

Institution: University College London
Unit of Assessment: 17B – Geography, Environmental Studies & Archaeology: Geography
Title of case study: The use of names to establish geo-genealogy and cultural, linguistic and ethnic affinity
<p>1. Summary of the impact (indicative maximum 100 words)</p> <p>UCL research has created a groundbreaking names classification tool for use by healthcare organisations, local government and industry. This improved the effectiveness of public service delivery to different cultural, linguistic and ethnic groups, in applications such as A&E admissions and GP referral patterns. It was used by the leading provider of commercial geodemographic segmentation of neighbourhoods as a more differentiated source of ethnicity information than Census sources alone. The public was engaged with research through popular websites and extensive media coverage, and the research has provided interactive tools through which science museums have improved public understanding of genetics and family history.</p>
<p>2. Underpinning research (indicative maximum 500 words)</p> <p>UCL Geography has longstanding interests in profiling the distinctive characteristics of neighbourhoods and the individuals who live in them, in order to understand better the structure and form of urban and regional systems. Since 2003, Paul Longley (Professor of GI Science, 2000 to present) and researchers (including Visiting Industrial Professor Richard Webber 2003-7) have investigated geographic concentrations of individual family names at scales from the local to the global. Important applications to migration were developed by Dr Pablo Mateos (ESRC Fellow from 2007, Lecturer 2008 to present).</p> <p>This included the development of a methodology to identify clusters of names in the British Isles [a]. Further research [b] showed how ‘naming networks’, constructed from forename-surname pairs in 17 countries, provide a valuable representation of cultural, ethnic and linguistic population structure around the world.</p> <p>Since 2007, the scope of the work has been broadened (part-funded by an ESRC Impact Award) to include data from many countries, and a names classification software was developed that is nearing the point at which it offers truly global coverage. A key product is Onomap, which allows users to classify lists of names into groups of common cultural, ethnic and linguistic origin using surnames and forenames (http://www.onomap.org). Although led by Longley and Mateos and funded via UCL Geography, ongoing work is conducted in association with Dr James Cheshire (UCL Centre for Advanced Spatial Analysis; CASA).</p> <p>This software development is the culmination of applied geographical analysis originally carried out through a Knowledge Transfer Partnership (KTP) between UCL and Camden Primary Care Trust. A ‘birth place geocoder’ was developed to improve the quantity and quality of ethnicity assignments on Camden’s medical records. A subsequent KTP with Southwark PCT developed this work into a formal names classification tool to target public health initiatives and healthcare delivery: the research was used to analyse GP referral patterns [d] and usage of accident and emergency facilities [c].</p> <p>Related work, including an ESRC Business Engagement Award with ESRI Inc., developed novel approaches to web mapping of names and Census of Population data. Underpinning research also developed out of two ESRC-co-funded CASE studentships. The first was in partnership with ESRI (UK) Ltd. (ESRC, 2008–11) and developed a number of novel applications of the family names databases using ESRI software. The second was in partnership with Southwark PCT (ESRC, 2008–11) and included the development of strategic applications of names-based classification of ethnicity.</p> <p>Most recently, the names classification was used by Longley and Dr Muhammad Adnan (Research Associate 2011 to present) to establish the probable ethnicity and age characteristics of a large (80 million) sample of Twitter microblog users. The academic motivation for this work is to examine patterns of segregation of different groups at different times of the day and week.</p>
<p>3. References to the research (indicative maximum of six references) (UCL authors [at time of research] in bold)</p>

Impact case study (REF3b)

[a] **Cheshire, J. A. & Longley, P. A.** (2012) Identifying spatial concentrations of surnames. *International Journal of Geographic Information Science* 26, 309–25. doi: [10.1080/13658816.2011.591291](https://doi.org/10.1080/13658816.2011.591291). (ISI Journal Impact Factor [JIF]: 1.613)

- Identifies the geographical 'footprint' of regionally and locally concentrated family names.

[b] **Mateos, P., Longley, P. A. & O'Sullivan, D.** (2011) Ethnicity and Population Structure in Personal Naming Networks. *PLoS ONE* 6(9), e22943. doi: [10.1371/journal.pone.0022943](https://doi.org/10.1371/journal.pone.0022943). (JIF: 3.730; Citations: 6)

- Describes the development of an operational classification of names into cultural, linguistic and ethnic groups.

[c] **Petersen, J., Longley, P. A., M. Gibin, P. A., Mateos, P. & Atkinson, P.** (2011) Names-based classification of accident and emergency department users. *Health and Place* 17, 1162–9. doi: [10.1016/j.healthplace.2010.09.010](https://doi.org/10.1016/j.healthplace.2010.09.010). (JIF: 2.419)

- A&E application of names classification written with KTP sponsor.

[d] **Lewis, D. J. & Longley, P. A.** (2012) Patterns of patient registration with primary healthcare in the UK National Health Service. 2012. *Annals of the Association of American Geographers* 102, 1135–45. doi: [10.1080/00045608.2012.657500](https://doi.org/10.1080/00045608.2012.657500). (JIF: 2.110)

- GP referral application written with ESRC CASE PhD student.

Evidence of the quality of research is provided by the series of external grants received (total £713,118 plus 1 ESRC CASE studentship) for this research, and publications in major peer-reviewed journals. Research grants, leading to [c] and [d] include:

2004–7 Economic and Social Research Council/Camden Primary Care Trust: a Knowledge Transfer Partnership to develop and utilise GIS for neighbourhood profiling and assist in targeting public health and health care delivery (KTP000037: £331,584). Official grant evaluation grade A (Outstanding).

2006–9 Economic and Social Research Council/Southwark Primary Care Trust: a Knowledge Transfer Partnership to develop systems to measure and monitor GP referrals and to target health promotion campaigns (KTP000666: £331,584). Official grant evaluation grade A (Outstanding).

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

Research conducted at UCL has improved the provision of targeted local health services through Camden and Southwark Primary Care Trusts. The underpinning names classification has subsequently been licensed to healthcare and government organisations, as well as to CACI Ltd. to improve the industry-leading ACORN commercial neighbourhood classification system used throughout business, government and public services. The public has been engaged with the research through innovative websites that allow searching and tracking of surnames across the UK and the world. Museums (such as London Science Museum) and heritage organisations (such as the National Trust) have presented these data to the public through interactive exhibits or their own websites.

Licensing classification tools to public sector organisations: Names classification software was applied, in partnership with two London Primary Care Trusts (PCTs), to analyse the ethnic backgrounds of those seeking screening and care, and to target interventions accordingly. This occurred, as described in section 2, through Knowledge Transfer Partnerships (KTPs), with Camden PCT and Southwark PCT, two of the country's most ethnically diverse boroughs [c, d], where vagaries in the recording of ethnic background hampered effective targeting of health interventions. Building on the KTP, Southwark PCT hosted a 2011 pilot project seeking to increase extremely low rates of breast cancer screening amongst women of African Caribbean descent. As part of this project, the names classification was used to identify the ethnic groups of women who missed screening, and then targeted resources and information accordingly, leading to an increase in the uptake of screening among African Caribbean women [1]. These partnerships, in turn, had an impact on how these and other PCTs understand GP referral patterns and admissions to A&E. For instance, PCT staff worked with UCL researchers to use surname data and challenge a commonly held perception that A&E usage differs by ethnicity [c in section 3 above].

As awareness of the names classification tool has increased, over 15 PCTs, strategic health

authorities and other government organisations used it under licence between 2008 and 2013. For example, the Health Protection Agency, England (now Public Health England) used the software for ethnic classification in sentinel surveillance of hepatitis and other blood-borne viruses in 2011 and 2012 [2]. NHS Lothian, Scotland, licensed the software in August 2010 to code patient records by ethnic group and determine differential disease prevalence, as well as the level of uptake and accessibility to public health prevention services such as smoking cessation or cancer screening. NHS Lothian officials also used it to assess need and usage of interpretation in GP surgeries by Polish speakers, and to assist in developing a new interpretation service (ITS) contract [2].

In the business sector, the classification was licenced to CACI Ltd for its ACORN classification – one of the most widely used general purpose geodemographic classifications in the UK today. ACORN is licensed to a very wide range of customer-facing organisations seeking effective communication with target groups. It has over 500 core licensees with long-term use of the complete dataset throughout their organisations, including government departments, local authorities, hospitals, banks, etc, and many others making more restricted use. Culture, ethnicity and linguistic group is profoundly important in shaping the neighbourhood geography of the UK as the country becomes more multi-ethnic and its ethnic minorities become more geographically dispersed. Recognising the increasing importance of effective classification of ethnicity, and cognisant of prospects for future UK population censuses, UCL research was used by CACI to provide a more differentiated source of ethnicity information than Census sources alone [3]. In March 2013, the latest ACORN classification was released, using this new approach, and was made available to CACI's huge customer base.

Engaging a UK and international public with research: Through online maps that graphically present the research described in section 2, research has engaged a global public with an interest in family histories and historic migrations and, at the most basic level, the question: where do we come from? During the impact period, the website collectively was accessed by over 4 million unique visitors.

The Public Profiler website (gbnames.publicprofiler.org) was originally developed, in 2006, out of the initial investigation into the geographic patterning of names in Great Britain. This was accessed by 1.6 million unique users in its first year alone, and between 1 September 2007 and 31 July 2013 had over 1.3 million unique visitors, each of whom spent an average of nearly five minutes on the site and looked at 15 pages: this indicates a substantial degree of user engagement [4]. As a result of this widespread interest, and to mark the centenary of its founding legislation, the National Trust licensed a version of this website for three years (August 2007–2010) [5].

When the initial analysis was extended to 25 other countries, an international surname mapping site (worldnames.publicprofiler.org) was created in 2007. Google Analytics show that this site was visited by more than 3.6 million unique users between 2008 and 31 July 2013, with each spending over 3 minutes viewing 6 pages [4]. Visitors originate predominantly from the USA and UK, with significant numbers from the European continent. James Cheshire was commissioned by *National Geographic* (US print circulation 5m) to draw upon the research to create a map of surname distribution in the United States, which appeared in February 2011. [7]

A Twitter names map of London (twitternames.publicprofiler.org) was launched in December 2012 as part of the EPSRC Uncertainty of Identity project and building on names classification data. This was reproduced widely on high-circulation news websites, such as the *Mail Online* (24 April 2013; 170 comments, demonstrating engagement with research; 2.1m daily visitors) and *Guardian* (13 Dec 2012; nearly 1.4m daily visitors), collectively reaching up to 3.5m readers.

Interactive educational exhibits for museums, increasing public understanding of science: Capitalising on the popularity and user-friendliness of the research and the website, science museums have utilised the research in interactive exhibits and to improve their own presentation of science topics such as identity and genetics. To date, over four million visitors to various museums are estimated to have been exposed to this interactive exhibit, and its significance may be gauged by the relationships cemented with various museums during the impact period.

The At-Bristol science discovery centre used the research to develop an interactive exhibit for the Inside DNA exhibition (December 2007 to August 2008), making the understanding of genetics both engaging and personally relevant by enabling visitors to trace migration patterns by the

Impact case study (REF3b)

spread of their own surnames. Inside DNA subsequently became a 5-year travelling exhibition and appeared at museums including National Museums Liverpool (September 2010 to February 2011; 1.2m visitors in this period), MOSI Manchester (May 2011 to November 2011; 536,000 visitors in this period), and Thinktank Birmingham (May 2012 to November 2012) [8]. Another interactive exhibit was developed with London Science Museum (approximately 2.7m visitors per annum) where the exhibit was included in the 'Who Am I?' interactive exhibition (2010–15); visitors engage with complex issues of identity, genetics and inheritance through reference to the changing geography of visitors' own family genealogies. In the first year alone (2010–11), an evaluation found the redeveloped 'Who Am I?' gallery had over a million visitors (twice the expected number) [9]. In 2012, this research was selected for a display on Genetic Maps at the Royal Society's prestigious Summer Exhibition, showing how genetic makeup, names and facial characteristics are distributed across the UK (11,120 visitors, 3–8 July).

Most recently, the Glasgow Science Centre's £1.9 million BodyWorks exhibition has the Family Names interactive exhibit, based on Public Profiler, as a key part of its exhibit, which "aims to allow visitors to learn about genetics, inheritance and cell biology". From its opening in April 2013 to August 2013, the exhibit received "overwhelmingly positive feedback" and attracted over 143,000 visitors and over 32,000 school children [10].

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

[1] Patient Navigation Pilot Project, July 2011. Selection of Southwark citing KTP research (p. 4) and conclusions from pilot (p. 26) <http://bit.ly/17o6fSu> [PDF].

[2] Licensees include: Camden PCT, Islington PCT, Southwark PCT, NHS Lothian, NHS Scotland / General Register Office Scotland (GROS), Health Protection Scotland, Health Protection Agency, England, Glasgow Centre for Population Health, West Midlands Cancer Register, London School of Hygiene and Tropical Medicine, Dartford Council, and Loughborough University (for a study for Loughborough Local Authority).

NHS Lothian ITS assessment and new contract: see presentation by NHS Lothian researcher, Dr Fatim Lakha (p. 11; use of Onomap noted on p. 8) <http://bit.ly/GzGcha> [PDF].

Public Health England, 2011 sentinel surveillance of blood-borne viruses (published January 2013): <http://bit.ly/18oXF5z>.

Sentinel surveillance of HepC 2008-2012, in the 'Hepatitis C in the UK 2013 report' (see page 61 for usage for East European population) <http://bit.ly/1eWVxan>.

[3] ACORN Technical Document, March 2013. See especially p. 14, item 4.10 for note on UCL research contribution: <http://bit.ly/1boNimT>. See also statement from CACI Ltd on use of the names classification system.

[4] Google Analytics data compiled and available at: <http://bit.ly/17o6HQE> (Annexes 2 and 3).

[5] At the end of the licence term, web traffic was temporary redirected to the publicprofiler website, but this arrangement has since lapsed. An archived version of the National Trust site may be viewed at the Wayback Machine. See, for example, the archived page from 10 July 2008: <http://bit.ly/1hi7ZhA>.

[7] Media usage of research: Daily Mail <http://dailym.ai/16UnZAz>; Guardian <http://bit.ly/15IBDrc>; National Readership Survey website figures for July 2012–July 2013: <http://www.nrs.co.uk/nrs-data-tables/>. National Geographic 'What's in a surname? A new view of the United States based on the distribution of common last names.' February 2011, p. 20–21 and online at <http://bit.ly/1hi88li>.

[8] Inside DNA travelling exhibition page <http://bit.ly/1felne>. Visitor figures for national museums from Department of Culture, Media and Sport <http://bit.ly/1dyhNI7>. Others derive from the annual reports of individual museums. Royal Society Summer Exhibition: <http://bit.ly/18KwqTO>.

[9] Statement provided by Science Museum on 'Who Am I?' exhibit.

[10] Statement provided by Project Manager, Body Works, Glasgow Science Museum, Visitor data supplied was for 31 Mar to 31 Aug 2013.