

<b>Institution:</b> University College London		
<b>Unit of Assessment:</b> 14 – Geography and Environmental Studies		
<b>Title of case study:</b> Innovative acquisition, analysis, and visualisation of geodemographic data to inform policy, practice and planning decisions for government, economy and society		
<b>Period when the underpinning research was undertaken:</b> 2010-2019		
<b>Details of staff conducting the underpinning research from the submitting unit:</b>		
<b>Name(s):</b>	<b>Role(s) (e.g. job title):</b>	<b>Period(s) employed by submitting HEI:</b>
Paul Longley James Cheshire	Professor of Geographic Information Science Professor of Geographic Information and Cartography (2019-present) Associate Professor of Quantitative Human Geography (2016-2019) Lecturer in Advanced Spatial Analysis and Visualization (2011-2016)	2000-present 2011-present
Guy Lansley Oliver O'Brien Jens Kandt	Senior Research Fellow Senior Research Associate Research Associate (2013-2017) Lecturer in Urban Geography and Data Science (2017-2019) Associate Professor in Urban Geography and Data Science (2019-present)	2012-2019 2008-present 2013-present
<b>Period when the claimed impact occurred:</b> 2013-2020		
<b>Is this case study continued from a case study submitted in 2014?</b> N		
<b>1. Summary of the impact</b> (indicative maximum 100 words)		
<p>Research by the UCL Geospatial Analytics and Computing (GSAC) group advancing the acquisition, analysis, and interpretation of consumer data has unlocked the power of consumer 'Big Data', providing critical and rigorous evidence to realise and enable the effective use of data. This research equips those working in the public, private, education, and research sectors with data products, classification approaches, and visualisation tools that have tackled questions relating to social inequalities and have informed policy, practice and planning decisions. The range of benefits from this research have included an informed understanding of the ethnic-variant impacts of COVID-19 for health agencies, improved urban transport planning, widened participation in sports scheme investments, and enhanced use of mapping in secondary school education.</p>		
<b>2. Underpinning research</b> (indicative maximum 500 words)		
<p>What is now understood as Geographic Information Systems (GIS) research was established in UCL Geography in the 1970s. Building upon this formative legacy, the Geospatial Analysis and Computing (GSAC) group has, since 2014, advanced the core principles and concepts of today's geospatial data science, and its innovative application in social science. Led by Professor Paul Longley, the group is a global leader in the development of geospatial techniques and tools for analysing the geographies and dynamics of human behaviour using large and complex datasets derived from both conventional survey and novel consumer data sources. The latter increasingly arise as digital traces of the interactions between humans and the technologies that record the activities and transactions of our everyday lives. Although a huge potential resource, poor understanding of issues of data provenance, access, control, and ethics limit the use of consumer data in research that can improve understanding the nature and functioning of society. In February 2014, Longley secured ESRC funding [ii] to establish the UCL division of the Consumer Data Research Centre (CDRC), prompting an upscaling of GSAC's research capacity and broadening of its scope. CDRC's core research mission is to develop new or novel approaches to the acquisition, custodianship, mining, and analysis of consumer data. This in turn has driven delivery of open access visualisations of geospatial demographic data. Members of GSAC have developed new approaches to acquire, create, validate, link, and maintain new nationwide consumer datasets. These have been used to deliver cutting-edge, highly granular, timely, and policy-</p>		

relevant descriptions of what is going on in society today. These accessible representations of the world are underpinned by advanced techniques, online infrastructure and analysis-ready consumer data products.

Chronologically, GSAC's contribution includes: i) **acquisition or harvesting** data in partnership with business, government and the third sector; ii) **geodemographic analysis** through data validation and linkage; and iii) **geovisualisation** through new techniques and services.

i) GSAC has developed data licencing agreements with more than 30 external partners along with application programming interfaces to acquire and distribute geospatial data. GSAC research has extended the capability and rigour of data mining from global and national networks and databases, ranging from bikeshare schemes to national consumer registers. Technical developments realised in O'Brien's work [R1] include synthesising data in the form of functional databases, modifying classification methods for improved geovisualisation, and expanding the capability of real-time processing to reveal spatial and temporal understanding of human behaviour (e.g. bike use). Kandt [R2] has established ethically safeguarded methods sensitive to subjective interpretation to relate names to ethnicities, using the secure research facilities of the Office for National Statistics (ONS) Secure Research Service (SRS).

ii) Analysis-ready 'geodemographic' classifications of neighbourhoods [R3/R4] continue more than two decades of substantive and methodological innovation by Longley, pioneering the sourcing, validation and linkage of consumer data sources. GSAC research has recognised both the shortcomings of conventional demographic approaches, and the potential of consumer data in improving on them. GSAC has developed and implemented a programme of internationally leading technical innovations [R5] to link diverse data sources with different data structures across space and time. Advancing the analytical methods that permit connections between open, consumer and administrative data, GSAC has defined operational frameworks for the robust delivery and analysis of population- and people-centred data [R4/R5] leading to improved understanding of changing individual activity patterns, household structures, and neighbourhood attributes.

iii) Until recently, digital map design and production remained the preserve of the GIS specialist. GSAC have developed novel cartographic and web-mapping tools and protocols to integrate and visualise demographic and activity data, making novel use of open and accessible base-maps, at national scales [R6]. Research has focused on linking social structure to the built environment and developing novel methods to represent traffic flows, using consumer data alongside national statistics. These are visually intuitive, and they readily-improve communication and enable local area rescaling to link the local to the big picture. These principles are embodied in the DataShine web-based mapping platform, created by O'Brien and Cheshire [R6], funded by ESRC [i]. Launched in 2014, DataShine was initially released visualising specific variables from the England and Wales 2011 Census. Following feedback from users, and continued research, additional functionality has been integrated into the toolkit to deliver a wider range of scales and mapping of other national datasets (e.g. DataShine Scotland, which maps census data for Scotland).

### 3. References to the research (indicative maximum of six references)

- R1. O'Brien, O., Cheshire, J., Batty, M.** (2014). Mining bicycle sharing data for generating insights into sustainable transport systems. *Journal of Transport Geography*, 34, 262-273. [doi:10.1016/j.jtrangeo.2013.06.007](https://doi.org/10.1016/j.jtrangeo.2013.06.007)
- R2. Kandt, J., Longley P.** (2018). Ethnicity estimation using family naming practices. *PLoS ONE (Public Library of Science)* 12. [doi:10.1371/journal.pone.0201774](https://doi.org/10.1371/journal.pone.0201774)
- R3. Singleton, A., Longley, P.** (2015). The internal structure of Greater London: a comparison of national and regional geodemographic models. *Geo: Geography and Environment*, 2:1. [doi:10.1002/geo2.7](https://doi.org/10.1002/geo2.7)
- R4. Singleton, A., Longley, P.** (2019). Data infrastructure requirements for new geodemographic classifications: The example of London's workplace zones. *Applied Geography*, 109. [doi:10.1016/j.apgeog.2019.102038](https://doi.org/10.1016/j.apgeog.2019.102038)

- R5. Lansley, G., Li, W., Longley, P.** (2019). Creating a linked consumer register for granular demographic analysis. *Journal of the Royal Statistical Society: A*. [doi:10.1111/rssa.12476](https://doi.org/10.1111/rssa.12476)
- R6. O'Brien, O., Cheshire, J.** (2016). Interactive mapping for large, open demographic data sets using familiar geographical features. *Journal of Maps*, 12:4. [doi:10.1080/17445647.2015.1060183](https://doi.org/10.1080/17445647.2015.1060183)

All outputs were peer reviewed.

The quality and scale of the research is also evidenced by its funding:

- i. 2013-2016: Economic and Social Research Council awarded [PI Cheshire] GBP243,145 to develop the Datashine geovisualisation website (Big, Open Data: Mining and Synthesis (BODMAS): ESRC Future Research Leaders Award, ES/K009176/1).
- ii. 2014-2022: Economic and Social Research Council awarded [PI Longley] GBP7,635,199 (plus two sets of continuation funding since February 2019 totalling GBP2,012,143) to develop the Consumer Data Research Centre (Retail Business Datasafe: ES/L011840/1).

#### 4. Details of the impact (indicative maximum 750 words)

GSAC research has improved understanding of the nature of spatial data, and has produced accessible, timely, and policy-relevant information, services and tools to support government, economy and society. It has benefitted stakeholders across public and private sectors both globally and in the UK (local and national government). GSAC's scientifically robust and effective reuse of consumer data and development of novel geovisualisation tools and classifications have supported local government decision-making, informed understanding of COVID-19 in health agencies, enabled effective operation of bikeshare systems, widened participation in sports scheme investments, and improved the use of mapping in secondary school education.

##### Supporting Local Government Decision-Making

GSAC's research developing neighbourhood classifications of social and physical conditions and characteristics [R3,R4,R5] in the Output Area Classifications (OAC), developed in collaboration with the Office for National Statistics (ONS), has informed multiple resource allocations and policy initiatives made by local governments. A prime example is in the commissioning of GSAC by the Greater London Authority (GLA) to build the London Output Area Classification (LOAC). The national OAC does not sufficiently capture the distinctive characteristics of London, particularly in terms of its ethnic structure, so GSAC developed the LOAC to classify all census output areas in London. The GLA used the LOAC "to provide up-to-date and improved estimates of small population groups and small area estimates and improve [its] knowledge of population migration and churn" [A1]. This helped the GLA identify areas experiencing rapid residential densification (leading to potential impacts on public service resource and provision) and informed the Mayoral plan for public services. LOAC also benefitted GLA school roll forecasting for boroughs to "support school place planning", and provided "the evidence base for other community services" [A1]. Without these data, the GLA would "either be over dependent upon conventional statistics, which are not updated with sufficient frequency, or would be reliant solely upon administrative sources which have some weaknesses and biases" [A1]. The LOAC also underpinned the Transport for London (TfL) Transport Classification of Londoners (TCOL); supplementing the LOAC with information from surveys that captured travel demand, behaviours, and preferences (2012-15), TfL developed a segmentation tool that categorises Londoners based on the travel choices they make, and the motivations for making those decisions [B]. The tool enables better planning and has informed the Mayor's Transport Strategy 2017-41, the Liveable Neighbourhoods Scheme, Healthy Streets Check, and cycling infrastructure model. As the Mayor's Transport Strategy Supporting Evidence report put it, "TCOL helps us identify what type of people are living in London, what their preferences are, how amenable they are to change and what might be effective in persuading them to change" [A2].

##### Improving understanding of COVID-19 and ethnicity

Ethnic inequalities in COVID-19 infections and outcomes are of national concern. GSAC research developing ethically rigorous ethnicity estimation methods [R2] has provided data to government bodies in England and Wales to inform understanding of the relationship between ethnicity and

COVID-19 infections and outcomes. Longley collaborated with the Welsh Government COVID-19 BAME Advisory Group and Public Health Wales, utilising the GSAC names classification software [R2] to, first, impute missing ethnicity information, and, second, examine ethnic variations in outcomes for patients hospitalized with COVID-19. The analysis added value to hospital data and enhanced understanding of ethnic disparities in hospitalization with COVID-19, specifically showing that ethnic minorities had increased risk of admission to intensive care units (ICUs) but not a greater likelihood of fatality. The analysis reinforced the continuing need: to break down COVID-19 impacts by ethnic group; to improve understanding of apparent ethnicity effects alongside neighbourhood deprivation; to guide policy and practice at the national and local levels; and to improve crisis communication. As an epidemiologist at Public Health Wales states, this “expert advice” and “rapid analysis” “inform[ed] the ongoing epidemic response in Wales” [B]. The research was developed with the Welsh Government COVID-19 BAME Advisory Group, presented as evidence to the Welsh Government Cross Party Group on Race Equality, and discussed with community groups and the Race Council Cymru. The same GSAC ethnicity classification approach has also been integral to the UK Government Joint Biosecurity Centre’s (JBC) pandemic response as it profiles the country for risk in the fight against COVID-19. As the JBC Programme Manager explains, “as COVID-19 impacts people from different ethnicities in very different ways the UCL data has been invaluable in helping us protect different communities through gaining a deeper understanding of the relationship between ethnicity and the spread of the virus” [C].

### **Facilitating Effective Bike Share Operations Globally**

The impacts of GSAC research advancing capabilities in data mining [R1] is clearly evidenced in the Bike Share Map [bikesharemap.com] that synthesises and visualises Big Data, in this case, global bikeshare systems in real-time. This delivers strategically valuable information to city authorities and transport operators worldwide. The Bike Share Map was the first online resource to integrate user data to compare c.500 bikeshare systems from around the world [R1], and has had over 1 million page views by more than 258,000 unique users since 1 August 2013 [D]. Users include Transport for London (TfL), San Francisco’s Municipal Transportation Agency (SFMTA), and PBSC Urban Solutions (a world leader in modular bikeshare solutions). In some cases, this is the only tool available for authorities to monitor their own bikeshare schemes, which can operate on a massive scale. An Assistant Engineer at SFMTA, where over 1.8 million trips were served by San Francisco’s bikeshare system in 2019, states that “the O’Brien Bike Share Map is the only way for us to know real-time numbers of bikes on the street, numbers of stations that are full and empty, and 24 hour historical dock/bike numbers at stations”. As this Assistant Engineer confirms, the Bike Share Map forms the basis of “our weekly conversations with the operator about how to improve operations to be in compliance with our contract” [E]. Operators, suppliers and transport authorities also use the Bike Share Map to compare usage between cities. TfL “primarily use [the site] for monitoring performance of our and other cycle hire services around the world, especially in London and the USA” [F]. For TfL’s Head of Cycle Hire, the map “has increased my knowledge of schemes around the world and provided insight into what Dockless cycle providers in London in particular are doing and how they are performing” [F]. The impacts of this work are on-going; for example, the coronavirus pandemic has driven a need within UK government for transport operators to examine bikeshare use during lockdown to frame changes in transport policy. As the Mobility Lead on the Data and Dashboard Team for the Cabinet Office’s COVID-19 taskforce explains, bikeshare has been integrated into their dashboard, which “serves as a central source of information for senior decision makers’ and that “the charts produced using this data were used in briefings to ministers and the Prime Minister” [G].

### **Improving the Effectiveness of Sports Investment**

DataShine maps created by GSAC [R6], including the original visualisation of 2011 Census data [6] and subsequent customised versions, have had c. 871,000 views and 488,000 unique users since August 2013 [H]. Bespoke versions of DataShine have been commissioned by the Scottish Government, National Records for Scotland, English Cricket Board and Birmingham County Football Association (BCFA). In 2018, GSAC produced an internal-facing mapping tool for BCFA which visualized the demographic composition of the Birmingham region to inform effective regeneration and engagement activities at local clubs. As BCFA’s Business Insights Manager,

describes, the tool was “widely used for the analysis and planning of activity for 2019/20” [I]. It guided decision-making in BCFA’s Club Improvement Project to support allocation of funding to increase sustainability of grassroots football clubs. The Datashine-based tool “enables us to identify areas of greatest need or benefit based on certain criteria and award grants accordingly” [I]. To date, this has steered decisions relating to GBP20,000 investment in pitches and facilities, across 795 teams, and c.17,000 players. As the Business Insights Manager points out, without the mapping tool, BCFA would “run the risk of awarding grants incorrectly” [I].

### **Enhancing Secondary School Pedagogy in Geography**

Datashine has been integrated into continuing professional development programmes for secondary school teachers to enhance their pedagogy by enabling access to national demographic datasets and providing a tool for the integration of GIS and digital mapping into the curriculum. The Data Skills and Partnerships Manager at the Royal Geographical Society (RGS) uses Datashine in national CPD training for Geography teachers and considers it to be “a brilliant tool” that “has widened the range of our CPD workshops, giving us a much-needed tool to help teachers develop the use of online maps to view powerful datasets and teach about a range of topics including development and changing place”. Without Datashine, he states, the RGS “would be struggling to help some teachers get started with using interactive maps, online data and GIS, as this is the perfect way for them to use such a tool, access relevant data and incorporate it into their lessons” [J].

**GSAC has proven the value of geodemographic data and more specifically has demonstrated how progressive advance in the technical and methodological approaches to the mining, integration, classification, and visualisation of these data transform their benefit to government, economy and society. In particular, in the innovative integration of data gathered through the contrasting endeavours of national censuses and the digital signature of consumerism, GSAC’s research has realised and enabled the effective use of data for the social good.**

### **5. Sources to corroborate the impact** (indicative maximum of 10 references)

- A. (1) Testimonial statement, Demography and Policy Analysis Manager, City Intelligence Unit, Greater London Authority and TfL Report: Transport Classification of Londoners (TCol): Presenting the Segments, 2017 <https://bit.ly/3q6f53d>; (2) Mayor’s Transport Strategy: Supporting Evidence, 2017; Mayor’s Transport Strategy, March 2018
- B. Testimonial statement, Epidemiologist, Communicable Disease Surveillance Centre, Public Health Wales
- C. Testimonial statement, Programme Manager, Joint Biosecurity Centre, Department of Health and Social Care
- D. Media coverage, views and users from Bike Share Map
- E. Testimonial statement, Assistant Engineer, Bike Share Program, San Francisco Municipal Transportation Agency
- F. Testimonial statement, Head of Cycle Hire, Transport for London
- G. Testimonial statement, Mobility Lead, Data and Dashboard Team, C-19 Taskforce, Cabinet Office
- H. Analytics data showing views and users from Datashine
- I. Testimonial statement, Business Insights Manager, Birmingham County Football Association
- J. Testimonial statement, Manager: Data Skills and Partnerships, Royal Geographical Society