Accepted: 18 November 2018
DOI: 10.1002/psp.2227

RESEARCH ARTICLE

WILEY

Bringing historical British Population Census records into the 21st century: A method for geocoding households and individuals at their early-20th-century addresses

Nigel Stephen Walford

Department of Geography and Geology, Kingston University, Kingston upon Thames, UK

Correspondence
Nigel Stephen Walford, Department of Geography, Geology and the Environment, Kingston University, Penrhyn Road, Kingston upon Thames KT1 2EE, UK.
Email: n.walford@kingston.ac.uk

Funding information
British Academy, Grant/Award Number: SF110183

Abstract

Increased use of geographical information systems (GISs) for social scientific research has highlighted the analytical possibilities offered by unlocking historical data sources that are impractical with contemporary data. Genealogists and social scientific researchers have used information relating to individuals, households, and their addresses contained in the British historical census records for a number of years. This article takes geographical exploration of these a stage further by developing a method of attaching grid references to the census addresses, thereby opening up new research possibilities including spatial analysis. The resultant geocoded census data enable address-, household-, and individual-level historical research and aggregation to contemporary and historical spatial units for exploration of demographic and socio-economic change. The paper focuses on the development of a method of geocoding 1901 and 1911 Census data in respect of six study areas in the historical counties of London and Middlesex with over 260,000 individuals within some 60,000 households or communal establishments in both 1901 and 1911. Successful semi-automated matching of historical census addresses with a contemporary address database is related to population density, change from property naming to numbering, residential development, thoroughfare name changes, and transcription error. Complete geocoding was completed by manual digitising with the aid of ancillary information sources and geographical information embedded in census records. The method outlined has the potential to be replicated in other areas and be adapted to use with other contemporary address databases that contain grid references. The paper outlines the factors that would influence transferability of the geocoding method.

KEYWORDS

address matching, geocoding addresses, historical census data, historical GIS, London and Middlesex

1 | INTRODUCTION

Developments in the handling, processing, and analysis of geospatial and statistical attribute data from population censuses have gone hand in hand with the increased sophistication and availability of geographical information systems (GIS) in the United Kingdom and other countries over a period of at least 40 years (Marx, 1986; Scholossberg, 2003). Much of this work has concentrated on the visusalisation, analysis, and interpretation of these demographic and socio-economic statistics in aggregate form (Caruthers, 1985; Gaits, 1969; Shepard, 1984), although the availability of sampled household and individual data has also allowed researchers to model population characteristics (Boyle & Shen, 1997; Shouls, Congdon, & Curtis, 1996). The physical size and number of spatial units available for such investigations have respectively decreased and increased over the period. Table 1, going back to 1971 when information technology software started to make analysis and geovisualisation of census statistics for small areas feasible, summarises these changes showing an increase in the number of small areas and the impact of the transition from enumeration districts (EDs) to output areas (OAs) across England, Wales, and Scotland in 2001 together with a reduction in mean area and total population (there was a small increase in mean population in 2011). Despite the ability to handle, process, and analyse this increasing number of zones and an expanded quantity of statistical information, the opportunity to examine the detailed patterns created by where people live, in other words the addresses of the dwellings they occupy, has remained elusive. For obvious and legitimate reasons, the census authorities are precluded from releasing census microscale data for the geocoded locations where individuals and households live at the present time and these records are embargoed for 100 years.

Different forms of geospatial analysis, such as those involving areal interpolation and dasymetric mapping, have attempted to locate populations by aggregation to their residential locations using secondary data sets on land cover and applying weights to assign proportions of the population to urban, suburban, countryside, and other categories of land area (Eicher & Brewer, 2001; Mennis & Hultgren, 2006). Such methods have also been used to address the issues arising from the changes to the boundaries of the spatial units employed in different enumerations, especially in the United Kingdom (Norman, Rees, & Boyle, 2003; Walford, 2013). However, changing our focus of attention from the present to the past enables us to benefit from being able to examine and analyse microscale data in respect of addresses, households, and individuals and small spatial units that were collected by census enumerations carried out 100 years or more ago. These historical census records, for a number of years the mainstay of genealogical research, provide the opportunity to explore the spatial patterns of earlier eras and to investigate the persistence and mutability of the demographic and socio-economic character of areas, streets, and individual properties by comparing such historical data with aggregate statistics for the present day. The potential benefits arising from such microanalyses have been demonstrated for Canada (St-Hilaire, Moldofsky, Richard, & Beaudry, 2010) and parts of the United States (Logan, Jindrich, Shin, & Zhang, 2011; Xu, Logan, & Short, 2014).

TABLE 1 Changes in the numbers of small areas and their mean total population and area in the British Population Censuses 1971 to 2011

Census year	Enumeration districts or output areas	Mean total population	Mean area (ha)
1971	125,476	430.2	182.5
1981	130,431	405.4	175.5
1991	151,719	344.9	150.9
2001	218,038	261.9	105.0
2011	227,759	269.5	100.5

Source: Office for National Statistics.

The main aim of this paper is to develop a semi-automated method for geocoding the addresses held on the historical census records from the British 1901 and 1911 Population Censuses for a selection of contrasting areas within what is now the Greater London Authority (GLA). Section 2 examines the characteristics of the historical and contemporary geospatial and census data sources used to develop and test the method, which is described in Section 3. Section 4 considers the results of applying this procedure in a selection of areas in the former London and Middlesex counties. Section 5 considers the broader implications of the research including the potential to extend the method to earlier and, in due course, subsequent British historical census records and to the issues potentially facing researchers in the 22nd century should records from a census-style enumeration no longer be available.

2 | BACKGROUND AND DATA SOURCES

Each census reflects the economic, social, and technological conditions of the time when it is conducted, and the British Population Censuses of 1901 and 1911, the last held before enactment of the 1920 Census Act, represented the culmination of over 100 years of enumerating the British population on a decennial basis. The head counts of the early-19th-century censuses were superseded by recording of a wider range of information in enumerator's books from 1841 onwards, which reflected the growing need for statistics as governmental bureaucracy and the population itself grew in size and complexity. Historians and geographers have made use of historical census records in their research for many years. Anderson's research in the 1970s transcribing a sample of records from the 1851 Census (Anderson, 1972, 1987; Anderson & Collins, 1973; Anderson, Collins, & Stott, 1977) and Southall and Gregory's work in the 2000s digitising of the geographically aggregated census statistics back to 1801 and mapping population change (Gregory, 2002; Gregory, Bennett, Gilham, & Southall, 2002; Gregory, Dorling, & Southall, 2001; Southall, 2003, 2006, 2014) constitute seminal examples. Local historical studies have charted demographic change in settlements, for example, Tilley and French's (1997) work in Kingston upon Thames, and between urban and rural areas (Hinde, 1985). Researchers have used GIS as a framework for capturing and analysing these census data in order to reveal historical patterns and processes. The Integrated Census Microdata (ICEM) project (Higgs, Jones, Schürer, & Wilkinson, 2013; Schurer & Higgs, 2014),

by connecting with The National Archives' (TNA's) commercial partner FindMyPast (part of Bright Solid), has crowned Anderson's sampling of census records by creating a data resource of British census records for the decennial enumerations spanning the period 1851 to 1911, although the open access version of the ICEM does not permit users to view details of individuals' geographical location. The method presented here makes use of the original census records in conjunction with the ICEM data. The following subsections examine characteristics of the data sets used to develop the geocoding procedure before outlining the method itself and assessing its success. Researchers now have more immediate access to a wider range of modern and historical data sources than was the case in the past (Marten, 1971; Stilwell, 2005; Walker, 2016). Validating the quality of these sources remains an important issue, especially where they have been obtained through "crowdsourcing" or are held on voluntarily maintained websites, such as those run by local history societies. Some sources of data, such as those operating with the support of public research funding (e.g., the U.K. Data Service) or where public bodies work with commercial partners, may be considered reliable, although researchers still need to assure themselves that the data provided contain information that is relevant to and capable of answering their research questions.

2.1 | Historical and modern census data sources

The method of geocoding addresses in the British 1901 and 1911 censuses potentially forms a starting point for extending this approach to both earlier and subsequent enumerations, respectively, in the 19th and 20th centuries and perhaps most significantly to those from the 1921 Census that will become accessible within 3 years. It is therefore important to reflect on the 1901 and 1911 censuses as part of a historical, evolutionary sequence of such enumerations. Early-20th-century censuses enumerated people where they were on "census night," the population present basis, rather than using the now familiar usual residence approach. The Censuses carried out overnight on March 31/April 1, 1901, and April 2/3, 1911, have some key demographic variables in common with those held on April 28/29, 2001, and March 27/28, 2011, as well as with those carried out over the intervening century. Common variables include those that yield counts of the numbers of males and females and persons in different age and occupational groups. However, even some of the common variables may have undergone changes in their definition over the period, such as in respect of the categories used for household types and occupational groups. During the 20th century, there was generally an increase in the number of census topics, but the 1911 census expanded on those included 10 years earlier by asking about people's industry of employment and by uniquely in the history of British censuses surveying married women's fertility. Furthermore, the 1901 Census had introduced new procedures for obtaining information from people in private households or communal establishments or on vessels. Table 2 details the topics covered by the 1901 and 1911 Censuses and indicates whether the same topic was also present in the 2001 and 2011 enumerations, irrespective of whether they were asked with different wording and/or defined in a different way.

TABLE 2 Topics included in 1901 and 1911 British censuses in England with an indication of their presence in 2001 and 2011

Addressed to	1901	1911	2001/2011
Households	Address	Address	Yes
	Number of rooms if less than 5	Number of rooms	Yes
		Building type	Yes
Individuals	Name	Name	Yes
	Relationship to head of family	Relationship to head of family	Yes
	Marital status	Marital status	Yes
	Age	Age	Yes
	Sex	Sex	Yes
	Occupation	Occupation	Yes
	Birthplace (level of geographical detail sought depended on whether birthplaces were in England, Wales, Scotland/Ireland, British colony or dependency, or a foreign country)	Birthplace (level of geographical detail sought depended on whether birthplaces was in England, Wales, Scotland/Ireland, British colony or dependency, or a foreign country)	Yes
	Medical disabilities (deaf, deaf/dumb, blind, lunatic, imbeciles, and the "feeble-minded")	Medical infirmities (deaf, deaf/dumb, blind, lunatic, imbeciles and the "feeble-minded")	No
		Marital fertility (total live births to women in their present marriage, number still alive, and number who had died)	No
		Age at marriage	No
		Nationality of people born outside of the country	Yes
		Employment status	Yes
		Whether working at home	Yes
		Industry or service of employment	Yes

Users of modern census statistics are familiar with the principle of cross-tabulating the categories of one household or individual characteristic (variable) with those of one or more others to produce counts of different population units (e.g., households, persons living in communal establishments, and persons aged 16–64) in tables containing cells of these combinations that are capable of being aggregated to a range of spatial units. Traditionally predefined sets of cross-tabulations were published by the British census authorities originally on paper and subsequently also digitally. Generation of such tables dynamically "on the fly" from the household and individual data was a notable innovation for 2011 output. Some census analysts are also familiar with the samples of anonymised records that have been randomly selected from the U.S. Censuses since 1970 and from British Censuses since 1991 and retrospectively back to 1961. These samples are composed of household and individual data records for a specific sampling fraction and in effect are equivalent to large sample surveys, although they are drawn from databases that include data about the entire statistical and human population in a country or occasionally a region. In other words, a very substantial percentage of the total information collected and processed is discarded

together with low-level (detailed) geographical coding to create a sample comprising typically 1% or 5% of the population.

Nevertheless, the documents on which members of the population record their data, the census forms or schedules, are preserved as part of the nation's archive of records for future generations. Traditionally, these records were passed over to the Public Records Office, which was reconstituted as TNA in 2003 and preserved as paper documents that were available for consultation by members of the public and researchers 100 years after a particular census took place. Such historical records may experience a degree of physical degradation over time as a consequence of fire or water damage or other mishap, such as the fire that destroyed 1931 English and Welsh records in 1942. Recently, TNA and similar archives responded to the growing use of the Internet to access information and a wish to arrest further deterioration of their collections by entering into partnerships with academic and commercial organisations to undertake a programme of digitisation or scanning to make their records more accessible to researchers and the public. As the remainder of the 21st century unfolds, the preservation of records that were only collected by digital means will present archives with new challenges until such time when the public records collected during the digital era pass their 100-year closure period and need to be made accessible. This will be the case with the forthcoming 2021 British Population Census, which will be mainly be conducted online (ONS, no date).

The collection of people living at an address may, for census purposes, be separated into more than one household, because the census authorities define a household as a group of people living together behind one front door or eating a meal together at least once a day. Tables 3 offers examples of the level of detail that can be obtained through consultation of the 1901 and 1911 Census records. The starting point is an address in Deptford, 29 Amersham Vale, which was randomly selected on the basis that at least one head of household present in 1901 also appeared in the 1911 census records for Deptford at a different address. Three households occupied 29 Amersham Vale in 1901. The first household comprised William and Phoebe Ward, who were recorded as having their granddaughter Grace M Edge aged 5 years staying with them on census night, who was presumably Mrs Ward's daughter's child on account of the different surname. The second household at 29 Amersham Vale in 1901 were a husband and wife (William and Sarah Bagwell). The third household comprised Mary Andrews, recorded as married but without her husband being identified and her daughter Grace Andrews. William and Phoebe Ward remained as one household at 29 Amersham Vale in 1911, and the second household present comprised Grace M Edge, now 10 years older, together with her father and older and younger sisters, including one born in 1901 and another in 1903. However, a further twist in her story emerges when using this information to search for her parents in 1901 as the census documents show them present at 46 Etta Street Deptford with not only her two older and two younger sisters but also Grace M Edge herself (aged 5 years). This example illustrates the type of inconsistency and possible double-counting that may be embedded within the census records that are difficult to uncover. The third household from 1901 (Mary and Grace Andrews) had moved away from Deptford. The Bagwell household, which had moved to a different address in the same local authority during the intercensal period, also increased in size as a result of the births of three children, a son in 1902, a daughter in 1904, and a second son in 1909. This enlargement of family size undoubtedly contributed to the decision to seek larger accommodation. Examination of the year of marriage column in 1911 reveals that Mr and Mrs Bagwell had been married for 10 years, shortly before the 1901 census, and started married life together in two rooms, but by 1911, the enlarged household occupied four rooms. Mrs Bagwell remained outside the paid workforce, whereas her husband had changed employment from a night porter and watchman to an assistant reliefing (sic) officer in "sheltered accommodation" known as Mary Ann's Buildings. This change appears to have offered him the opportunity to secure a larger dwelling space for his family, which seems likely to have been connected with his new employment as they also lived in Mary Ann's Building.

2.2 | Historical and modern topographic mapping and addresses

Using thematic maps as a means of visualising the changes revealed by successive censuses has become an important way of conveying the dynamics and characteristics of populations to policymakers, researchers, and members of the general public. Since the 1960s, but especially after the burgeoning of computer mapping and GIS technology in the 1970s and 1980s, there has been widespread growth in the dissemination of population information via digital media, including interactive mapping. Many of these maps portray census counts either for the physical areas to which they have been aggregated (e.g., Champion et al., 1996) or for stylised shapes using cartograms (Dorling, 1994). However, the uniform shaded areas depicted on choropleth maps are supported by a topographic underlay comprising the properties and addresses where households and individuals were documented on the census records. The British national mapping agency, the Ordnance Survey (OS), commenced surveying the country to produce the first County Series topographic maps showing building outlines at 6 in. to the mile (1:10,560) scale in 1840 and extended this to the more detailed 1:2,500 (25 in. to the mile) scale in 1854. Almost as soon as the first editions of maps at these scales had been published in the 1890s, the OS had started to resurvey the country in order to publish a revised edition, which was finished before World War I. Further resurveying started in 1907 and continued well into the mid-20th century (1940s), although it was not completed and the third edition or second revision maps were only published for areas experiencing significant change (Harley, 1975). The OS has maintained a nearly complete (93%) archive of its historical topographic maps, and in 1995, the Landmark Information Group started a programme of work to scan these maps digitally at 300 dpi and georeference them to the British National Grid. Some years later, the seamless mosaic of these scanned map tiles became available to the higher education community to browse and to download from Edina at the University of Edinburgh.

The historical topographic maps current at the time of the 1901 and 1911 Censuses show the names of road, street, and other thoroughfares but do not record the names or numbers of individual

TABLE 3 Linked household-, family-, and individual-level transcriptions of 1901 and 1911 census records

1901

Schedule, address and rooms			Name and relationship		Particulars of marriage	Age and sex		Profession or occupation	Birthplace	
Schedule number	House number /name, road/street name	Number of rooms if less than 5	Name and surname	Relationship to head of family	Condition as to marriage	Sex	Age last birthday	Profession or occupation	Birthplace	If: (1) deaf and dumb, (2) blind, (3) lunatic or (4) imbecile, feeble-minded
473	29 Amersham Vale	4	William Ward	Head	Married	Male	48	Potman	Whitechapel	-
			Phoebe Ward	Wife	Married	Female	63	-	Bermondsey	-
			Grace M Edge	Granddaughter	Single	Female	5	-	Deptford	-
474		2	William Bagwell	Head	Married	Male	25	Night Porter & Watchman	Camberwell	-
			Sarah Bagwell	Wife	Married	Female	23	-	Bermondsey	-
475		2	Mary Andrews	Head	Married	Female	28	Dressmaker	S Norwood, Surrey	-
			Grace Andrews	Daughter	Single	Female	6	-	City Road, London	-

1911

Schedule, address and rooms			Name and relationship		Age and sex		Particulars of marriage and fertility				Profession or occupation and employment					Birthplace and nationality		
Schedule number	House number/ name, road/ street name	Number of rooms	Name and surname	Relationship to head of family	Male	Female	Marital status	Completed years of present marriage	Total children born alive	Children still living	Children who have died	Personal occupation	Industry or service	Employment status	Whether working at home	Birthplace	Nationality for persons born in foreign country	If: (1) deaf and dumb, (2) blind, (3) lunatic or (4) imbecile, feeble-minded
336	Relief Office Mary Ann's Buildings, High Street	4	William Bagwell	Head	35		Married	-	-	-	-	Assistant Relieving Officer	Board of Guardians	Worker	No	Camberwell London	-	-
			Sarah Bagwell	Wife		33	Married	10	3	3	0	-	-	-	-	Bermondsey London	-	-
			Albert Bagwell	Son	9		-	-	-	-	-	-	-	-	-	Deptford London	-	-
			May Bagwell	Daughter		7	-	-	-	-	-	-	-	-	-	Deptford London	-	-
			William Bagwell	Son	2		-	-	-	-	-	-	-	-	-	Deptford London	-	-

Note. Dashes (-) denote blank entries on the census record indicating not applicable; household-level data have not been duplicated for individuals.Source: The National Archive RG13/527/168/3; RG14PN2635 RG78PN89 RD28 SD2 ED12 SN336.

addresses. However, the naming of certain types of residential address such as vicarages, farms, schools, hospitals, police or fire stations, significant houses, and similar properties does occur. Nevertheless, the geocoded historical map tiles by including building outlines over which address data can be overlain have helped with geocoding addresses in the 1901 and 1911 Censuses. The need to start resurveying for the revised or second edition maps in the 1890s reflects the rapid urbanisation and suburbanisation taking place at that time. Figure 1 illustrates the local impact of such land cover change in the Borough of Hammersmith, west London. The inset from the First Revision County Series sheet TQ2280 published on January 1, 1896, covers land on the fringe of the built-up area in Hammersmith and shows the presence of late Victorian residential streets towards the south (streets such as Ellerslie Road and Ethelden Road were not on the first edition map published on January 1, 1871). It also shows a farm, fields, used and disused brick yards, a school, and old clay pits. The second revision map, published on January 1, 1916, shows that most of the southern half of the same inset had undergone further residential development

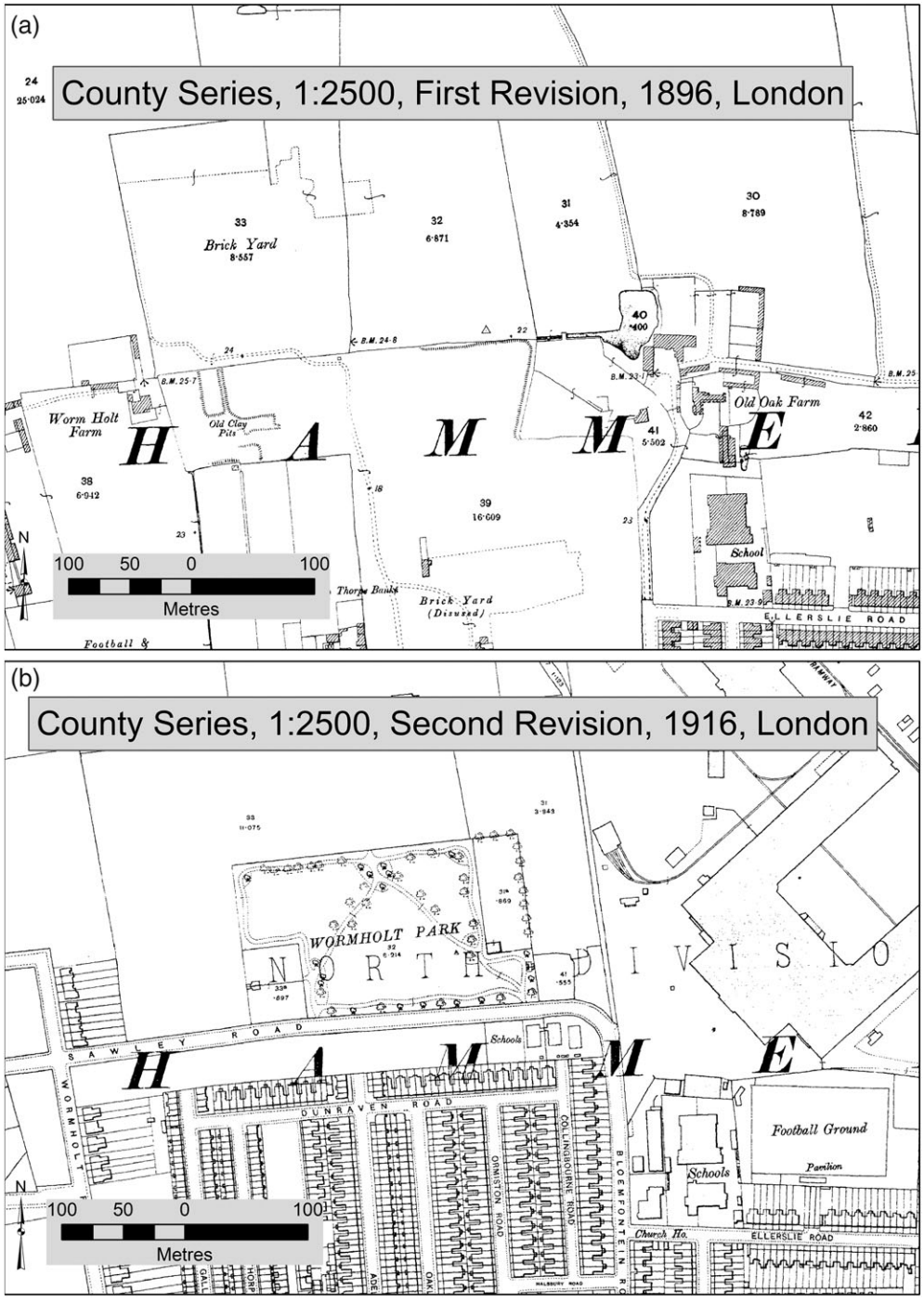

FIGURE 1 Comparison of Ordnance Survey base topographic mapping for an area in Hammersmith west London. (a) County Series, 1:2,500, First Revision, [1896, London]. (b) County Series, 1:2,500, Second Edition, [1916, London].
Source: Ordnance Survey, Edina. © Crown copyright and/or database right 2018 OS

along new roads (e.g., Dunraven Road and Wormholt Road) and new schools on Bawley Road where it joins Bloemfontein Road. The farm had disappeared, and one field had become a football ground, and parts of other fields had been converted into Wormholt Park. Part of the expanding underground rail system, here running overground, had arrived in the north-east quadrant of the inset, and additional buildings had been erected west of the school.

Many of the man-made features depicted on modern topographic maps are unlikely to be relevant as an aid to geocoding historical census addresses. However, modern address databases have the potential assist with historical geocoding. This is the case in localities where historical residential properties continue to be occupied as private or communal addresses and have not been deliberately demolished or reconfigured and not destroyed by enemy action during World War II and where persistence of address numbering can be verified or changes determined. The OS's main digital topographic database, MasterMap, includes Address Layer 2 (MMAL2), which contains 1-m National Grid coordinates, full address details, and postal geography for all addresses in the country together with similar location information for nonpostal point features. Some of the other layers in the MasterMap database (e.g., topography and integrated transport) are openly available for teaching and research purposes to U.K. Higher Education institutions through Edina through a service known as Digimap. However, the MMAL2 is not included, mainly because of its value as a product for commercial customers, although, as in the present project, this layer can be used for academic research under special licence from the OS with approval from Royal Mail. The MMAL2 should be viewed as an example of an address database containing grid references to each address, and other options will be available in different national contexts, and within the United Kingdom, the method outlined here could be adapted to work with other contemporary address databases.

The following sections describe the methods developed to link these data sources in order to geocode historical census addresses.

3 | METHOD FOR GEOCODING HISTORICAL CENSUS ADDRESSES

3.1 | Selection of study areas

The historical and contemporary census and geospatial data sources just outlined exist in a fairly consistent format and structure for most of the United Kingdom and excepting MMAL2 are accessible to researchers through service providers such as Edina and the U.K. Data Service and a TNA partner organisation in the case of scanned historical census records. However, attempting to geocode the national sets of 1901 and 1911 census addresses would have been impractical with the resources available, and the research should be viewed as a proof of concept. The development and testing of the geocoding procedure reported here represents a case study that provides a starting point for potentially embarking on such a large-scale endeavour. London and other cities had already become major centres of population by the start of the 20th century as a result of industrialisation and associated rural-to-urban migration in the 19th century. The geocoding procedure was developed using a selection of six areas from across the former counties of Middlesex and London, which are themselves now almost entirely within the GLA area. The combined population of Middlesex and London was 5.3 million in 1901 and 5.6 million in 1911, and the population of the equivalent area was 5.7 million in 2011, which represented some 69.1% of the GLA.

Throughout England and Wales in the first decade of the 20th century, there was a two-tier system of local government comprising

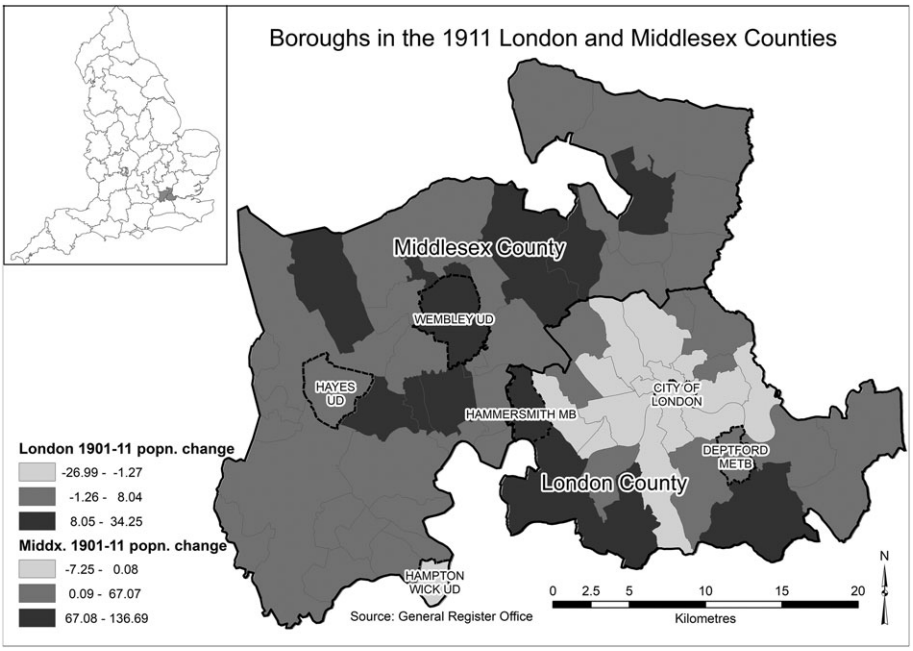

FIGURE 2 The 1901–1911 population change in former London and Middlesex boroughs and districts highlighting case study areas in relation to present-day Greater London Area.
Source: Ordnance Survey, Edina. © Crown copyright and/or database right 2018 OS

"higher" level counties within which there were municipal boroughs (MBs), county boroughs, urban districts (UDs), and rural districts, with these different labels partly reflecting historical status and partly their function and population density. The term metropolitan borough was introduced in 1900 with respect to the local authorities within the County of London, and the terms municipal borough and metropolitan borough are typically abbreviated as MB. The historical counties of Middlesex and London between them contained each of these types together with the special case of the City of London CC, considered for nonadministrative purposes as a separate county: There were 28 metropolitan boroughs (MBs) plus the City of London CC in the County of London and 37 MBs, UDs, and rural districts in Middlesex. Figure 2 shows that all but one area in Middlesex experienced population growth during the first decade of the 20th century, whereas increases occurred in less than a third of London's areas. These different experiences of population growth together with the intention of including localities north and south of the River Thames and within modern inner and outer London provided the starting point for selecting areas for inclusion. Six areas were chosen, namely, City of London CC, Deptford MB, Hammersmith MB, Hampton Wick UD, Hayes UD, and Wembley UD: The first three had 4.9% of London County's population in 1901 and 1911; and the latter three had 4.8% of Middlesex's (the same percentages also apply to the number of households). These percentages translate as 260,408 and 267,927 individuals and 55,911 and 61,005 households or communal establishments in 1901 and 1911, respectively. These areas experienced contrasting trends of population change 1901–1911, decline, stability, or substantial growth (see Figure 2). Another consideration when selecting these areas was that the number of addresses to be geocoded would be feasible with the available resources: There were 37,361 and 39,894 addresses across the six areas in 1901 and 1911, respectively. The spread of local authority types allows the selected areas to realistically be treated as a template for geocoding similar types of area across other historical counties in England and Wales.

3.2 | Geocoding method

The method of unlocking the geovisual and spatial analytic potential of historical data sources commonly starts with capturing X,Y coordinate point data. One option for achieving this is to use an online geocoding service, but these will not find points that no longer exist. Historical gazetteers offer an alternative approach, but these may be spatially and temporally partial in their coverage. Navickas (2016) suggested that a combination of approaches might yield greatest accuracy and completeness. Hitchcock et al. (2015) achieved 66% success for typeset documents but only 38% for manuscripts when matching in their Locating London's Past project. However, Plewe (2003) argued that some uncertainty must pertain in the absence of independent verification.

The modern MMAL2, as a database of addresses, unit postcodes, easting/northing grid references, and associated information, has the potential for its individual, unique residential address records to be matched directly with those captured from the historical census documents provided that the structure of the address data field is identically formatted in each. However, there are a number of reasons why such matching might fail or be inaccurate as detailed below. It was not anticipated that all or even perhaps a substantial number of the historical census addresses would be geocoded by matching with the MMAL2 database, but it was expected that there would be a measure of success that varied between different types of local authority and populated area. Attempting to quantify the extent of this variation was an important objective of the research that would guide attempts to geocode historical census addresses in other parts of England and Wales. The urban landscape and morphology of the majority of London County including parks and open spaces were already in place by the first decade of the 20th century. This provided a legacy of buildings with addresses that could potentially be "rolled forwards" to match with modern address data. However, during the century, processes of regeneration and redevelopment, in part following destruction arising from enemy bombing during World War II, eroded parts of this urban heritage. In contrast, Middlesex was still mainly an area covered by countryside at the start of the 20th century comprising scattered settlements of various sizes interspersed with agricultural, extractive, and light industrial land use and employment typical of many such areas.

A four-stage method was developed to geocode the historical census addresses:

1. Attempt to match addresses transcribed from the 1901 and 1911 census records and stored in a standard format with identically formatted modern addresses obtained from MMAL2.
2. Identify unmatched addresses and correct anomalies arising from road name changes, inaccurate transcription, and so on, and repeat matching procedure.
3. Visualise matched addresses from Stages 1 and 2, locating these on historical topographic mapping before manually digitising new point features for unmatched addresses using ancillary information.
4. Merge geocoded addresses with thematic variables from historical censuses to enable analysis and aggregation to a range of geographies.

This method emerged following a series of iterative tests in Hammersmith and Hampton Wick, respectively representing the more and less densely populated areas in the set. This early experimental work enabled the other areas to be completed in a timely fashion as the main issues had been encountered by this stage, although each area and year presented its own specific challenges.

Points representing successfully matched historical census addresses during Stage 1 were viewed over the historical topographic maps, which enabled streets or individual apparently residential properties to be identified where matching had not occurred. Historical census addresses not geocoded during Stage 1 were identified as a residual list of unmatched addresses. Four main types of inconsistency occurred. Some thoroughfares were named identically on the historical map image and in the MMAL2 but were incorrectly transcribed from the census records. For example, some addresses in Zampa Road Deptford were inadvertently transcribed as Lampa Road from the handwritten census records, whereas the correct road name appears

on the historical OS map. A similar issue arose with thoroughfares and individual residential properties whose names included an abbreviation, most typically St. or St, which were both used as a shortened form for Saint. The MMAL2 invariably stores such abbreviations with a full stop, whereas such consistency was not present in the transcribed census addresses. Straightforward editing of the census address records corrected both these types of error once they had been detected. The majority of thoroughfares retained the same name over the 100-year period between the 1901 and 1911 Censuses and the compilation of the records in MMAL2. Changes in the names of London's thoroughfares can be determined from number sources (e.g., old to new street names, http://www.maps.thehunthouse.com/Streets/Old_to_New_Abolished_London_Street_Names.htm). Various terms are used to indicate a type of thoroughfare; examples are alley, close, crescent, gardens, grove, mews, road, and street; and a minor change was where one type had been replaced by another, but some naming information persisted (e.g., Clifton Road in Hammersmith became Clifton Avenue). Major changes were defined as the wholesale replacement of an historical thoroughfare name leaving no obvious clue to its former identity in the MMAL2 (e.g., High Road in Deptford transformed into Lewisham Way). These discrepancies were corrected by adding the new name for the same thoroughfare as an additional data field. Wholesale redevelopment of parts of the urban landscape resulted in an entirely different layout of thoroughfares and addresses, which might in some instances further confound the matching process by relocating a seemingly correctly matched address to a new position. Correction of these errors involved removal of incorrect geocodes. Once these data edits were completed, the first residual list of addresses were rematched with MMAL2, and those successfully geocoded at this stage were added to the first set.

Two further sources of data (street directories and local history society websites) and the geographical information embedded within the structure and sequencing of the census records themselves assisted with locating the historical census addresses, especially in resolving the "problem cases" (unmatched addresses). Street gazetteers and commercial directories were published for many cities and towns by the start of the 20th century. These can potentially help with geocoding historical census addresses by linking the names of small business proprietors with information on the census records. One such directory or almanack had been published for Wembley in Middlesex and detailed the sequence of not only commercial but also residential addresses and the occupier's name along thoroughfares (Wembley Urban District Council, 1906). The growing interest of communities in their past and the historical development of their settlements has led to a number of local history societies trawling through historical records and in some cases building openly accessible websites. Such a website has been created for one of the study areas, Hampton Wick (http://www.brickbybrick.org), which lists modern addresses along the streets in the town with images and the dates when they were built. It also includes street maps and links from the older addresses to the publicly available historical census records. Many of the historical census addresses in Hampton Wick used house names rather than numbers, and the maps on the local history society website allowed the modern house numbers to be connected with these names of Edwardian properties, therefore enabling matching with the MMAL2 records to be achieved. Although these supplementary data sources assisted with the geocoding, their presence arose fortuitously during the research rather than being a prerequisite for selecting areas; therefore, their availability here does not significantly challenge the transferability of the geocoding method to other local authorities. It would be entirely feasible to apply the geocoding method without the supplementary sources available here.

Another aid to geocoding the historical census addresses lay in the structure and geographical information embedded in the census records themselves. Each census record for 1901 and 1911 included an address and schedule number, which were recorded sequentially along individual thoroughfares in each ED. Descriptive definitions of the geography of the EDs are available in the historical census records and indicate that boundary lines typically followed the centre lines of roads, watercourses, and other clearly visible features. For example, ED 19 in the London City Registration District in 1901 was bounded by Finsbury Pavement, Finsbury Circus, Moorgate Place, and Moorfields, respectively, to the north, east, south, and west and contained "Moorgate St Buildings, Pete White's Alley, Gt Swan Alley, Little Swan Alley, Cross Key Court, Black Swan Alley, Moorgate Place, Moorgate Court, West Street, Finsbury Pavement, Finsbury Circus, Moorfields Nos. 2-52 (even Nos.), Short Street" (TNA, n.d., RG13/263). This description has been used to create a digital representation of ED 19 within St Stephen Coleman Street Ward, and ED 18 was created "by default" as it formed the rest of the ward (Figure 3). The schedule numbers of addresses on either side of streets forming the boundary were independently sequenced in respect of the ED in which they were located, whereas those for thoroughfares entirely within an ED possessed a sequential structure that could be detected on scrutiny of the data. This type of information was especially useful where gaps in the matching of census addresses with the MMAL2 occurred and the historical topographic maps revealed the presence of early-20th-century residential properties.

The residual set of nongeocoded historical census addresses were then captured by manual digitising using these supplementary information sources and the geographical detail embedded in the census records and by careful examination of the historical topographic maps to identify property names. Geocoding a proportion of the addresses in each area by successful matching with the MMAL2 enabled transfer of not only the grid coordinates but also the unit postcode to the historical census addresses. The unit postcode for the manually digitised addresses was assigned by allocation from the nearest known MMAL2 (unmatched) address. Having geocoded the census addresses, the remainder of the census data for households and individuals was attached, thus allowing analysis for these entities as well as aggregation to historical and modern spatial units (features) including 1901 and 1911 census EDs, thoroughfares, and the OAs used in the 2001 and 2011 Censuses.

4 | RESULTS AND ANALYSIS

This paper focuses on the method and results of geocoding addresses from historical British censuses rather than substantive findings arising

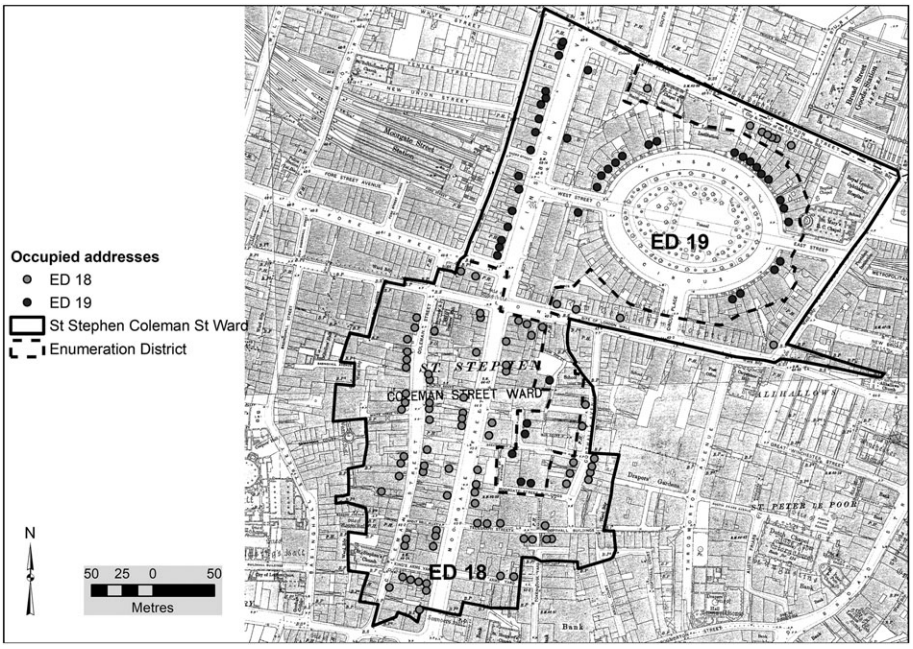

FIGURE 3 Enumeration Districts 18 and 19 in St. Stephen Coleman Street Ward within the London City Registration District, St. Botolph Registration Subdistrict in 1901 illustrating creation of digital boundaries for historical censuses overlain on topographic mapping.
Source: Ordnance Survey, Edina. © Crown copyright and/or database right 2018 OS

from the analysis of demographic and socio-economic topics capable of investigation with these data. Table 4 summarises the results of applying the geocoding method in the six boroughs/districts for 1901 and 1911. It is evident that there were differences in the geocoding success rate across these areas and between those in the more and less densely populated London and Middlesex counties. As indicated earlier, using a standardised format for both the historical census and MMAL2 addresses the geocoding procedure started by attempting to match the 1911 address records including those that were adjusted after Stage 1 before manually digitising unmatched addresses. The successful matches from 1911 helped with geocoding the 1901 addresses before comparing additional 1901 addresses with the MMAL2 and finally digitising any remaining 1901 addresses. Despite there being relatively small numbers of addresses from the 1901 and 1911 Censuses to be matched with the MMAL2 records in Hampton Wick and Hayes, the percentages linked and geocoded at the first pass were generally low in these areas. It is difficult to determine the precise reason why any individual address in the census records failed to be matched with one contained in the MMAL2, although Table 5 attempts to specify some of the identifiable reasons. The use of residential property names rather than numbers in Hampton Wick and Wembley accounted for 83% and 60% of the unmatched addresses, respectively, and this was also significant in Hayes, although it was unimportant in the London local authorities. The absence of a name or number property identifier on some roads was important in Hayes (20% of unmatched addresses), and here, the sequencing of census records was especially useful. Nonpermanent accommodation (vessels and sleeping rough) also accounted for small numbers of unmatched addresses in most areas. The difficulty of determining a reason for unmatched addresses in the London authorities is likely to be accounted for by factors such as redevelopment including World War II damage resulting in modern residential addresses not corresponding with thoroughfares that were inhabited at the turn of the 20th century and, in the case of the City of London CC, its relative decline as a residential area. During the 1960s and 1970s, the area around the Barbican was redeveloped from a

TABLE 4 Overview of the results of matching and geocoding methods for case study areas in Middlesex and London counties

	Addresses in population census records		First stage: geocoded after matching with MMAL2, *n* (%)		Second stage: geocoded after road/address checking and editing and matching with MMAL2, *n* (%)		Third stage: geocoded after digitising of unmatched addresses, *n* (%)	
	1901	1911	1901	1911	1901	1911	1901	1911
City of London CC	4,028	3,179	103 (2.6)	537 (16.9)	514 (12.8)	546 (17.2)	4,028 (100)	3,179 (100)
Deptford MB	16,287	16,503	4,384 (26.9)	5,234 (31.8)	4,686 (28.7)	5,558 (33.7)	16,287 (100)	16,503 (100)
Hammersmith MB	15,212	16,685	8,501 (55.9)	9,419 (56.5)	9,562 (62.9)	10,523 (63.1)	15,212 (100)	16,674 (100)
Hampton Wick UD	502	507	34 (6.8)	190 (37.4)	37 (7.3)	250 (49.3)	502 (100)	507 (100)
Hayes UD	544	849	26 (4.8)	128 (15.1)	26 (4.8)	130 (15.3)	544 (100)	849 (100)
Wembley UD/MB	788	2,184	76 (3.5)	1,079 (49.5)	76 (3.5)	1,102 (50.5)	788 (100)	2,182 (100)
Total	37,361	39,894	13,124 (35.1)	12,777 (32.0)	13,705 (36.7)	18,109 (45.4)	37,361 (100)	39,894 (100)

TABLE 5 Reasons for failing to match 1911 census addresses with contemporary address database in case study areas in Middlesex and London counties

	Property names replaced by numbers, *n* (%)	Absence of name or number, *n* (%)	Vessels, homeless, etc., *n* (%)	Not determined including redevelopment, *n* (%)
City of London CC	152 (5.8)	0 (0.0)	35 (1.3)	2446 (92.9)
Deptford MB	223 (2.0)	0 (0.0)	1 (0.0)	10721 (97.5)
Hammersmith MB	124 (2.0)	0 (0.0)	16 (0.3)	6022 (97.7)
Hampton Wick UD	213 (82.9)	3 (1.2)	0 (0.0)	41 (16.0)
Hayes UD	213 (29.6)	141 (19.6)	3 (0.4)	362 (50.3)
Wembley UD/MB	652 (60.3)	6 (0.6)	2 (0.2)	422 (39.0)
Total	1573 (7.2)	150 (4.8)	57 (0.2)	20014 (91.8)

Note. Estimates exclude mismatches arising from street name changes.

collection of residential streets, transport infrastructure, and commercial buildings into a group of modern multistorey residential blocks, a concert hall, theatre, the Guildhall School of Music, offices, and ancillary service facilities (see Figure 4). In 1911, relatively few of the addresses in the area were occupied as private households, but the redevelopment resulted in 1750 separate addresses in 2012.

The highest percentages of addresses that were matched and geocoded at the first stage in the procedure occurred in Hammersmith in 1901 and 1911, partly because the numbering of addresses had already started along roads in this rapidly developing area. However, the percentage of successful matches at Stage 1 was somewhat depressed in other areas by changes in the names of thoroughfares over the century leading to inconsistency. Overall, the percentage of successful matches at Stage 1 was higher in 1911 than 1901 and was the least successful in Hayes and the City of London. The notable improvement in matching addresses in Wembley was undoubtedly due to suburban residential development during the decade. Checking and editing of addresses to compensate and correct for transcription errors, road name changes, and abbreviations raised the number of successful matches overall by 1.6% in 1901 and 13.4% in 1911. The greatest improvement occurred in Hampton Wick in 1911 where the percentage of successful matches with the MMAL2 data increased from 37.4% to 49.3%. The third (digitisation) stage resulted in the capture of grid coordinates for all of the census addresses in each area by employing a combination of the techniques to locate missing addresses mentioned previously.

The full sets of geocoded addresses for the six study areas in 1901 and 1911, which are underpinned by the disaggregated household and individual census data records, are shown in Figure 5a,b. It reveals that some parts of the three Middlesex areas were starting to show evidence of the relatively regular suburban pattern of residential addresses along streets, which was already a common feature in Deptford and Hammersmith in 1901. Even at the scale shown in Figure 5a, comparison of the 1901 and 1911 patterns for Deptford and Hammersmith reveals evidence of residential development, for example, the south-west corner of Deptford MB and two areas on the western side of Hammersmith MB, having occurred between 1901 and 1911. In the latter case, the northern part of these areas corresponds with the enlarged inset shown in Figure 1. There is a more fragmented pattern evident in the City of London CC in 1901 and 1911, which reflects the mixed use of this local authority including financial, commercial, and trading industries alongside a declining residential population. Comparison of the geocoded addresses in Wembley (Figure 5b) also reveals some growth in the amount of housing, especially to the west of the town centre. There was some churn or turnover in the residential addresses occupied in 1901 and 1911 in each of these areas (see Table 6), although the geographical detail

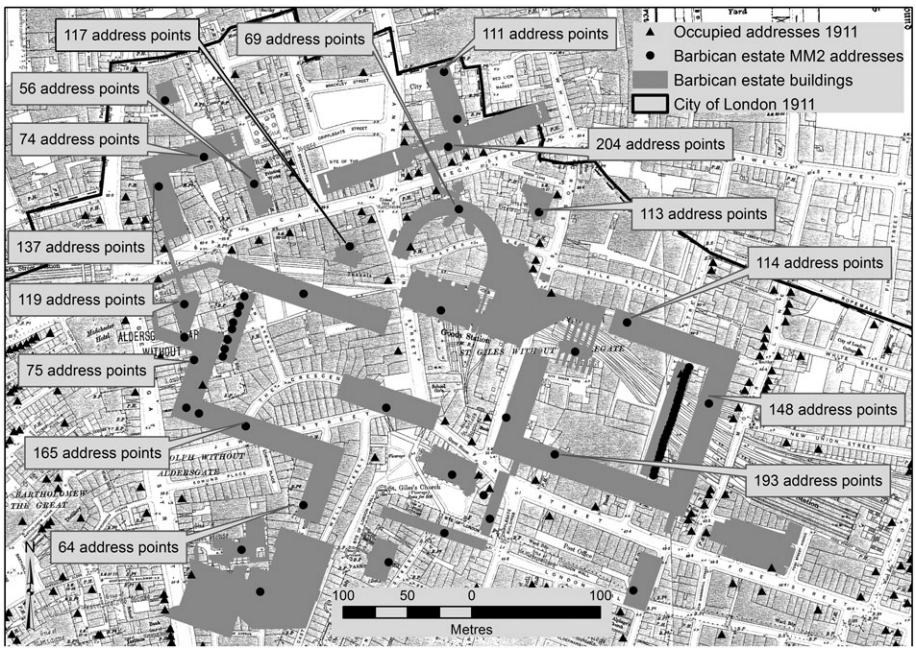

FIGURE 4 Example of historical nonresidential locality in the City of London CC redeveloped for residential purposes. *Source:* Ordnance Survey, Edina. © Crown copyright and/or database right 2018 OS

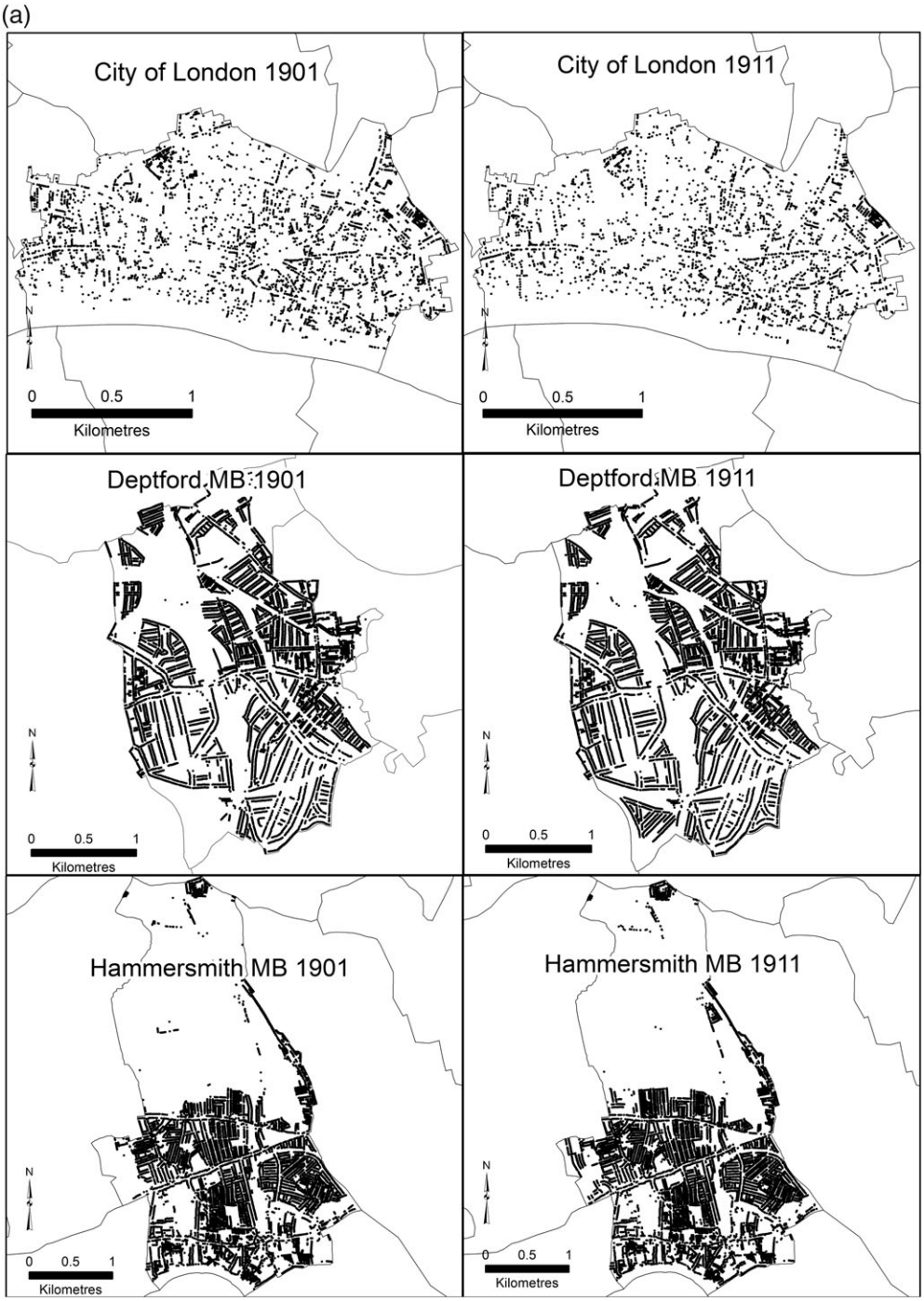

FIGURE 5 Geocoded historical 1901 and 1911 Census addresses for selected local authorities in London and Middlesex. (a) Full set of 1901 and 1911 addresses in selected districts in London County. (b) Full set of 1901 and 1911 addresses in selected districts in Middlesex County. *Source:* Ordnance Survey, Edina. © Crown copyright and/or database right 2018 OS

is not apparent at the "overview" scales used in Figure 5. Overall 26,311 addresses were occupied in the 1901 and 1911 censuses accounting for 70.4% and 66.0% of the respective totals. The highest figures occurred in Deptford and Hammersmith, and the lowest were in the three districts in Middlesex. Part of the difficulty in these areas were addresses imprecisely specified in the scanned paper records, which was especially notable in some of village and hamlet settlements where census records might simply refer to a household as living at Botwell, Hayes. These addresses were coded in the transcribed data as X01Xn and X11Xn where n is a numerical sequence from 01 to n for the imprecise addresses in the same locality or on one thoroughfare (e.g., X01X01 Botwell and X01X02 Botwell). High percentages of addresses (over 60.0%) were uniquely occupied in 1901 in the City of London, Hayes, and Wembley, with similar or even higher figures in the three Middlesex districts in 1911, which was connected with substantial increases in the number of addresses especially in Wembley.

These address points provide considerable flexibility as to how the associated census data might be reconstituted, first, to explore the demographic and socio-economic geography of these areas and

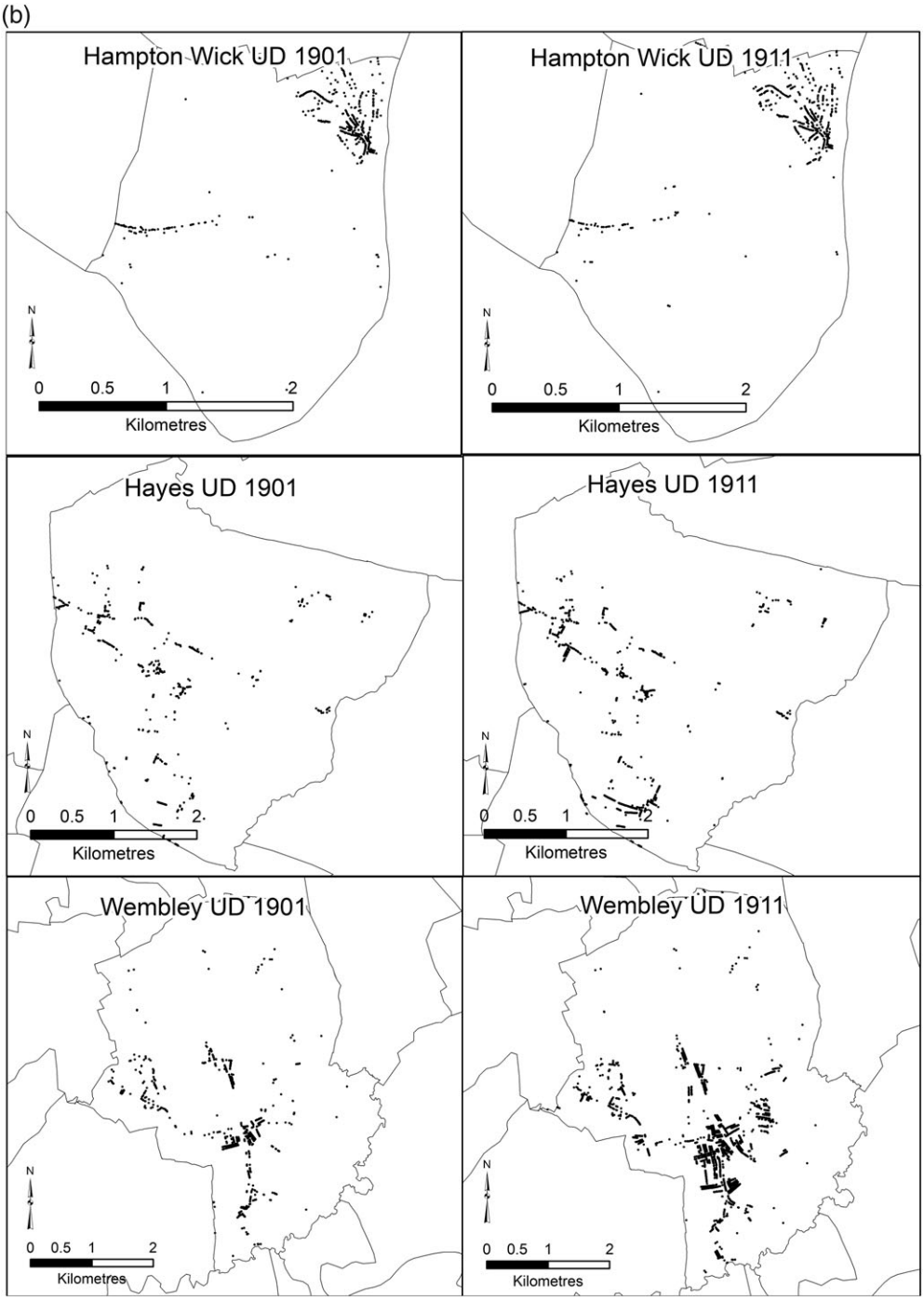

FIGURE 5 Continued.

the short-term changes 1901–1911 and, second, to allow analysis of change over the 100-year period between the first decades of the 20th and 21st centuries. It is beyond the scope of this paper to carry such analyses, and a brief review of what is feasible and the issues involved will suffice. EDs were linked to Census Registration Districts at the time of the 1901 and 1911 censuses, and these areas commonly spanned more than one local authority area, although EDs themselves were not split between the latter, whereas late-20th-century and early-21st-century small area census geographies (EDs and OAs) have been constrained to fit local government geography. Apart from summarising the historical census data for the 1901 and 1911 EDs, it is also possible to aggregate to those sections of thoroughfares with residential addresses and calculating population density along these linear features. This will allow more detailed analysis of demographic and socio-economic patterns than is possible with modern data aggregated to OAs based on unit postcodes even if the nonbuilt land is excluded. The opportunity to investigate the characteristics of individuals, households, and addresses is perhaps one of the most important benefits arising from the geocoding of historical census addresses. It is possible to aggregate the 1901 and 1911 census data to contemporary census units and to create cross-tabulations similar to those available from modern censuses. Inevitably, there will be complications in making comparisons between 1901–1911 and 2001–2011, for example,

TABLE 6 Turnover in occupation of residential addresses between 1901 and 1911 in case study areas in Middlesex and London counties

	Addresses occupied in 1901 and 1911	Imprecise addresses 1901, *n* (%)	Addresses occupied in 1901 but not in 1911, *n* (%)	Imprecise addresses 1911, *n* (%)	Addresses occupied in 1911 but not in 1901, *n* (%)
City of London CC	1,514 (37.6% in 1901; 47.6% in 1911)	17 (0.4)	2,514 (62.0)	0 (0.0)	1,665 (52.4)
Deptford MB	13,384 (82.2% in 1901; 81.1% in 1911)	0 (0.0)	2,903 (17.8)	0 (0.0)	3,119 (20.1)
Hammersmith MB	11,171 (73.4% in 1901; 67.0% in 1911)	0 (0.0)	4,041 (26.3)	0 (0.0)	5,503 (33.0)
Hampton Wick UD	82 (16.3% in 1901; 16.2% in 1911)	4 (0.8)	416 (83.7)	31 (6.1)	394 (77.7)
Hayes UD	47 (8.6% in 1901; 5.5% in 1911)	139 (16.3)	358 (65.8)	141 (25.9)	661 (77.9)
Wembley UD/MB	113 (14.3% in 1901; 5.2% in 1911)	366 (46.4)	309 (39.2)	9 (0.4)	2,060 (94.4)
Total	26,311 (70.4% in 1901; 66.0% in 1911)	526 (1.4)	10,541 (28.2)	181 (0.5)	13.402 (33.6)

Note. Imprecise addresses removed when calculating number and percentage of addresses uniquely occupied in 1901 or 1911.

when comparing the percentage of the population who were born in the United Kingdom over these decades. In the earlier period, the United Kingdom included the whole of the island of Ireland prior to the creation of the Republic of Ireland in 1922.

5 | CONCLUSIONS

The progressive assimilation of digital data and information and communication technologies into academic research and people's everyday lives raises the challenge of searching out historical data sources that can be similarly coerced into contemporary information retrieval procedures. The digital capture of the British Population Census documents has not only fuelled enthusiasm for genealogical research but also prompted researchers from different disciplines to reopen the debate about topics of social and economic history and prompted geographers to explore the underlying spatial relationships contained in these historical records. Rather than address specific substantive research questions, this paper has sought to develop and test a method for geocoding the address, household, and individual records from the 1901 and 1911 census records for areas in the former counties of London and Middlesex. Six local authority areas were selected, three from each of these counties, with the intention of assessing the geocoding method in respect of different types of locality. These included essentially "semirural" areas that contained small town and village settlements alongside dispersed dwellings and farms as well as areas within London that had been developed for housing in the late 19th century and were continuing to "suburbanise" during the first decade of the 20th century. The City of London was also included, which contained a significant mainly scattered residential population in the Edwardian era that subsequently decreased as the 20th century unfolded, only to experience some modest increase as city centre living returned in recent times.

Geocoding historical records at its most basic level is concerned with accurately attaching coordinate grid references to the entities contained in these sources such that they are capable of being mapped and analysed spatially within GIS software. The method developed for geocoding the historical census records reported here sought to investigate the opportunity for "borrowing" such grid references from contemporary geospatial data sources in conjunction with other ancillary data. In particular use was made of the OS MMAL2 data, although the method could be adapted to use other address databases, together with historical topographic mapping and geographical details embedded within the census records themselves. The resultant hybrid method combined address matching with digitisation of unmatched records in order to produce fully geocoded address, household, and individual data sets of the six selected areas. The results presented here have shown that the scope for automating the process of geocoding depends on a number of factors. Successful matching between 1911 census addresses and contemporary ones was achieved in over 50% of cases in some areas (Hammersmith and Wembley), whereas others were considerably less successful, at less than 15% (City of London CC and Hayes).

A number of factors accounted for this variability. First, areas where early-20th-century addresses, compared with areas where dwelling names predominated, were composed of a number and street name resulted in a higher level of success in matching. Second, successfully matched addresses in 1911 assisted with the process of matching the 1901 addresses; therefore, working back through the census records offers a higher chance of success. Third, even in areas where property numbers were prevalent, unmatched addresses could arise because of street name changes and incorrect or inconsistent transcription of the census records. Fourth, once a "residue" of unmatched addresses from the census records remained, it was possible to use geographical information embedded within the census records, such as the schedule number, to identify the correct sequence of addresses along individual thoroughfares after successfully matching a small number of dwellings with name and street name from the historical OS maps. Fifth, the set of six areas included two in which ancillary data sources, a local history website in Hampton Wick and Street Almanack in Wembley, fortuitously helped to match some of the addresses. The potential to transfer the geocoding method outlined here to other areas and to substitute different contemporary grid referenced address databases relates to the prevalence of the factors that contributed a match not occurring. In general, a lower degree of matching was achieved in areas with a

dispersed population in small settlements or clusters of relatively isolated properties (e.g., Hayes in this case), whereas more successful matching occurred in areas that had already experienced relatively intensive, possibly late Victorian suburbanisation (e.g., Deptford and Hammersmith). The prospect of applying the geocoding method to the 1921 census records when they are released in 2021 would appear feasible, although going further back in time, the late-19th-century census records (e.g., 1881 and 1891), although not untenable, might achieve lower levels of address matching if address naming rather than numbering prevailed.

The potential benefits of exploring the microscale demographic and socio-economic changes that took place during the first decade of the 20th century and of making connections through to the equivalent decade at the start of the 21st century with respect to comparable areas are considerable. Three areas of investigation with the potential to advance our understanding of population geography of the early-20th-century relate to geographical variation in household composition and place of birth and to the development of flexible spatial boundary systems based on household and address rather than postcode characteristics. It is already evident from informal examination of the census records that spatial clustering of different types of household, multioccupancy dwellings, and people born in a range of other countries exists within the six London and Middlesex areas. The intention is for further analysis of the geocoded census records for these areas, possibly with modest extension to one or two other London and Middlesex local authorities and making the data available to other researchers. The flexibility offered by geocoded addresses presents opportunities for other types of aggregation including the inhabited sections of thoroughfares, historical EDs, and residential land parcels and of drilling down to addresses to investigate household composition and structure. It is to be hoped that the possibility of researchers in the future undertaking similar address matching in the early decades of the 22nd century will not be inhibited by a failure to retain and then release household and individual records for those people living at addresses in the United Kingdom at the present time.

ACKNOWLEDGEMENTS

Office for National Statistics, 2011 Census: Aggregate data (England and Wales) [computer file]. U.K. Data Service Census Support. Downloaded from http://infuse.mimas.ac.uk. These data are licensed under the terms of the Open Government Licence (http://www.nationalarchives.gov.uk/doc/open-government-licence/version/2). Office for National Statistics, 2011 Census: Aggregate data (England and Wales) [computer file]. U.K. Data Service Census Support. Downloaded from http://infuse.mimas.ac.uk. These data are licensed under the terms of the Open Government Licence (http://www.nationalarchives.gov.uk/doc/open-government-licence/version/2).

Office for National Statistics, 2001 Census: Aggregate data (England and Wales) [computer file]. U.K. Data Service Census Support. Downloaded from: http://infuse.mimas.ac.uk. These data are licensed under the terms of the Open Government Licence (http://www.nationalarchives.gov.uk/doc/open-government-licence/version/2). This work is based on data provided through EDINA UKBORDERS with the support of the ESRC and JISC and uses boundary material that is copyright of the Crown. It also uses data provided under licence by the OS. The work on which this article is based was carried out under a British Academy, Senior Research Fellowship (SF110183) entitled *The Geography of London's Population: Populating Places with Historical Census Statistics*. The author is grateful to two anonymous referees who provided helpful comments on an earlier version of this article.

OS MasterMap Address Layer 2 was used under Academic Research Licence. Ordnance Survey ©Crown Copyright. All rights reserved.

CONFLICT OF INTEREST

There is no known conflict of interest.

ORCID

Nigel Stephen Walford https://orcid.org/0000-0002-9182-0252

REFERENCES

Anderson, M. (1972). Standard tabulation procedures for the census enumerators' books 1851–1891. In E. A. Wrigley (Ed.), *Nineteenth-century society* (pp. 134–145). Cambridge University Press. https://doi.org/10.1017/CBO9780511896118.005

Anderson, M. (1987). *National sample from the 1851 Census of Great Britain: Introductory user guide*. Dept. of economic and social history, University of Edinburgh, Edinburgh, September.

Anderson, M. & Collins, B. (1973). *National sample from the 1851 Census of Great Britain: Sample procedures: Background paper 2*, Department of Sociology, University of Edinburgh, Edinburgh, July.

Anderson, M., Collins, B., & Stott, C. (1977). The national sample from the 1851 Census of Great Britain: Sampling and data handling procedures. *Urban History*, *4*, 55–59. https://doi.org/10.1017/S0963926800002522

Boyle, P., & Shen, J. (1997). Public housing and migration: A multi-level modelling approach. *Population, Space and Place*, *3*(3), 227–242. https://doi.org/10.1002/(SICI)1099-1220(199709)3:3<227::AID-IJPG69>3.0.CO;2-W

Caruthers, A. W. (1985). Mapping the population census of Scotland. *The Cartographic Journal*, *22*(2), 83–87. https://doi.org/10.1179/caj.1985.22.2.83

Champion, A. G., Wong, C., Rooke, A., Dorling, D., Coombes, M., & Brunsdon, C. (1996). *The population of Britain in the 1990s: A social and economic atlas*Clarendon Press.

Dorling, D. (1994). Cartograms for visualizing human geography. In D. Unwin, & H. Hearnshaw (Eds.), *Visualization and GIS. Belhaven Press* (pp. 85–102).

Eicher, C., & Brewer, C. (2001). Dasymetric mapping and areal interpolation: Implementation and evaluation. *Cartography and Geographic Information Science*, *28*(2), 125–138. https://doi.org/10.1559/152304001782173727

Gaits, G. M. (1969). Thematic mapping by computer. *The Cartographic Journal*, *6*(1), 50–68. https://doi.org/10.1179/caj.1969.6.1.50

Gregory, I. N. (2002). The accuracy of areal Interpolation techniques: Standardizing 19th and 20th Century census data to allow long-term comparisons. *Computers, Environment and Urban Systems*, *26*(4), 293–314. https://doi.org/10.1016/S0198-9715(01)00013-8

Gregory, I. N., Bennett, C., Gilham, V. L., & Southall, H. R. (2002). The Great Britain Historical GIS Project: From maps to changing human geography. *The Cartographic Journal*, *39*(1), 37–49. https://doi.org/10.1179/000870402787288045

Gregory, I. N., Dorling, D., & Southall, H. R. (2001). A century of inequality in England and Wales using standardized geographical units. *Area*, *33*(3), 297–311. https://doi.org/10.1111/1475-4762.00033

Harley, J. B. (1975). Ordnance Survey maps: A descriptive manual. Ordnance Survey, Southampton.

Higgs, E., Jones, R., Schürer, K. & Wilkinson, A. (2013). Integrated Census Microdata (I-CEM) Guide. Available from: http://www.essex.ac.uk/history/research/ICeM/documents/icem_guide.pdf. Accessed 29 January 2016.

Hinde, P. R. A. (1985). Household structure, marriage and the institution of service in nineteenth-century rural England. *Local Population Studies*, *35*, 43–51.

Hitchcock, T., Shoemaker, R., Emsley, C., Howard, S. & McLaughlin, J., et al. (2015). The Old Bailey Proceedings Online, 1674–1913. Available at: www.oldbaileyonline.org, version 7.2 [Accessed: 16 May 2017]

Logan, J. R., Jindrich, J., Shin, H., & Zhang, W. (2011). Mapping America in 1880: The urban transition historical GIS project. *Historical Methods: A Journal of Quantitative and Interdisciplinary History*, *44*(1), 49–60. https://doi.org/10.1080/01615440.2010.517509

Marten, E. J. (1971). The 1971 Canadian census data access programme. *Population Index*, *38*(4), 424–427. https://doi.org/10.2307/2733685.

Marx, R. W. (1986). The TIGER system: Automating the geographic structure of the United States census. *Government Publications Review*, *13*(2), 181–201. https://doi.org/10.1016/0277-9390(86)90003-8

Mennis, J., & Hultgren, T. (2006). Intelligent dasymetric mapping and application to areal interpolation. *Cartography and Geographic Information Science*, *33*, 179–194. https://doi.org/10.1559/152304006779077309

Navickas, K. (2016). Political meetings mapper with British library labs: Mapping the origins of British democratic movements with text-mining, NLP, geo-parsing and crowd-sourcing. Institute of Historical Research Digital History Seminar. Available at https://www.youtube.com/watch?v=zgaUkTn-T-E. [Accessed 16 May 2017]

Norman, P., Rees, P., & Boyle, P. (2003). Achieving data compatibility over space and time: Creating consistent geographical zones. *International Journal of Population Geography*, *9*(5), 365–386. https://doi.org/10.1002/ijpg.294

Office for National Statistics. (2011). Census transformation programme. Available at: https://www.ons.gov.uk/census/censustransformationprogramme [Accessed 24 September 2018]

Plewe, B. (2003). Representing datum-level uncertainty in historical GIS. *Cartography and Geography Information Science*, *30*(4), 319–334. https://doi.org/10.1559/152304003322606229

Scholossberg, M. (2003). GIS, the US Census and neighbourhood scale analysis. *Planning, Practice and Research*, *18*(2–3), 213–217. https://doi.org/10.1080/0269745032000168269

Schurer, K. & Higgs, E. (2014). Integrated Census Microdata (I-CeM), 1851-1911. [data collection]. UK Data Service. Sn: 7481, http://doi.org/10.5255/UKDA-SN-7481-1

Shepard, D. S. (1984). Computer mapping: The SYMAP interpolation algorithm. In G. L. Gaile, & C. Willmorr (Eds.), *Spatial statistics and models* (pp. 133–145). Netherlands: Springer. https://doi.org/10.1007/978-94-017-3048-8_7

Shouls, S., Congdon, P., & Curtis, S. (1996). Modelling inequality in reported long term illness in the UK: Combining individual and area characteristics. *Journal of Epidemiology and Health*, *50*(3), 366–376. https://doi.org/10.1136/jech.50.3.366

Southall, H. R. (2003). A vision of Britain through time: Making long-run statistics of inequality accessible to all. *Radical Statistics*, *82*, 26–43.

Southall, H. R. (2006). Electronic resources for local population studies: A vision of Britain through time: Making sense of 200 years of census reports. *Local Population Studies*, *76*, 76–89.

Southall, H. R. (2014). Rebuilding the Great Britain Historical GIS, Part 3: Integrating qualitative content for a sense of place. *Historical Methods: A Journal of Quantitative and Interdisciplinary History*, *47*(1), 31–44. https://doi.org/10.1080/01615440.2012.664101

St-Hilaire, M., Moldofsky, B., Richard, L., & Beaudry, M. (2010). Geocoding and mapping historical census data: The geographical component of the Canadian Century Research Infrastructure. *Historical Methods: A Journal of Quantitative and Interdisciplinary History*, *40*(2), 76–91. https://doi.org/10.3200/HMTS.40.2.76-91

Stilwell, J. (2005) Providing access to census-based interaction data in the UK: That's WCID! Proceedings of 3rd international conference on computing, Communications and Control Technologies.

The National Archives (n.d.) Census of England and Wales, 1901, Description of Enumeration District, Document Reference RG 13/263.

Tilley, P., & French, C. (1997). Record linkage for nineteenth-century census returns: automatic or computer aided? *History and Computing*, *9*(1–3), 122–133. https://doi.org/10.3366/hac.1997.9.1-3.122

Walford, N. S. (2013). Development and design of a web-based interface to estimate population statistics for consistent spatial units across recent British censuses. *Environment and Planning a*, *45*(7), 1713–1733. https://doi.org/10.1068/a45377

Walker, K. (2016). tigris: An R Package to access and work with geographic data from the US Census Bureau. *R Journal*, *8*(2), 231–242. https://journal.r-project.org/archive/2016/RJ-2016-043/index.html

Wembley Urban District Council (1906). Wembley Directory and Almanack for 1906, Wembley Urban District Council, Wembley. Pp. 58-77.

Xu, H., Logan, J. R., & Short, S. E. (2014). Integrating space with place in health research: A multilevel spatial investigation using child mortality in 1880 Newark, New Jersey. *Demography*, *51*(3), 811–834. https://doi.org/10.1007/s13524-014-0292-y

How to cite this article: Walford NS. Bringing historical British Population Census records into the 21st century: A method for geocoding households and individuals at their early-20th-century addresses. *Popul Space Place*. 2019;25:e2227. https://doi.org/10.1002/psp.2227