails: jamie.o'brien@ucl.ac.uk (J.O'B.), a.hudson-smith@ucl.ac.uk (A.H.-S.), m.austwick@ucl.ac.uk (M.Z.A.)

²Space Syntax Laboratory, Bartlett School of Architecture, University College London, London, NW1 2BX, UK;
E-Mails: s.psarra@ucl.ac.uk (S.P.), m.serra@ucl.ac.uk (M.S.)

³UCL Computer Science, University College London, London, WC1E 6AA, UK; E-Mail: a.hunter@cs.ucl.ac.uk

* Corresponding author

Submitted: 11 March 2016 | Accepted: 5 June 2016 | Published: 27 June 2016

Abstract

In this paper we outline the methodological development of current research into urban community formations based on combinations of qualitative (volunteered) and quantitative (spatial analytical and geo-statistical) data. We outline a research design that addresses problems of data quality relating to *credibility* in volunteered geographic information (VGI) intended for Web-enabled participatory planning. Here we have drawn on a dual notion of credibility in VGI data, and propose a methodological workflow to address its criteria. We propose a 'super-positional' model of urban community formations, and report on the combination of quantitative and participatory methods employed to underpin its integration. The objective of this methodological phase of study is to enhance confidence in the quality of data for Web-enabled participatory planning. Our participatory method has been supported by rigorous quantification of area characteristics, including participant communities' demographic and socio-economic contexts. This participatory method provided participants with a ready and accessible format for observing and mark-making, which allowed the investigators to iterate rapidly a system design based on participants' responses to the workshop tasks. Participatory workshops have involved secondary school-age children in socio-economically contrasting areas of Liverpool (Merseyside, UK), which offers a test-bed for comparing communities' formations in comparative contexts, while bringing an under-represented section of the population into a planning domain, whose experience may stem from public and non-motorised transport modalities. Data has been gathered through one-day participatory workshops, featuring questionnaire surveys, local site analysis, perception mapping and brief, textual descriptions. This innovative approach will support Web-based participation among stakeholding planners, who may benefit from well-structured, community-volunteered, geo-located definitions of local spaces.

Keywords

community participation; data credibility; geo-spatial quantification; participatory methods; quality of data; urban planning; volunteered geographic information

Issue

This article is part of the issue "Volunteered Geographic Information and the City", edited by Andrew Hudson-Smith (University College London, UK), Choon-Piew Pow (National University of Singapore, Singapore), Jin-Kyu Jung (University of Washington, USA) and Wen Lin (Newcastle University, UK).

© 2016 by the authors; licensee Cogitatio (Lisbon, Portugal). This article is licensed under a Creative Commons Attribution 4.0 International License (CC BY).

1. Introduction

Effective urban planning must reflect citizens' experiences of socio-spatial inequalities (cf. UN-Habitat, 2009, p. xxiii), which relate to place-specific factors of social distinction and cultural identity (Cassiers & Kesteloot, 2012). Such inequalities are widening due to major economic and infrastructural changes in many cities around the world (cf. UN-Habitat, 2012, p. 83), and reveal uneven distributions of economic, social and cultural resources (cf. Marcuse, 2002, pp. 11-34; UN-Habitat, 2009, pp. 31-39). The effects of spatial inequalities include the urban population's uneven levels of access to their city's resources, resulting from both physical barriers and social exclusions (cf. Grant, 2010, pp. 5-9).

Communities as urban forms are distinguished by their network interactions at spatial, social and semantic levels. This means that people in a segregated spatial enclave, for example, develop bounded networks that sustain the community, based on proximity, needs and resources and cultural affinities. Spatial inequalities are apparent in community environments as people's quality-of-life capabilities are affected by, for example, their self-identification or sense of entitlement relating to a place; yet these inequalities are not always directly visible to community participants (discussed by Dorling, 2012, pp. 220-222).

Community participation may assist urban planners in understanding the non-discursiveness of community formations, especially in highly localised spaces. There has been renewed emphasis on empowering community participants to articulate their positions in terms of their localized definitions (Chambers, 1995). Many indicators of inequalities that are highly salient relate to visible urban-fabric or socio-cultural differentiations (cf. Chokor, 1991; Veiga, 2012). Less salient, however, are information-based indicators such as social connectedness and access to economic opportunities (cf. Morsey, 2012; United Nations, 2013, p. 77). Current analyses of spatial inequalities are generally based on broad (and costly) economic and social surveys (e.g. at the city level). Community participation based on volunteered geographic information (VGI) provides a cost-effective means of gathering a wide range of data relating to localities and their inter-relationships with urban infrastructures and political agencies (Haklay, Antoniou, Basiouka, Soden, & Mooney, 2014).

While participation based on VGI offers benefits to both researchers and participants in urban planning, it also poses challenges in its uptake and use in professional and scientific domains (Tulloch, 2007). One challenge relates to the diversity of community among data producers, who diverge in their levels of expertise and the nature of their motivations for participating (Coleman, Georgiadou, & Labonte, 2009). Another challenge relates to the credibility of using volunteered information, which has been addressed by Flanagan and

Metzger (2008). Credibility as a quality may stem from the data *accuracy* of scientific enquiry, or from its *believability* for social or experiential engagement. However, believability based on perception of quality may result from the participant's 'rule-of-thumb' cognitive heuristics, rather than a well-defined rationale for evaluation (Flanagan & Metzger, 2008). The authors maintain how information abundance, resulting from Web participants' multifarious engagements with geo-location technologies, is associated with paucity in credibility and a so-called 'context deficit' for data management. Hence, the range of data sources and styles of metadata generation (for example, of image tagging) have resulted in low credibility of voluntary data overall. This paucity has, in some cases, been overcome by 'credentializing' participants in the required domain of expertise (such as, for example, in wildlife identification; cf. Goodchild, 2007).

Upholding data quality may thus place a burden of responsibility on data providers in geospatial data-sharing communities (Goodchild, 1995, pp. 413-425). Hence Lush, Bastin and Lumsden (2012) have outlined a requirement for providing standard indicators of geospatial metadata credibility. The authors have offered a range of indicators to uphold data quality in decision-making domains, including possibilities for credentialing the data provider through a so-called 'community advice' process. In such a process, peer review of geospatial data could allow users and experts to provide commentary on datasets provided, including any limitations and problems associated with the data.

2. "Visualizing Community Inequalities"

The work outlined in this paper stems from a broader project at University College London, "Visualizing Community Inequalities" (VCI) (supported by the Leverhulme Trust), which aims to integrate an urban model of local community identities in their urban-network and geo-demographic contexts. The challenge is to integrate into the model urban communities' definitions of their local spaces, based around composite or multi-dimensional markers of these spaces. The investigators also found that a major challenge of integration relates to the variability in VGI participants' levels of engagement, confidence and skill, which we addressed through a structured workshop format (outlined below). Through this we have sought to identify what spatial structures are employed by urban communities to define their local spaces. Our approach is in contrast to a notion of local community spaces being formed around distinctive boundaries or binary definitions of what lies 'inside' and 'outside' the community space. Instead, we surmise that local markers of community spaces stem from 'relational artefacts' of physical and conceptual inter-dependencies, which underpin spaces for connectivity, separation or interaction. Such

local spaces form around community foci and street-network functionalities within spatial-structural and geo-demographic contexts. The overarching research challenge is to integrate urban structural, demographic and semantic layers into the urban model.

The research design of the VCI project has similarly addressed the need for data credibility by including in the methodology a stable and rigorous format for participatory point-data generation. Here point data has been gathered in terms of well-defined socio-economic and urban structural contexts. In this approach, we aim to develop a Web-enabled platform for credible, volunteered point-data generation. Yet the process of characterizing urban communities for integrated digital modelling has posed a challenge of complexity to the investigators. This led us to develop a 'pen-and-paper' prototyping methodology (based around participatory workshops). The aim of this was to generate diverse data pertaining to socio-spatial characteristics of the communities sampled at both the community (aggregated) and individual (disaggregated) levels. We subsequently selected only the most relevant and salient of the data for inclusion in the platform's data model. Here we report on the qualitative participatory methods and quantitative analytical methods for ensuring the quality of data capture. We present a case study of our research in Liverpool, (Merseyside, UK), which provides a compelling background for research into urban community formations in highly contrasting socio-economic contexts.

3. Deriving Urban Community Data for Planning

Urban communities involve subtle and irresolvable interplays of social meanings and spatial structures, forming their physical and symbolic boundaries (Logan, 2012). Community spaces also include effects from within the broader urban network (Hillier & Vaughan, 2007; Sampson, Morenoff, & Gannon-Rowley, 2002), positioned with multiple layers of spatial, social and effective properties (Grannis, 2009), thus posing a significant challenge to consistent data gathering.

Community spaces are special features of the urban environment, formed through the socio-spatial configurations by which people achieve 'nearness' at many levels of the home, street and public space. Community spaces comprise relational complexes of object and abstract artefacts (Hillier, 2007, pp.67-68), which we term physical and conceptual artefacts respectively. Homogeneity and separateness in the urban environment can enforce stereotypes of 'self' and 'other' (Sibley, 1995). These may also be reflected in socio-spatial structures and behaviours, including neighbourhood boundaries and place-bound identities or values (Mckenzie, 2015).

Community spaces that are conceptualized and defined as neighbourhoods afford the benefits of family life, social experiences and economic opportunities (Kearns & Parkinson, 2001). Yet neighbourhood identi-

ties are not spatially or temporally fixed. Community members' concepts of their neighbourhood vary according to, for example, their age, gender, level of ability, socio-economic standing or stage in life (Lupton, 2003).

The intermediaries of community spaces, their topo-geometric and topographic properties, provide the means by which urban actors both *think of* and *think with* their environments (Hillier, Turner, Yang, & Park, 2010). For example, the theoretical and methodological field of space syntax has shown how cities are arranged topo-geometrically into foregrounds of economic movement and backgrounds of controlled, residential zones. In these contexts the observer sees the 'other' city (whether background or foreground) relative to his or her position (Hillier & Vaughan, 2007). We *think of* these networks in terms of theoretical and professional discourse; we *think with* them in terms of quotidian actions based upon spatially embedded meanings.

Elsewhere the anthropologist Tim Ingold has similarly outlined an epistemological split between professional and 'inhabitant' spatial cognitions, stemming from distinctions of analytic and embodied knowledge respectively. Ingold has argued provocatively for a notion of 'way-faring' in spatial cognition, whereby the inhabitant develops his or her local and embodied knowledge through tangible encounters with flows of environmental information (Ingold, 2011, pp. 146-155). So too, various urban sociologies have described the embeddedness of local structures in community formations. For example Grannis (1998) has highlighted the importance of supposedly 'trivial' streets that interconnect blocks of tertiary streets, thus providing a path for novel community links to form. Elsewhere, Power and Houghton (2007) have identified spaces of community connectivity and urban innovation, which are not always visible to the official planner (Power & Houghton, 2007, pp. 158-159), and can lead to separation from social cores such as family homes (Power & Houghton, 2007, p. 55). In this way the authors observe how infrastructural projects have undermined or replaced 'community anchors', such as places for stopping and chatting or for children's play (Power, 2007, pp.58-59).

Considering the embedded, embodied or non-discursive nature of community formations around physical artefacts at a more theoretical level, Conroy and Bafna (2003) have drawn on seminal work of Lynch (1960) to re-define urban taxonomies in terms of their 'imageability'. In a similar vein, Palaiologou and Vaughan (2012) have offered a synthesis of urban structures that control movement among individuals and communities, namely those of *boundary, threshold and interface*, and their potential role to *divide, connect and allow interaction*. The VCI project sought to address these themes to case-study sites in areas of polarized socio-economic inequalities, with a view to revealing how communities in these contrasting areas variously employ local features to demarcate their local spaces.

4. Case Study

In our current work, we invited urban communities in Liverpool to volunteer information about their local spaces in terms of connectivity, separation and interaction. Liverpool is the UK's third city by regional population. While enjoying higher than-average economic growth in the period 2009–2014 (Liverpool City Council [LCC], 2016a), the city has among the highest levels of multiple deprivations of any UK local authority (LCC, 2015). The Liverpool region is 'a place of contrast and social and spatial disparities', bearing a range of spatial inequalities reflected in zonal concentrations of wealth and poverty (Sykes, Brown, Cocks, Shaw, & Couch, 2013, p. 6). It can be characterized historically as an area of prolonged industrial decline, reflected in overall above-average unemployment (5.7% at the time of writing), neighbourhood dereliction (cf. Hilditch, 2014), and low business density (LCC, 2016b). The Mersey estuary region is now attracting massive brownfield infrastructural investments, with groundwork currently under way. Within the Liverpool region, community inequalities within neighbourhoods at the peripheries of these developments present among the UK's most polarized spatial inequalities (findings of a study by Dorling et al., 2007). The region's transport authority has also advanced a progressive policy agenda, 'Connected Communities', which highlights the crucial importance of transport in revitalizing the urban region. A round of stakeholder consultations with urban practitioners and policy makers in Merseyside revealed the suitability of focusing on secondary schools in the city region, as they provide a stable environment and baseline for community-based research within a (planning) domain-sensitive context.

5. Area Characterization Methods

A preliminary characterization of all possible target schools in Merseyside area was carried out with the objective of creating a solid quantitative contextual background for the research. This preliminary step created a well-informed sample of schools and produced general contextual data, against which the much more detailed and contingent data gathered at the workshops could be compared and interpreted. The complete population of 79 state-sector secondary schools in Merseyside was characterized according to two broad sets of attributes: one covering demographic and socio-economic aspects, and another covering urban-structural characteristics.

Demographic and socio-economic attributes were quantified for the Lower Super Output Areas (LSOAs) where each school is located, using the 2015 Index of Multiple Deprivation (IMD). Super Output Areas are geo-located units devised by the UK's Office of National Statistics (ONS) to represent population aggregations by place of residence. In the present context, LSOAs

represent locations of around 1500 people in small residential areas. IMD provides a measure of relative deprivation, by which LSOAs are scored and ranked from least- to most-deprived. IMD is constructed by weighting indicators covering seven aspects of deprivation, namely: income, employment, health deprivation and disability, education skills and training, barriers to housing and services, crime and living environment. IMD and its separated sub-domains are provided as open data sets by the ONS.

Urban form and structure attributes were computed in a GIS from Ordnance Survey (OS) data, including layers representing the full road network hierarchy and the footprints of all buildings. Access points to public transport modes were also reckoned, using the UK Department for Transport's NaPTAN data set. The various urban-form and road-network structure indicators were computed through vector manipulation and quantified for the areas within circular buffers of 1 km radius, centred on each school's postcode centre point. These morphological indicators cover six aspects of urban form and structure, namely: geometry and topology of the street network (e.g. total length, number and type of junctions), geometry and topology of blocks (e.g. number of blocks, area/perimeter ratio of blocks), density and grain of buildings (e.g. number of buildings and total built-up area), network centrality measures (e.g. maximum betweenness centrality value) and density of public transport access points (e.g. number of bus and taxi stops).

Regarding the demographic and socio-economic characterization of Merseyside schools, we have made use of a set of geo-statistical clustering methods known as Local Indicators of Spatial Association (LISA), applied to IMD values encoded into the geography of LSOAs in the Liverpool Region. More than just quantifying the level of deprivation of each single LSOA containing a school, we were in fact interested in the *embeddedness* of such LSOAs within potentially larger geographical patterns of relative deprivation. Indeed, it is well known that the shape and size of the spatial units into which data is aggregated, has an unavoidable biasing effect (known as the 'modifiable areal unit problem') on the resulting aggregated value. Therefore, to merely collect the IMD values (or of its sub-domains) for each LSOA containing a school, would produce only coarse and not very reliable data.

LISA methods allow researchers to evaluate the variability of geographical attributes not only for each feature, but also for the features within a given neighbourhood around each one. In other words, LISA tell us the degree to which a feature has a particularly high or low value, according to the attribute itself *and* the location of the feature in question. Therefore, one may evaluate the degree to which a school located within a high-, medium- or low-deprivation LSOA is also located within a larger area of relative deprivation respectively. If, for in-

stance, high deprivation is verified across a reasonable range of distances (e.g. 0.5 km, 1 km, 1.5 km and so on) around a deprived LSOA containing a school, one may say with a very high level of certainty that such a school does indeed lie within an area of consistent and systematic deprivation. Using this method, we have made a characterization of the socio-economic embeddedness of every public-sector secondary school in Merseyside, targeting those lying in areas of consistent deprivation patterns, as described by IMD and its sub-domains.

The sampling of schools using as criteria their contrasting socio-economic characteristics also results in a strong variability regarding the morphological characteristics of the urban environments where those schools are located. To account for such variability becomes a fundamental step, in order to control for its potential impacts on the phenomena studied through the data gathered in the school workshops. The general physical and spatial morphological characterization of the schools' environments is therefore expected to provide complementary information to their socio-economic characterization, and also to serve as benchmark by which the information gathered in the workshops may be evaluated.

The morphological indicators mentioned above will be fed into a methodological data-flow model (which we present in Section 8, below) at a later stage of the research project. These indicators were chosen for their acknowledged support in distinguishing relevant aspects of urban form (such as building and network densities or network connectivity and geometry), which change significantly across urban areas and historical periods. Each morphological indicator results in a single figure for each area surrounding each school; together, they can be analysed through multivariate statistical methods in order to obtain summarized but consistent morphological descriptions and classifications. The use of network centrality measures allows us to generally characterize the urban movement potential of a given area, and therefore to make basic inferences about the probability of co-presence and social interaction on that area. Finally, the reckoning of the number and type of public transport access points, provides information about an area's accessibility degree, exposure to city-wide social networks and of the affordance of public space.

6. Participatory Research Methods

Once the sampled areas had been characterized, the investigators sought to understand how local populations made use of urban structures within socio-economic contexts to formulate the demarcations of their community spaces. In other fields, geographers have highlighted the significance of people's mental maps of urban landscapes (cf. White & Gould, 1986). There seems to be a task outstanding to identify the relational complexity of community spaces resulting, that

is, from the conceptual subsuming of physical artefacts into locally embedded or everyday processes (cf. O'Brien & Psarra, 2015).

In order to study this complex field and to ensure data credibility in a participatory VGI context, the investigators devised a workshop format whereby participants would work creatively within a structured framework. The workshops focused on themes of affordances for connectivity, separation and interaction in an area of 1 square mile surrounding the school, with the school located at the centre of the area sampled. Participation was invited from all state-sector secondary schools (excluding special schools) across the Metropolitan County of Merseyside, via an invitation submitted to teaching staff located across the county (Figure 1). Our intention was to include in the study schools representing the range of multiple deprivations in the city, while maintaining a broad balance in gender profile.

In total, 14 secondary schools participated, representing the range of areas in Liverpool by various indicators of deprivation (Figure 2). Approximately 360 participants engaged in the workshops, comprising 51% male and 49% female students. Around 230 participants were aged 13–14, around 80 aged 15–17, and a far smaller number aged 18. As most participants were children, it was not possible to gather data about individual circumstances relating to deprivation. However, an assumption was made that participants have general experience of typical social, structural and environmental characteristics of these areas.¹

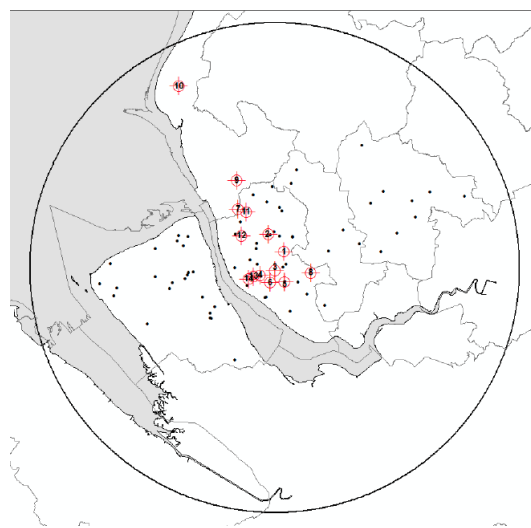


Figure 1. Locations of schools invited to participate in the workshops located within Merseyside. The schools selected for the sample (highlighted) were limited to contrasting areas within the city of Liverpool. The outer circle demarcates the geographic scope for statistical sampling (radius=25 km).

¹ Workshops were organized with the assistance of an external partner, Placed, a non-profit creative enterprise based in Liverpool (<http://www.placed.org.uk>).

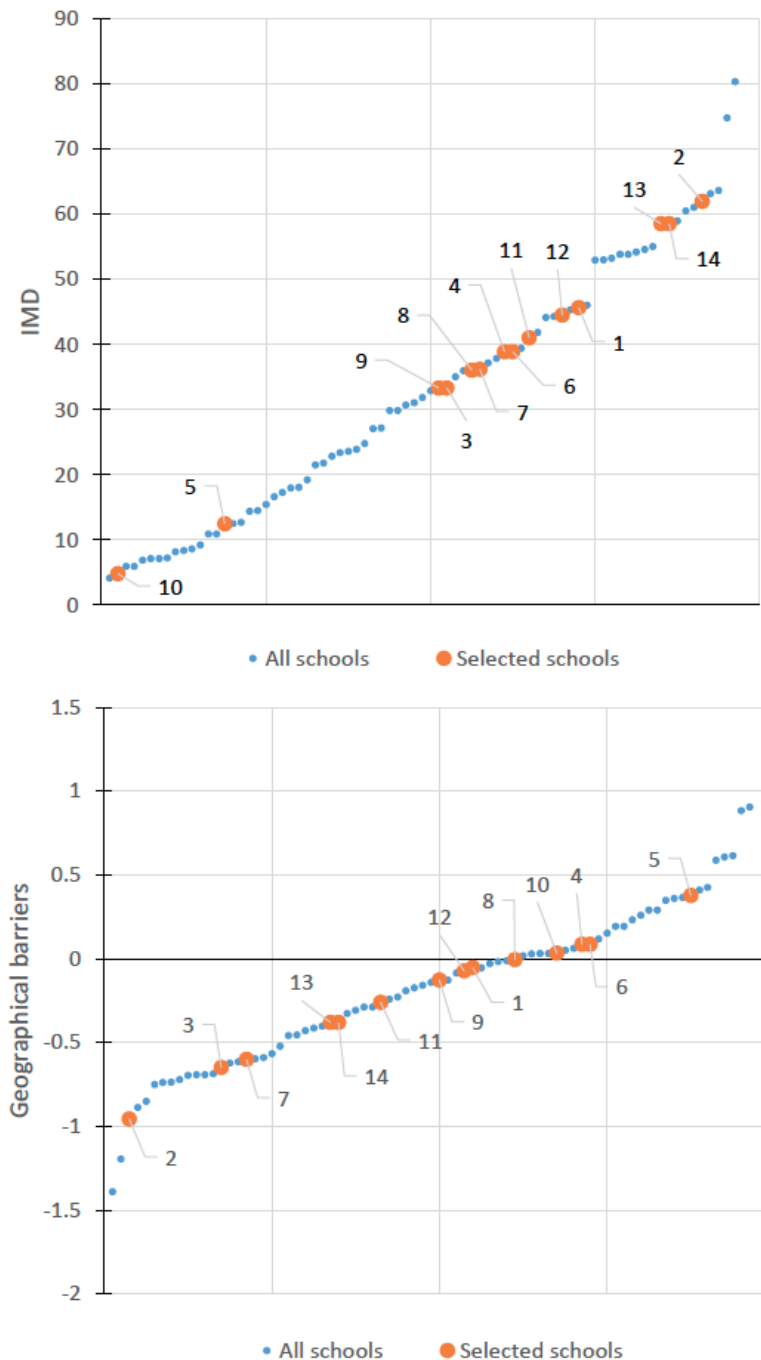


Figure 2. Two rank plots of the 14 Liverpool schools participating in the workshops (highlighted points) and of all state secondary schools in Merseyside (non-highlighted points). The y-axis denotes the values of the Index of Multiple Deprivation (IMD, top) and of the Geographical Barriers sub-domain (bottom), with high values meaning highly-deprived LSOAs; the x-axis denotes each LSOA rank value. IMD reveals a cluster of schools located within LSOAs of moderate IMD scores, with outliers in low- and high-deprivation areas. IMD sub-domains such as Geographic Barriers (bottom) reveal a more even distribution of sampled schools, although deprivation levels generally correlate to accessibility levels.

The investigators followed a methodology of community participation that seeks to equip the participants to analyse and describe their own ‘realities’ (Chambers, 1997). The aim was to bring about a ‘transformative’ participatory experience, whereby unfamiliar concepts are introduced and barriers to understanding are surmounted through close dialogues with facilitators (cf.

Meyer & Land, 2006). Hence, participants received basic ‘class-room’ training in the impact of the urban fabric for community connectivity, separation and interaction. Participants were arranged into age-based groups of 3–4 people, including groups of 12–13, 14–15, 16–17 and 18–plus year-olds. The session used image-based case studies of urban fabrics, and involved

participants in simple evaluation exercises. Following this, the participants undertook an evaluation of their local school area using A1 maps, colour-coded sticky dots to demarcate significant places and string to demarcate paths.

Participants then made use of their individual packs of workshop materials. An important element for capture of metadata was a questionnaire survey (extracted from a format devised by Scholes, Boniface, Stockton and Mindell (2016). This was devised to capture:

- Basic personal details;
- Level of social engagement;
- Sense of difficulty getting to/from school;
- Sense of the local area (weighted scale: positive-to-negative);
- Knowledge of local proximity to amenities and services;
- Range of districts visited in previous 12 months;
- Range of travel modalities employed.

The investigators also included in each pack a person character card, selected from a set that was devised by the researcher to reflect some typical experiences of people living in deprived areas (e.g. relating to health, employment, personal relationships, hazards and disorders). The purpose of this was twofold. Firstly, we wished to avoid the possibility of stigmatization (i.e. facilitators were restricted by the school Head Teachers from introducing themes of ‘deprivation’ in describing

the local area during the workshop). Secondly, we encouraged the participants to reflect on the broader needs of the community. For example, character scenarios invited the participants to think about local structures from another’s perspective, which also fits a ‘dialogic’ model of relational complexity (O’Brien & Psarra, 2015)². Participants filled out a brief questionnaire about that character’s needs in the local area, from which we may draw a set of broader community requirements.

Participants worked individually with A3 local maps representing 1 square mile (1.7 km²) surrounding the school (Figure 3). This sample represented typical distances between local amenities and services proximal to the school, such as shops and transport links. These distances are set within a radial scale relating to walkability, or to ‘velo-mobility’, that is typical of participants’ everyday mobilities. Participants were invited to apply emoticon stickers to any number of local structures that they deemed significant for affordances of connectivity, separation or interaction among people in the local area. Participants could select from a constrained set of up to 14 emoticons. Sets of emoticons comprised representations of primary emotions by a standard psychological schema, basic hand gestures and abstract signs for ‘hazard’ or ‘barrier’. Each participant was given one sheet of 5x sets of 14 emoticons. Participants made their own, individual interpretations of emoticons’ meanings.

² The psychologist James Wertsch (1993) has outlined how children use such ‘dialogic’ imagination to construct their sense of reality by thinking with the experiences of another.

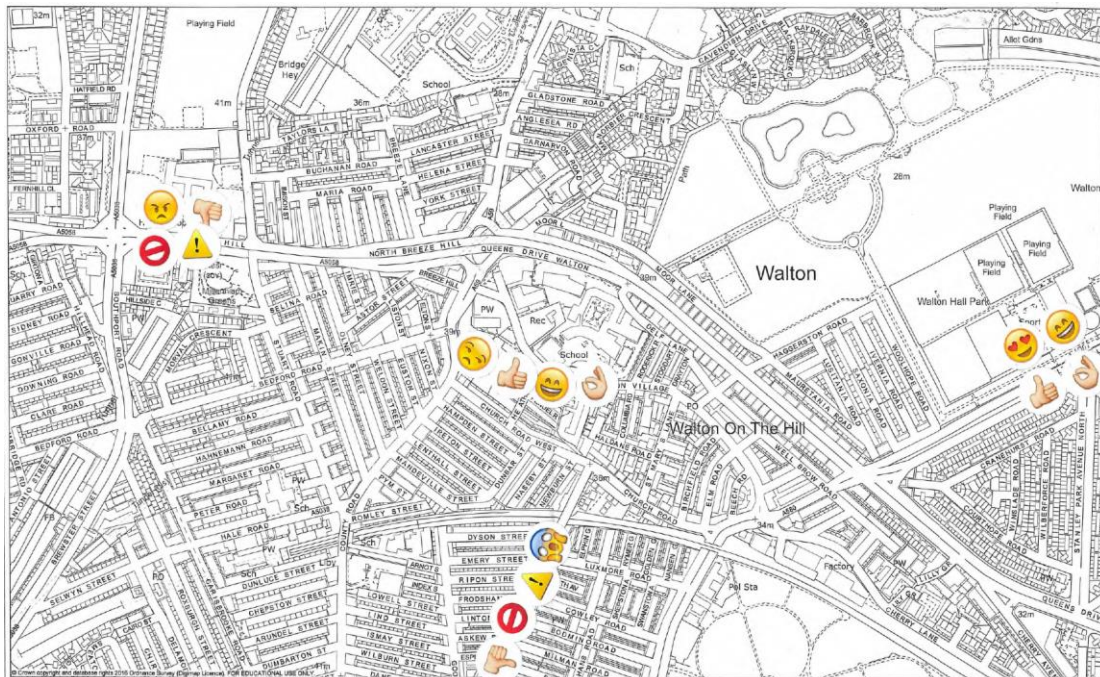


Figure 3. Example A3 map of local area (not to scale). Here a 13 year-old, female participant has selected four features that she regards as being significant for local community formations (including the school she attends, in the centre for the map). She has used several emoticon stickers to represent her range of thoughts, feelings or experiences, and so on, about these.

What YOU think

1 Write in here the local feature you selected

2 Stick down here the emojis you have used on your map

FEATURE 1 : Belvedere Academy	
😊	love this place. Good education and friends.
👉	It's ok. Some of the people here can be scary though.
😬	A few accidents happen around here involving the busy roads.
😬	The bus route takes a while to get to school because of traffic but we don't get in trouble for it.

3 Write in these boxes what each emoji means to YOU

FEATURE 2 : Belvedere roads	
⚠️	Can be dangerous.
😱	Oh My God! It's The amount of times I have almost been hit with a car is scary. It's dark at parent's evening and loads of cars are parked here.
🚫	Try to avoid when possible - Very scary.
🚫	Stop! The cars will hardly ever stop and some people park their cars so we can't see around them.

Figure 4. Example of completed emoticon table, with brief descriptions of each emoticon (not to scale). Here the participant has selected her school as being the local 'feature' most important for connectivity, and the local main road as that most important for separation. The participant has listed four emoticons for each feature, and written a text description of each emotion.

Participants were then invited to select two 'features' in the local area: one most important for connectivity and one most important for separation among local people. Following this, they were invited to fill out a table to list and describe the emoticons they used on their maps (see Figure 4). This involved listing the two most significant features for connectivity and separation, and sticking examples of four emoticons they had used down a left-hand column. Participants then filled in each table field with a brief description of what each emoticon meant to them with reference to the selected local feature (for example, 'fear' emoticon relating to a local park might be described as 'scary after dark').

The table was refined after four completed workshops to improve the quality of texts gathered. This was achieved by supplying the word 'because' in each field, compelling the participants to construct complete sentences around a causal factor (for example 'I was

scared at the park gates *because* a gang was there'). These refined tables also included a list of example sentences, which helped improve the quality of texts gathered, while resulting on only occasional 'parrotting' of the supplied examples.

7. Assessing Quality of Context for Data Production

The investigators' intention was to bring all participants, no matter their profile and background, to a broadly similar level of engagement in the data-gathering activity. This was measured through entry and exit surveys, which invited all participants to report on the impact of the workshop on their awareness and knowledge of urban design and communities, as well as their level of confidence and 'domain' knowledge in these areas (323 entry and 275 exit surveys were collected from 340 participants). Entry surveys demon-

strated stark differences in the confidence levels and aspirations of students; only 21% overall reported any prior experience in working with built environment themes. Yet after training and practical work the majority of participants reported improved knowledge, awareness and confidence overall in addressing matters of formation and intervention in the built environment.

Comparing entry and exit surveys among participants revealed the overall positive impact on skills and knowledge resulting from engagement in the workshop (albeit our counting was biased by the lower return rate of exit surveys). On entry 32% of participants agreed and agreed strongly that they were aware of the impact of urban design on people, increasing to 91% upon completion of the workshop. Similarly, 68% of participants reported being aware of their local urban community environments at entry, rising to 94% upon exit. Confidence in talking about architecture and the built environment rose from 17% to 79% upon completion of the workshop. The exit surveys revealed a generally improved levelling of participants' knowledge of the built environment, perhaps reflecting the criticality of the community training exercises as part of the workshop format. This general levelling helped us to ensure that the data produced through the workshops was from a stabilized basis of skills, knowledge and awareness of our study's key areas of focus.

8. Methodological Data Model

In generating a research design for the VCI project, the investigators surmised that all urban communities comprise three informational layers: urban-structural, geo-demographic and 'semantic'. Hence, the urban-structural layer relates to street-network functionality in terms of movements; the geo-demographic layer relates (in the present context) to distributive patterns of socio-economic deprivations; the 'semantic' layer relates to locally embedded and inter-dependent definitions of community spaces. Modelling the semantic layer has posed an outstanding challenge for research and we have described our approach to this through data gathering via a series of participatory workshops. These have comprised a wide-ranging prototyping phase, from which we have gathered high-dimensional data. The complexity of the model being tested benefits from a diagrammatic model (see Figure 5).

The project data model reveals data gathering from urban spatial sources (structural and distributive) and local semantic sources. The local semantic data pertains to community definitions of local spaces, categorised as points (the location of emoticons) and weights (the type of emoticons applied). Points data may be processed algorithmically to reveal clustering around significant local features (such as community foci and street-network functionalities), and these may be tak-

en as 'markers' of local structures for community life. Both weight and points data may be applied to take sections of the urban data; hence, horizontal sections may be taken at the city scale from categories of emoticons (for example, the section of data relating to all 'angry' emoticons), or vertical sections may be taken at the local scale relating to significant local structures. Once data sections have been taken, the investigators can then extract text descriptions of these markers. From these we may generate taxonomic categories of urban community localities. For example, where a park entrance has been widely demarcated with 'hazard' and 'angry' symbols, and described in terms of 'gang' activities, so we may taxonomize this as some kind of *barrier* to accessibility and community life.

9. Discussion

The phase of work reported in this paper forms one part of a broader project, which intends to develop a participatory, Web-enabled platform to support planning-domain processes. We adopted a 'pen-and-paper' prototype method for point-data capture, which allowed us to iterate rapidly a system design, to improve participants' levels of engagement, and to stabilize the data-generation environment. This prototype method would be readily transferable to any digital platform featuring, for example, Web-mapping capabilities and meta-data capture. The outstanding task for the investigators is to devise a 'linked' data model, which allows for the re-combination of point- and meta-data so as to produce novel, localized definitions of community spaces.

Fortunately some current developments in open-source platform technologies support this kind of participatory map-making. For example, GeoKey is an open source platform that provides server-side, customizable geographic data components, allowing modellers to build a framework for participatory map-making and to manage volunteered data³. Elsewhere, various libraries within the R programming environment support overlays of geo-spatial and graph-theoretical network analyses, which are publishable to the Web (such as, for example, Spatial R, iGraph and Shiny). Used in combination, these allow the modeller to observe or simulate urban communities' dynamic inter-relationships.

Urban community formation is a multiple-layered, super-positional process, involving urban structural, geo-demographic and semantic components. Urban communities also make use of a range of artefacts in their built environments to define their local spaces. The 'relationality' of these artefacts involves their incorporation into locally embedded frameworks, whereby

³ Developed by the Extreme Citizen Science Research Group at University College London (<http://geokey.org.uk>).

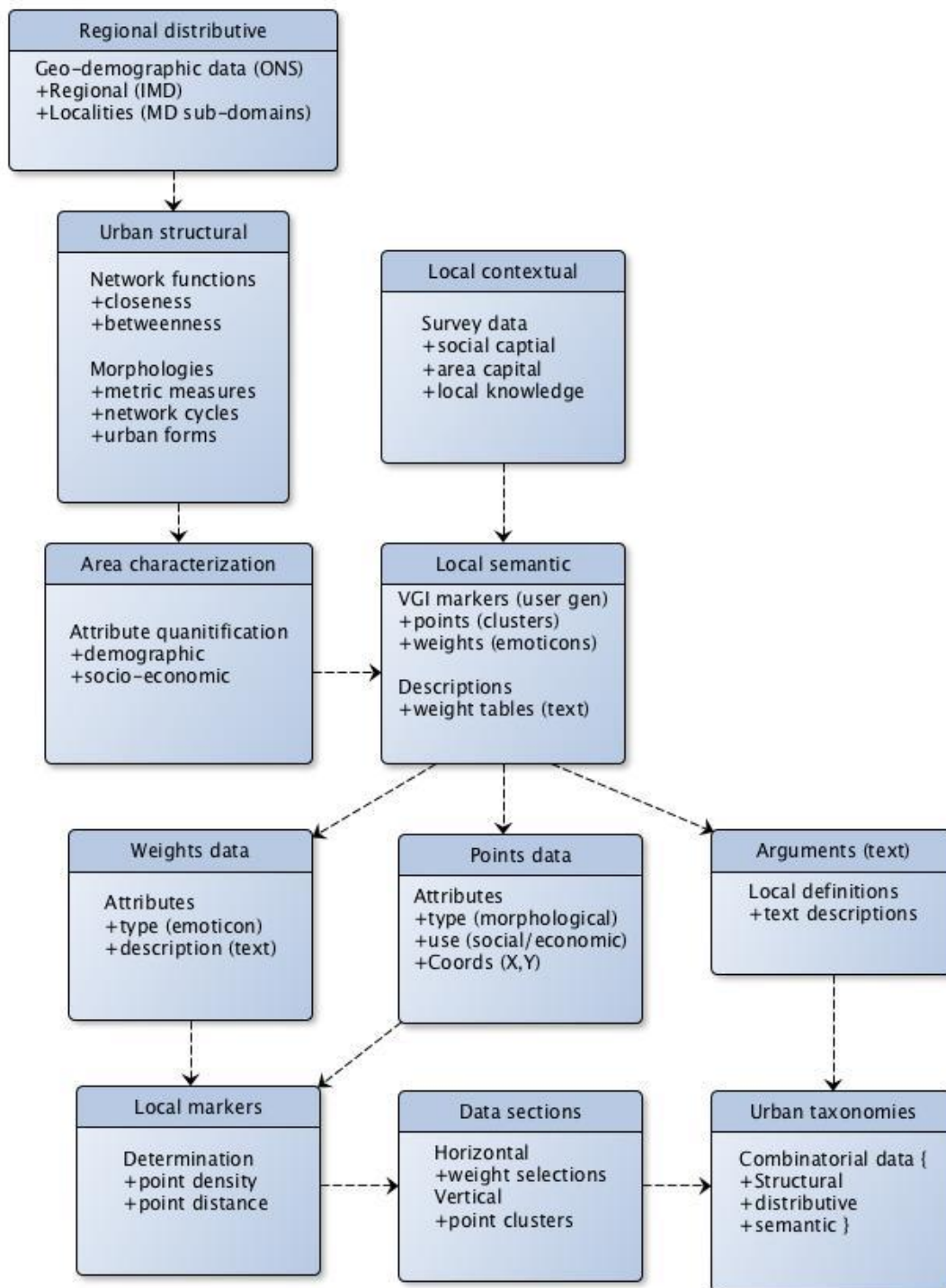


Figure 5. Methodological data model ('VCI project'). The model shows how a range of data flows into the study's urban community analysis, forming the basis for gathering voluntary point data and text data. It shows how 'local semantic' data is gathered from the participatory workshops, including the geo-locations (points) and characteristics (weights) of the emoticon stickers, the urban artefacts they represent, and their text descriptions. Finally, the processed points data will be used to taxonomize the range of artefacts selected in terms of comparative demographic and structural characteristics. In this way, the VGI data is captured traceably within a well-defined, multi-layered framework.

separate communities (or sub-groups within a community) may use the same physical artefacts yet attribute to them different sets of weights. Hence, one group's space for connectivity may be another's for separation.

Understanding how 'relational artefacts' are embedded in localised, everyday processes has posed significant challenges to the investigators. Firstly, we face the challenge of complexity in the multiple layers of urban

community formations. Secondly, we face the problem of ensuring quality of VGI-derived data. In meeting these challenges, we have employed two contrasting modes of data capture, one based on quantitative structural and demographic characteristics, and one on qualitative or participatory generation of VGI data.

The workshop format we adopted attended to the need for credibility in data capture stemming from 'credentialization' of data producers. Hence, the workshop included a training component that helped to bring all participants to a broadly even level of knowledge, skill and capability in working with urban affordances and possible interventions, which we measured on broad terms via entry and exit surveys. The population sample comprised a range of secondary state-sector schools, providing us with a stable and well-defined baseline for study, as well as participant segmentation based on gender and age, and community profiling based on transport and mobilities categorizations. Further analysis may recombine the sample sub-populations more subtly, based on their personal and community profiles in combination with their selective 'weights' (such as, for example, the group of pedestrians in high-deprivation areas who experience overall negative feelings towards their local environments). Where the capture of VGI data has presented a problem of quality and consistency in geo-spatial analysis elsewhere, we have found that a voluntary approach framed within a well-defined and 'transformative' participatory context serves to maintain quality. Moreover, the participatory context has also benefitted from systematic analysis of area characteristics, meaning that highly diverse weightings and descriptions of local community formations are traceable to their specific socio-economic and structural contexts.

10. Conclusion

The investigators set out to ensure credibility in VGI-based data generation in urban planning through an iterative research design. Our intention was to achieve credibility by framing the enquiry. We focused on accuracy through a systematic urban spatial analysis, which served to formulate the structural contexts of our sample populations, including morphological affordances for community life in often challenging urban environments. We focused on believability by stabilizing the participatory data-production environment through comparative evaluations of participants' levels of engagement, which served to formulate their experiential contexts based on these generative affordances. This dual approach to VGI data credibility has allowed us to address the core concern for the inclusion of participants' experiences of socio-spatial inequalities in urban-planning processes.

We propose that the various datasets generated for the study may flow into an integrated model, which

will underpin a connective (Web-enabled) platform intended for participatory urban planning. The employment of a participatory selection and weighting method means that we can show how community members are connected to their local spaces through various kinds of localised mediations. Hence, we intend to develop a layered graph data model of weighted and defined structures, which will comprise our next phase of work. Finally in the current project, the investigators intend to validate the study's multiple-layered model of urban community formations in a formal urban-planning domain.

Acknowledgment

The authors are very grateful to Jo Harrop, Director of Placed CIC, for providing additional information about workshop outcomes.

Conflict of Interests

The authors declare no conflict of interests.

References

- Cassiers, T., & Kesteloot, C. (2012). Socio-spatial inequalities and social cohesion in European cities. *Urban Studies*, 49(9), 1909-1924.
- Chambers, R. (1995). Poverty and livelihoods: Whose reality counts? *Environment and Urbanization*, 7(1), 173-204.
- Chambers, R. (1997). *Whose reality counts? Putting the first last* (2nd ed.). London: ITDG Publishing.
- Chokor, B. (1991). The perception of spatial inequalities in a traditional third world city. *Urban Studies*, 28(2), 233-253.
- Coleman, D., Georgiadou, Y., & Labonte, J. (2009). Volunteered geographic information: The nature and motivation of producers. *International Journal of Spatial Data Infrastructures Research*, 4, 332-358.
- Conroy-Dalton, R., & Bafna, S. (2003). The syntactical image of the city: A reciprocal definition of spatial elements and spatial syntaxes. In *Fourth International Space Syntax Symposium*. London: University College London.
- Dorling, D. (2012). *The visualisation of social structure*. Chichester: John Wiley & Sons.
- Dorling, D., Rigby, J., Wheeler, B., Ballas, D., Thomas, B., Fahmy, E., Gordon, D., & Lupton, R. (2007). *Poverty, wealth and place in Britain, 1968 to 2005*. Bristol: Joseph Rowntree Foundation/The Policy Press.
- Flanagin, A., & Metzger, M. (2008). The credibility of volunteered geographic information. *GeoJournal*, 72, 137-148.
- Goodchild, M. (1995). Sharing imperfect data. In H. J. Onsrud & G. Rushton (Eds.), *Sharing geographic information*. New Brunswick, NJ: Centre for Urban Policy Research.

- Goodchild, M. (2007). Citizens as sensors: The world of volunteered geography. *GeoJournal*, 69(4), 211-221.
- Gould, P., & White, R. (1986). *Mental maps* (2nd ed.). London: Routledge.
- Grannis, R. (1998). The importance of trivial streets: Residential streets and residential segregation. *American Journal of Sociology*, 103(6), 1530-1564.
- Grannis, R. (2009). *From the ground up: Translating geography into community through neighbor networks*. Princeton, NY: Princeton University Press.
- Grant, U. (2010). *Spatial inequality and urban poverty traps* (ODI Working Paper 326). London: Overseas Development Institute.
- Haklay, M., Antoniou, V., Basiouka, S., Soden, R., & Mooney, P. (2014). *Crowdsourced geographic information use in government*. London: Global Facility for Disaster Reduction & Recovery (GFDRR), World Bank.
- Hilditch, M. (2014, June 6). Long-term housing mission for Liverpool. *Inside Housing*. Retrieved from <http://www.insidehousing.co.uk/long-term-housing-mission-for-liverpool/7004041.article>
- Hillier, B. (2007). *Space is the machine*. London: Space Syntax. Retrieved from <http://discovery.ucl.ac.uk/3881>
- Hillier, B., Turner, A., Yang, T., & Park, H.-T. (2010). Metric and topo-geometric properties of urban street networks: Some convergences, divergences and new results. *The Journal of Space Syntax*, 1(2), 258-279.
- Hillier, B., & Vaughan, L. (2007). The city as one thing. *Progress in Planning*, 67(3), 205-230.
- Ingold, T. (2011). Against space: Place, movement and knowledge. In T. Ingold (Ed.), *Being alive: Essays on movement, knowledge and description*. Abingdon: Routledge.
- Kearns, A., & Parkinson, M. (2001). The significance of neighbourhood. *Urban Studies*, 38(12), 2103-2110.
- Liverpool City Council. (2015). *The Index of Multiple Deprivation 2015: A Liverpool analysis*. Liverpool: Liverpool City Council.
- Liverpool City Council. (2016a). *The city of Liverpool key statistics bulletin* (Issue 24, June 2016 Update). Liverpool: Liverpool City Council.
- Liverpool City Council. (2016b). *Liverpool Economic Briefing 2016: A monitor of jobs, business and economic growth*. Liverpool: Liverpool City Council.
- Logan, J. (2012). Making a place for space: Spatial thinking in social science. *Annual Review of Sociology*, 38, 507-524.
- Lupton, R. (2003). "Neighbourhood effects": Can we measure them and does it matter? (CASE paper no. 73). London: Centre for Analysis of Social Exclusion, London School of Economics and Political Science.
- Lush, V., Bastin, L., & Lumsden, J. (2012). Geospatial data quality indicators. In *Proceedings of the 10 International Symposium on Spatial Accuracy Assessment in Natural Resources and Environmental Sciences*, Florianopolis-SC, Brazil.
- Lynch, K. (1960). *The image of the city*. Cambridge, MA: MIT Press.
- Marcuse, P. (2002). The partitioned city in history. In P. Marcuse & R. van Kempen (Eds.), *Of states and cities: The partitioning of urban space*. Oxford: Oxford University Press.
- Mckenzie, L. (2015). *Getting by: Estates, class and culture in austerity Britain*. Bristol: Policy Press.
- Meyer, J., & Land, R. (2006). Threshold concepts: An Introduction. In J. Meyer & R. Land (Eds.), *Overcoming barriers to student understanding: Threshold concepts and troublesome knowledge*. Abingdon: Routledge.
- Morsey, H. (2012). Scarred generation. *Finance & Development*, 49(1), 15-17.
- O'Brien, J., & Psarra, S. (2015). The dialogic city: Towards a synthesis of physical and conceptual artefacts in urban community configurations. In *Proceedings of the Tenth International Space Syntax Symposium*, London, UK.
- Palaiologou, F., & Vaughan, L. (2012). Urban rhythms: Historic housing evolution and socio-spatial boundaries. In *Proceedings of the Eighth International Space Syntax Symposium*, Santiago, Chile
- Power, A. (2007). *City survivors: Bringing up children in disadvantaged neighbourhoods*. Bristol: Policy Press.
- Power, A., & Houghton J. (2007). *Jigsaw cities: Big places, small spaces*. Bristol: Policy Press.
- Sampson, R., Morenoff, J., & Gannon-Rowley, T. (2002). Assessing "neighborhood effects": Social processes and new directions in research. *Annual Review of Sociology*, 28, 443-478.
- Scholes, S., Boniface, S., Stockton, J., & Mindell, J. (2016). *Developing a questionnaire to assess community severance, walkability, and wellbeing: Results from the Street Mobility Project in London* (Street Mobility and Network Accessibility Series Working Paper 05). London: University College London.
- Sibley, D. (1995). *Geographies of exclusion: Society and difference in the west*. Abingdon: Routledge.
- Sykes, O., Brown, J., Cocks, M., Shaw, D., & Couch, S. (2013). A city profile of Liverpool. *Cities*, 35, 299-318.
- Tulloch, D. (2007). Many, many maps: Empowerment and online participatory mapping. *First Monday*. Retrieved from <http://www.firstmonday.org/ojs/index.php/fm/article/view/1620/1535>
- UN-Habitat. (2009). *Planning sustainable cities: Global report on human settlements 2009. Key findings and messages*. London: Earthscan.
- UN-Habitat. (2012). *State of the World's cities 2012/2013: Prosperity of cities*. New York, NY: Routledge.
- United Nations. (2013). *Inequality matters: Report on the world social situation 2013*. New York, NY: United Nations. Retrieved from <http://www.un.org/esa/socdev/documents/reports/InequalityMatters.pdf>
- Veiga, D. (2012). *Urban inequalities and segregation in Montevideo*. Paper presented at the Urban and Re-

gional Research Committee, RC 21 International Sociological Association Forum of Sociology ISA. Buenos Aires, Argentina.

Wertsch, J. (1993). *Voices of the mind: A sociocultural approach to mediated action*. Cambridge, MA: Harvard University Press.

About the Authors



Jamie O'Brien is Senior Research Associate at the Bartlett Centre for Advanced Spatial Analysis, University College London (UCL), where he is researcher on the project, "Visualizing Community Inequalities" (supported by the Leverhulme Trust). His research interests relate to visualization and integrative modelling of everyday domains, with applications in decision-making and cognitive processing. Jamie was EPSRC Research Engineer (EngD programme), Enterprise Scholar (UCL/London Business School), and is author of the book *Shaping Knowledge* (Elsevier-Chandos, 2014).



Miguel Serra holds an MSc in Planning and Design of the Built Environment and a PhD in Civil Engineering: Territorial and Environmental Planning, both from the Faculty of Engineering, University of Porto. He is a member of CITTA—Research Centre for Territory, Transports and Environment, University of Porto—and an Honorary Research Associate at the Bartlett School of Architecture, University College London. His current research focuses on regional and supra-regional spatial networks, geographic information systems and data analysis and visualization.



Andrew Hudson-Smith is a Professor of Digital Urban Systems and Director of The Bartlett Centre for Advanced Spatial Analysis (CASA) at University College London. He is Editor-in-Chief of *Urban Planning* and of *Future Internet*, and has a research focus on location-based digital technologies. Andy is course founder of the MRes in Advanced Spatial Analysis and Visualisation, Course Director of the MSc in Smart Cities and Urban Analytics and the MRes in Smart Cities, all at CASA.



Sophia Psarra is Reader of Architecture and Spatial Design at the Bartlett (UCL) and associate editor of the *Journal of Space Syntax*. She uses computer modeling to analyze spatial layouts in relation to social, cultural, cognitive and organizational performance. This analysis is combined with empirical data of users' activity to provide an account of how organic cities emerge through self-organizing processes. Sophia is author of the book *Architecture and Narrative: The Formation of Space and Cultural Meaning* (Routledge, 2009).



Anthony Hunter is Professor of Artificial Intelligence, and head of the Intelligent Systems Research Group, in the UCL Department of Computer Science. His research is primarily in the area of computational models of argument. This is concerned with developing formal models of the human cognitive ability to deal with incomplete and inconsistent information about the world through the use of argument, and to use these models in tools for decision-making and sense-making.



Martin Zaltz Austwick is Senior Lecturer in Advanced Spatial Analysis and Visualization. He holds degrees in physics, nanotechnology and quantum computing, and his public engagement work has led to an EPSRC-funded Public Engagement Fellowship, as well as Radio Academy Gold and Silver Awards. Martin's research relates to visualization and analysis of human data and spatial movement, including Bike Share Schemes, GPS tracks drawn from pedestrian movement in London, and shipping movements around the globe.