International Journal of Geo-Information

MDPI

Article

Geo-Referencing and Mapping 1901 Census Addresses for England and Wales

Tian Lan * and Paul Longley

Department of Geography, University College London, London, WC1E 6BT, UK
* Correspondence: tian.t.lan@ucl.ac.uk

Received: 21 May 2019; Accepted: 20 July 2019; Published: 24 July 2019

Abstract: Geocoding historical addresses is a primary yet nontrivial application of spatial analysis in historical geographic information systems (GIS) and spatial humanities. We demonstrate our endeavours of geo-referencing and visualising historical census addresses in England and Wales, by matching the residential addresses to a historical gazetteer and a contemporary address database of Britain. The results indicate that it is feasible to standardise and geocode a large share of unique addresses from the historical database. The historical gazetteer and the modern address registers are two complementary data assets that can be used to geo-reference both well-formatted addresses in urban areas and non-standard addresses such as place names or building names in rural areas. The geo-referenced historical census data open up new opportunities for a broad spectrum of geo-demographic research on historical population characteristics at the micro level in England and Wales.

Keywords: historical censuses; geocoding; address matching; historical geo-demographics; spatial humanities

1. Introduction

Historical censuses, first carried out in the British Isles in 1801, preserve snapshots of population size and characteristics. They are highly valued sources of information to understand the population and social structure in the past and their transition over time and are today crucial in historical GIS and digital spatial humanities. Many efforts in different countries have been made to digitise historical census records, which were initiated by genealogical researchers, organisations, and companies (e.g. FamilySearch (https://www.familysearch.org) and Findmypast (https://www.findmypast.co.uk)) in the past decades. For instance, the Censuses for Britain 1851–1911 [1], the United States enumerations 1850–1940 [2], and the 1911–1951 Canadian censuses [3]. These digital copies of historical censuses have opened up new opportunities for historians, historical geographers, and other social scientists to investigate migration, health, economic activity, and social mobility in the nineteenth and early twentieth centuries [2,4–8].

Historical census records often contain explicit spatial attributes, such as enumeration unit names (e.g., registration district, parish, etc.), home addresses, or birthplace names, which cannot be directly imported into GIS. In order to incorporate map visualisation and spatial analytics into socio-economic studies (see Longley and Singleton [9] and Singleton and Longley [10]), geo-referencing techniques are required to transform textual descriptions of locations and places into geographical coordinates. There exists a variety of modern geo-referencing Application Programming Interfaces (APIs) supplied by map data and service providers, for example, Google Maps (maps.google.com) and Open Street Map (www.openstreetmap.org). Each works very effectively with current postal addresses, which are in standardised and hierarchical formats. However, historical census addresses are structured much less clearly. Particularly in rural areas, household addresses can simply be place names or building

names such as "acre farm", "butcher's shop" or "clock inn", which bear neither house numbers nor thoroughfare names. In addition, some common street names across the country, for instance, "Church Street", "High Street", or "Station Street", create additional ambiguity for the geo-referencing process. Such uncertain and complex scenarios have posed significant challenges to geocoding historical census records and have impeded the spatial analysis of census data.

It is, therefore, a non-trivial task to geocode micro level census data before we can use them to investigate geo-demographics and internal migration of the late Victorian population. The main aim of this paper is to develop an automatic method of geo-referencing the household addresses from the 1901 Census in England and Wales. We demonstrate a fuzzy address matching method using two complementary address corpora and visualise the geocoded addresses in a series of maps. The results indicate that it is feasible to standardise and geocode a large share of unique addresses from the historical censuses at a national scale.

2. Related Work on Geo-referencing Historical Data

There are several significant contributions to geo-referencing practices in historical GIS [11–13] projects, which broadly fall into two methodological paradigms. The first strategy is to digitise historical locations, streets, or enumeration units directly on geo-referenced historical maps, in either a manual or semi-automatic way. Gregory et al. [8] have digitised administrative boundary maps in the UK from 1906 to 1910 and used the results to develop a series of thematic maps, for example, the mortality rate from lung disease by registration districts. To extract locations and social classifications of households in historical London, Orford et al. [14] have manually digitised about 120,000 points from the Charles Booth's poverty map. Logan et al. [4] have geo-referenced the addresses from the 1880 U.S. Census by editing the U.S. TIGER (Census Bureau's Topologically Integrated Geographic Encoding and Referencing System) files. Combining both modern census geography and various historical data sources, St-Hilaire et al. [3] have reconstructed the census subdivision polygons for the 1911–1951 Canadian censuses and associated other census variables to these polygons. The advent of crowdsourcing and web-GIS techniques has enabled public engaged geocoding practices. With the assistance of a web-based volunteered geographic information (VGI) system, Southall et al. [15] have created a historical gazetteer of Great Britain for the c. 1900s by pinpointing text annotations of jurisdictions, places, streets and other points of interest using a series of geo-referenced Ordnance Survey (OS) maps. Likewise, Cura et al. [16] have established an online VGI geocoding system of historical Paris based on their Historical Geocoder, to collaboratively enhance geocoding results of computer vision and machine learning algorithms. There are other geocoding packages similar to the Historical Geocoder, for example, Pelias and Nominatim.

Another strategy is to run text-based address matching between historical data sources and address databases that have already been geo-referenced. The most common geo-referencing method is to associate census attributes to specific enumeration geographies for mapping and analysing purposes. For instance, Carrion et al. [17] create a spatial database for medieval fiscal data in Italy by matching the place names to present-day geographies. Clough et al. [18] summarise several sources of existing geocoded data assets for linking the UK National Archive's data to geographical locations. There are also cases which geo-reference historical data at the record level rather than at the aggregated level. Daras et al. [19] have presented their results of geocoding 24 million births, marriages, and deaths records from 1855 to 1974 in the Digitising Scotland project. Walford [20] has devised a semi-automated geocoding method for addresses in six pilot study areas within the modern Greater London Authority area based upon 1901 and 1911 Census data. Lansley et al. [21] have linked the addresses from twenty years' residential addresses from consumer registers to the OS AddressBase product. These various projects demonstrate that a considerable amount of historical addresses can be matched to modern address databases, despite the fact that entire areas, as well as individual properties, may have been redeveloped over the intervening years.

The two strategies each have their advantages and disadvantages. Digitising and geo-referencing addresses on historical maps usually achieve higher spatial consistency, notwithstanding the cost of intensive human labour. VGI solutions mitigate costs. In contrast, the address matching strategy enables automated geo-referencing processes, suitable for large numbers of records, as with geocoding of entire nation census records. However, the address matching process is vulnerable to the quality of the address strings that may be ambiguous or error-prone, for example, when street numbers are changed or re-sequenced over time. In this paper, we aim to geo-reference millions of historical census addresses at the micro level for further spatial analysis, without the commitment of resources required in digital encoding from historical paper sources. Where available, we utilise existing assets of geocoded address corpora alongside the primary strategy of linking historical census addresses to existing databases.

3. Data and Methods

Twelve digitally encoded Great Britain census datasets were obtained from the UK Data Service (UKDS) under a special licence agreement, pertaining to England and Wales or Scotland over the period 1851–1911. The data are enhanced versions of the original Census return transcriptions, enriched by the Integrated Census Microdata project (I-CeM) [1]. In this paper, we take only the 1901 Census in England and Wales as an example to illustrate our geo-referencing process. There were 32,493,318 individuals enumerated in the 1901 Census in England and Wales. Residential addresses were reported in the unit of a household or an institution. Table 1 presents some address instances from the 1901 Census. Column "RECID" stores the unique record identifiers while "CONPARID" is the ID of the parish to which the record belongs. Some of the addresses in urban parishes are clearly formatted into street number(s) and thoroughfare names, such as "10 Gower Street" or "10 Cardiff Road". By contrast, others are more descriptive, such as the "house on police station". In rural parishes, some addresses are simply place names, for example "old mill". Many addresses in Wales are recorded in Welsh, as illustrated in Table 1.

Table 1. An illustrative sample of addresses in England and Wales as recorded in the 1901 Census.

RECID	Address	CONPARID
20661567	10 gower street	108040
1535692	100 & 102 mildmay grove	100002
20661576	12 / 14 gower street	108040
2282943	100 old ford road bonner road	100002
8488188	post office high street	100002
1932358	furniture shop 37 eagle street	100002
2389381	married quarter of police station 72, 74, 76, 78 leman street	100002
8496253	house on police station	100002
30628062	10 cardiff road	110645
30628201	2 court hill	110645
30628509	old mill	110645
30628524	tyla morris lodge	110645
30628159	llwyn yr eas	110645
30628179	llanmeas	110645
30628166	tyny caean	110645

In addition to the previously mentioned challenges, common street names pose another issue for address matching. To explore the diversity of street names in the Census address, we extract street names and summarise their frequency distribution. The 10 most frequently used street names in the 1901 Census are High Street, Church Street, Queen Street, George Street, King Street, Victoria Street, West Street, Station Road, Chapel Street, and Victoria Road. These reflect interesting features in the nomenclature of streets in the Nineteenth Century. For instance, to commemorate the monarchy, streets across the country were often named after the monarch or Royal Family. There are 703,263 unique

street names across England and Wales. Figure 1 plots the frequency distribution of the most popular 500 street names, and exhibits a truncated long tail pattern. Beyond the frequent occurrences, such as the 34,145 High Street and 15,874 Church Street addresses, there is a long tail of unique occurrences such as "furniture shop 37 eagle street".

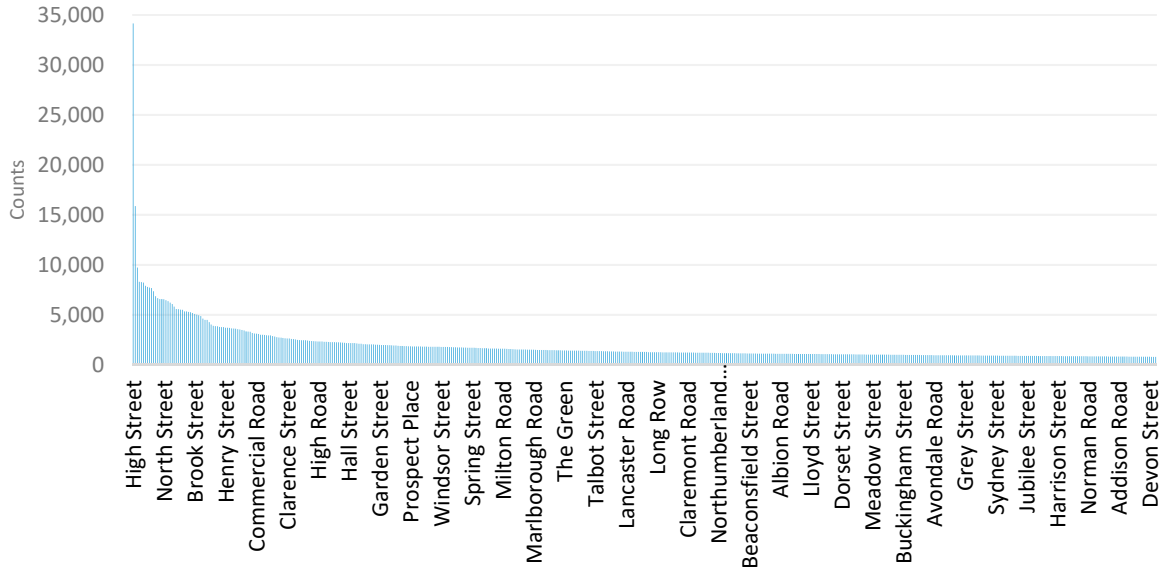


Figure 1. Frequency distribution of the 500 most frequently occurring street names.

Besides the historical Census addresses, we have two corpora of geocoded addresses: the contemporary OS AddressBase Premium (https://www.ordnancesurvey.co.uk/business-and-government/products/addressbase-premium.html) and the historical GB1900 Gazetteer (http://www.visionofbritain.org.uk/data/). OS AddressBase is the most comprehensive address database maintained for Great Britain, comprising 28 million Royal Mail postal delivery addresses plus inputs from the planning system maintained by local authorities. These addresses are consistently structured and are geocoded, providing a reliable spine against which historical address data may be matched. However, many historical addresses are messy, particularly in rural areas where addresses are frequently reported as place names, street names, or farm names, often in non-standardised, non-hierarchical formats. Moreover, past place names or street names may be altered or fall into disuse. In an attempt to identify and accommodate such instances, we incorporate the historical GB1900 Gazetteer. This gazetteer of street and place names is the outcome of transcription and geocoding of scanned OS County Series covering Great Britain from the early 1900s, using the contributions of thousands of volunteers over years in the GBHGIS project [15,22]. There are c. 2.6 million geo-referenced text strings, which have been transcribed and confirmed then by different transcribers to guarantee the data quality. This is likely to be the largest open historical gazetteer online so far [15] and can be accessed via the website 'A Vision of Britain through Time' (http://www.visionofbritain.org.uk/data/).

Figure 2 shows the flow of the two-stage data processing in this paper. We take the addresses from the 1901 Census in England and Wales, the OS AddressBase, and the GB1900 Gazetteer as the data input. We also assign historical parish identifiers to addresses in both the OS AddressBase and the GB1900 Gazetteer, by spatially joining them to historical parish boundary data, provided by the Cambridge Group for the History of Population and Social Structure (https://www.campop.geog.cam.ac.uk/). We adopt a geo-blocking strategy, searching only for candidate address pairs within the same parishes from the historical addresses and the geocoded addresses in the two address corpora. There are 4,872,707 address strings in total that are unique within each parish in England and Wales. We thus go some way towards resolving multiple occurrences of common street names and reduce the high computation cost of fuzzy string matching.

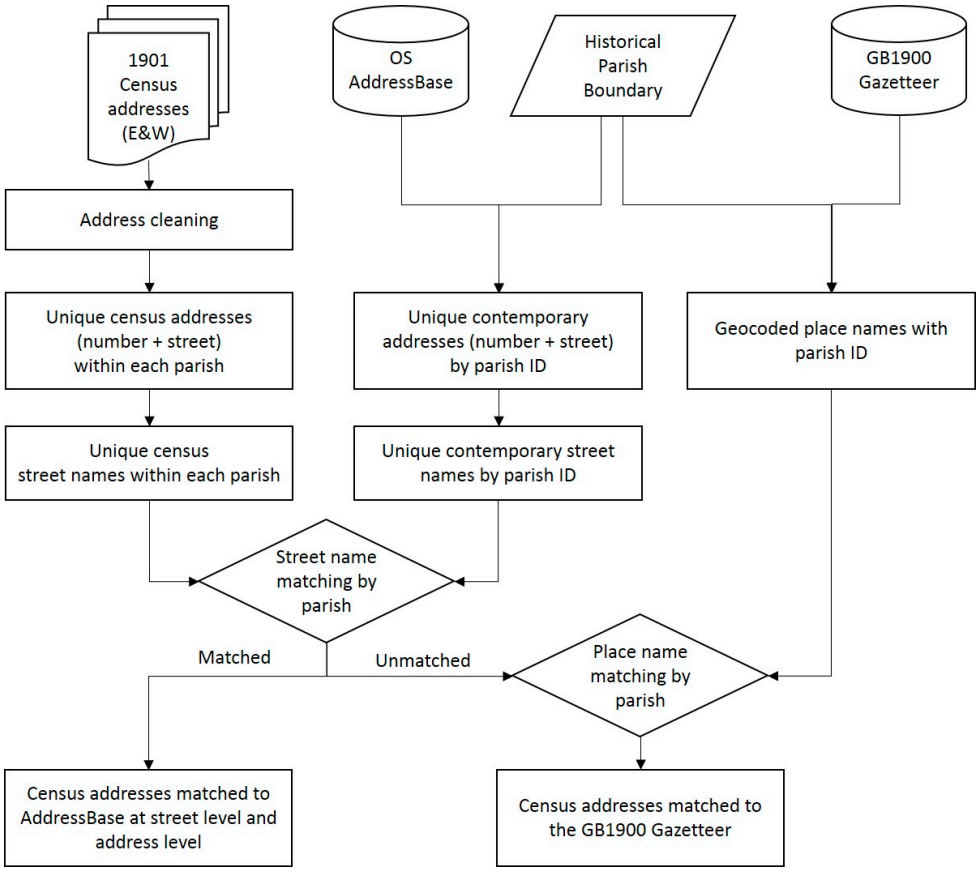


Figure 2. The workflow of the geo-referencing process.

After cleaning the historical addresses, we extract unique address strings within each parish from both the historical addresses and the OS AddressBase. We further split these address strings into street numbers and thoroughfare names in order to get unique street names by parish. Taking advantage of the open source package Fuzzywuzzy (https://github.com/seatgeek/fuzzywuzzy), we link the historical street names to their most probable matches within the relevant parish found in AddressBase. We then check whether street numbers can be matched in addition to street names. A historical address is considered matched at the address level if both street name and number are found in the same parish in AddressBase. Likewise, an address is considered as being matched only at the street level if naming or numbering is inconsistent in the same parish. Whilst this procedure does not accommodate street re-numbering, the results are deemed sufficiently consistent for our present purposes.

For historical addresses that are not matched at either level, we implement a parish-level fuzzy matching procedure as an additional stage. It is worth noting that the encoding of the address strings from the historical censuses and the geocoded address corpus should be unified in advance to facilitate the fuzzy string matching, since address strings in Welsh sometimes appear in Unicode, which may inhibit string comparisons.

4. Results

4.1. Overall Matching Performance

In accordance with the workflow, the geo-referencing results are evaluated in terms of the overall match rate, possible bias, the geographical variation in the match rate, and using a pilot study of the matching results at the address level. We first show some examples of the matched and unmatched addresses grouped by their match sources in the last column of Table 2. Here, the addresses matched to the OS AddressBase, the first three are matched at the address level while the others in that group

are matched only at the street level. There are instances of place names such as "twyn gwyn farm" and "big edge hill" linked to the historical gazetteer. We also find unmatched records which perhaps are either too vague or might have been demolished to be matched by our method.

Table 2. Examples of matched and unmatched addresses.

Record ID	Historical Addresses in the 1901 Census	Matches	Match Source
32307703	3 mount pleasant	3 mount pleasant	OS AddressBase
32306195	llwydiart arms 2 thomas street tanyard s row	2 thomas street	
32306970	wolverhampton house 81 & 83 market street	83 market street	
32309400	bont 10 summer hill	1 summer hill	
32305356	ty mawr 13 pump street	1 pump street	
32308153	pen terfyn 10 british terrace	1 british terrace	
30455564	american gardens penygarn	American Gardens	GB1900 Gazetteer
30410529	twyn gwyn farm	Twyn-gwyn Farm	
30466550	big edge hill	Big Edgehill	
30466849	big pond	N/A	N/A
22825269	blue boar hotel market place		
4243812	st georges villas 1		

Table 3 summarises both the numbers and percentages of addresses that are matched at different geo-referencing stages. Overall, approximately 85% of unique address strings in England and Wales are matched, of which c. 63% are matched to the contemporary OS AddressBase at either address level (25%) or street level (an additional 38%). A further 22% of the addresses are matched using the historical GB1900 Gazetteer. We further break down the statistics of these addresses by categorising the parishes they fall within into urban and rural groups. Here we follow the minimum threshold of an urban settlement, one person per acre, proposed by Law [23] to determine the urban or rural classification of a parish. We understand it is a complex concept to define urban and rural settlements and there are more rigorous definitions [23] using a variety of criterion such as the population size, density, map evidence, degree of nucleation, etc. We take the population density as a crude and convenient criteria in this specific case, calculated using the population of that parish from the census report divided by the acreage of the parish. As shown in Table 3, 86.9% and 12.8% of the addresses are categorised as urban and rural addresses respectively, with a 0.4% residue of the unclassified addresses caused by missing the relevant information from the Census for the calculation. Regarding the match rates across different stages compared within each category, the majority of urban addresses (41.9%) are matched at the address level to the modern OS AddressBase, while in the rural parishes, most of the matched addresses (32.5%) are linked to the historical gazetteer.

Table 3. Summary results of the overall match rates.

Urban/Rural Classification	No. of Addresses	Match Using OS AddressBase		Match Using the Gazetteer	Unmatched
		(Street Level)	(Address Level)		
Urban	4,232,683 (86.9%)	1,142,235 (27.0%)	1,775,116 (**41.9%**)	839,056 (19.8%)	476,276 (11.3%)
Rural	622,600 (12.8%)	89,230 (14.3%)	73,065 (11.7%)	202,519 (**32.5%**)	257,786 (41.4%)
Not classified	17,424 (0.4%)	2967 (17.0%)	3324 (19.1%)	4605 (26.4%)	6528 (37.5%)
Overall	4,872,707 (100%)	1,234,432 (25.3%)	1,851,505 (**38.0%**)	1,046,180 (21.5%)	740,590 (15.2%)

To explore the geographical variation of the match rates, Figure 3 presents the proportions of geocoded addresses from the 1901 Census by the historical parishes in England and Wales. Figure 3a shows considerable variations in the match success using AddressBase, many of which reflect

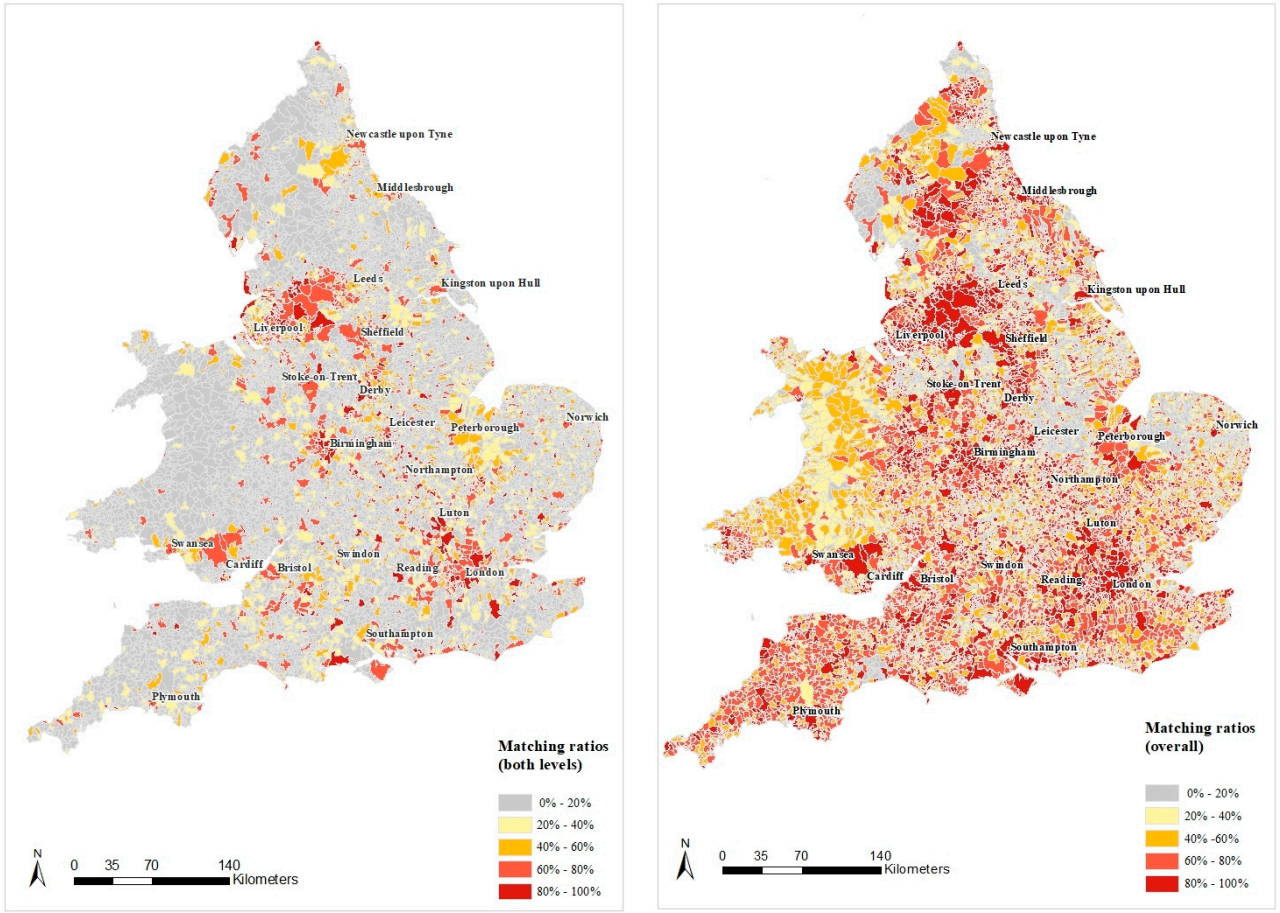
urbanity—with densely populated areas characterised by the highest match rates. It appears that fuzzy matching using AddressBase being much more successful in urban than rural areas. This indicates that address systems were more evolved in urban areas in 1901. In contrast, street names and street numbers in rural areas are less common as local landmarks were effective means of orientation. For example, many addresses in rural parishes were as vague as "old mill" or "house with a butcher's shop". These ambiguous addresses show the census respondents' perceptions of their home locations, which can only be linked to contemporary address registers with great difficulty. About 63% of the historical addresses are matched at this stage, most of which are in urban areas. Moreover, it is apparent that the portion of matches in Wales (2.4%) is much lower than that in England (60.9%) mainly because a large share of addresses are listed in the Welsh language.

Figure 3b presents the overall match rates by historical parishes in England and Wales, including matches from both the contemporary AddressBase and the historical gazetteer. Comparing Figures 3a and 3b, we observe that incremental improvements principally accrue in the geographically extensive rural parishes, though they only account for not more than 21.5% of all addresses. Parishes with high match rates (greater than 60%) are evenly distributed across England after this stage, other than concentrating in and around the major cities in Figure 3a. Significant improvement in address matching accrues in Wales using the historical gazetteer. Since the GB1900 Gazetteer was transcribed from historical map series, some of the place or street names were also annotated in Welsh. Match rates in both urban and rural areas in Wales are thus improved.

We take the example of the City of London to visualise the matched results at the address level in Figure 4. We overlay the geocoded address points from the OS AddressBase around Fleet Street on top of an OpenStreetMap base map layer. The address points are symbolised with green circles and labelled in red by the census address strings. Comparing the present thoroughfare names in black in the base map, we find the historical addresses align quite well with the OpenStreetMap streets. Although some of the addresses might not necessarily locate in the same points back in 1901 because of events such as urban redevelopment or the Blitz during wartime, we consider that they nevertheless present reasonable estimates of the original positions of the historical addresses.

4.2. Exploring Possible Social Gradient in the Geocoding Results

To explore the possible occupational bias in the geocoded records, we show the distributions of both the sub-population with geocoded addresses and the entire population with respect to the Historical International Standard Classification of Occupations (HISCO) occupational groups in Table 4. Detailed information about the HISCO occupational classification can be found on the website (https://historyofwork.iisg.nl/major.php). Around 80% of people who are associated with geo-referenced addresses. Only 65% of agricultural farmers, forestry workers, and fishermen are represented in the geocoded population, which is low relative to other groups and the national rate. These predominantly rural occupations aside, it appears there is no serious bias in the population that have been geocoded with our method among other groups.



(**a**) (**b**)

Figure 3. Match rates using AddressBase (**a**), supplemented with the historical gazetteer (**b**) by historical parishes in England and Wales.

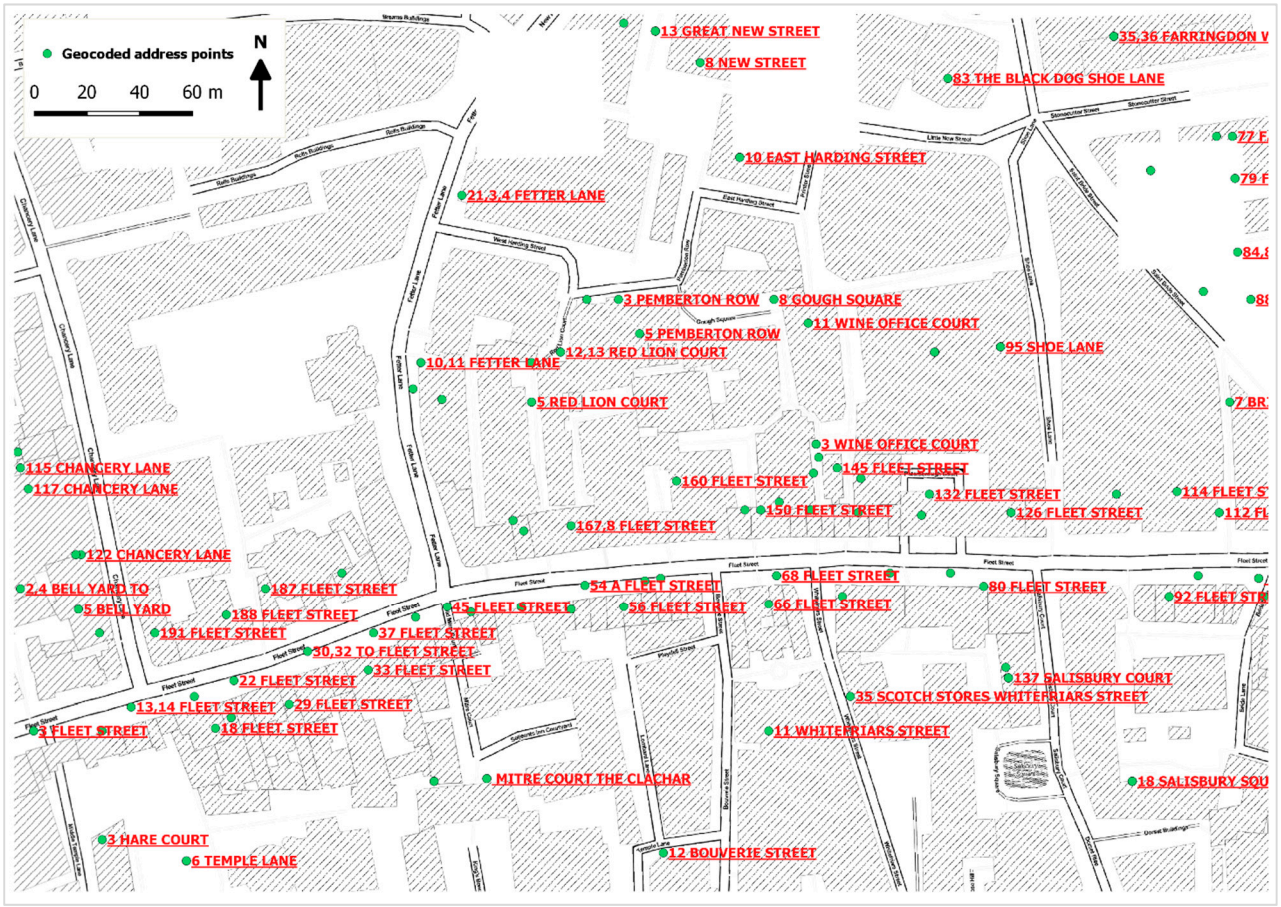


Figure 4. Examples of geo-referenced census addresses (green circles) and (red) labels along Fleet Street in the City of London. (Base map source: style using Stamen Black and White by Stamen Design, data © OpenStreetMap).

Table 4. Numbers and percentages of individuals with geocoded addresses by the HISCO occupational groups.

HISCO Occupation Groups	All Population	Geocoded Population	Percentage
Professional, technical and related workers	500,780	409,282	81.7%
Administrative and managerial workers	264,557	220,914	83.5%
Clerical and related workers	788,603	678,990	86.1%
Sales workers	1,366,246	1,143,359	83.7%
Service workers	2,526,565	1,930,180	76.4%
Agricultural, animal husbandry and forestry workers, fishermen and hunters	1,432,165	934,090	65.2%
Production and related workers, transport equipment operators and labourers	9,083,726	7,272,622	80.1%
Occupation not specified	16,530,676	13,292,585	80.4%
Total	32,493,318	25,882,022	79.7%

4.3. The Impact of Urban Changes on Address Matching

Although the majority of the historical addresses have been geo-referenced, c. 15% were not linked to either of the address corpora. Visual inspection of unmatched records suggested a range of possible causes, including vague or unrecognisable address strings, missing street names, or non-permanent accommodation such as vessels or non-fixed abodes [20]. Our underlying assumption of matching the historical addresses to the contemporary address registers in the proposed strategy is that streets and the addresses that delineate them remain largely unchanged over time. This might not be the case in some urban areas: in particular, where buildings or entire streets might be demolished during urban redevelopment. Names or alignment of streets could be changed during such redevelopment.

London is used here as an example to study the impact of urban changes on our geo-referencing strategy. We extract the street network of the modern Greater London Authority (GLA) from the OS Open Road data and snap the geocoded historical census address points falling within the GLA to the streets by thoroughfare names. The street network is visualised in a binary colour scheme in Figure 5. Road segments in green indicate that at least one address point has been matched onto the street segment, while streets coloured red identify streets in which no single address has been matched. It is not surprising that contiguous areas in the outer ring of the GLA areas are predominantly coloured red. The green segments largely delineate the extent and layout of the street network in 1901. Comparing this with the manually digitised streets in London around 1900 in an earlier study [24], we find that they broadly align with each other in terms of the geographic coverage.

We further superimpose the road network onto a geo-referenced historical map around the 1900s provided by the web map service by the National Library of Scotland (maps.nls.uk). Figure 6 displays an indicative locality of Clapham, London. Despite London's dynamic development over the last 118 years and the effects of World War 2, much of the street network has remained fundamentally unchanged, as indicated by the green road segments at the centre of the locality. Many addresses are almost identical between 1901 and the present day. However, the comparison also identifies streets superimposed upon previous green space, such as the road networks coloured in red towards the east of the Clapham Common. We also observe concentrations of failed matches, shown in red in the lower right corner of the map, presumably arising from redevelopment in these areas.

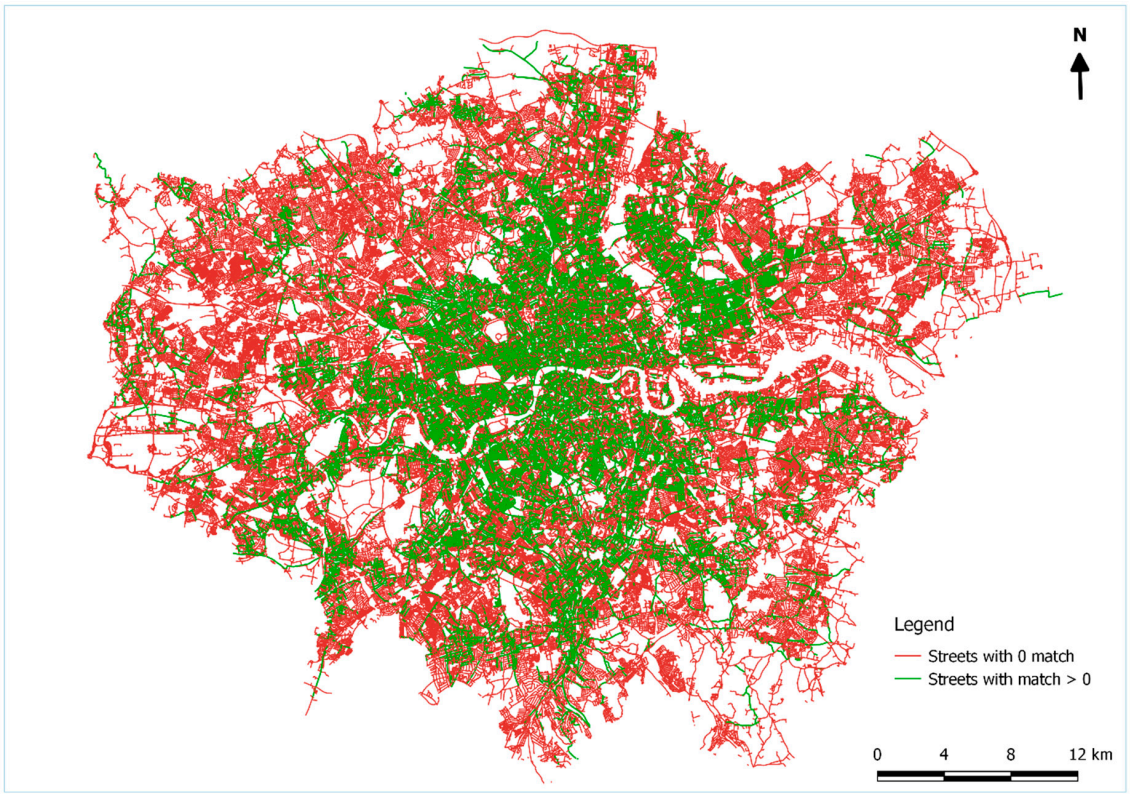


Figure 5. The street network of London in 2016, showing streets that have retained at least one identical address (green) and those that have no matches at all (red). (Source mapping: Street network from OS Open Road data).

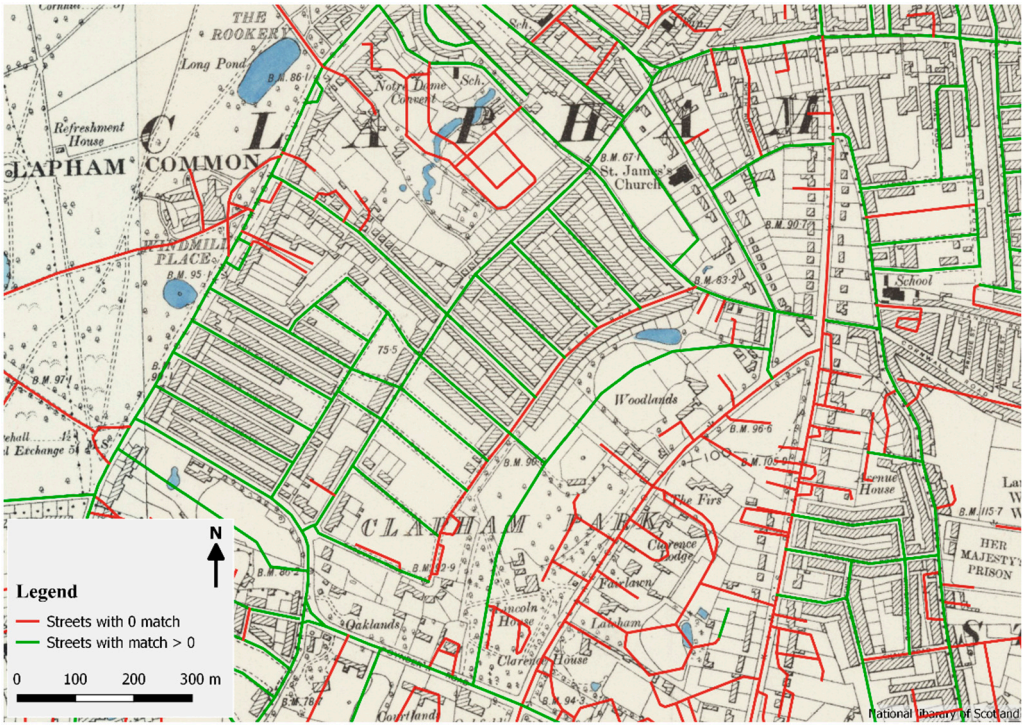


Figure 6. The street network of Clapham in 2016 overlaying a historical map from the early 1900s. (Source: Street network from the OS Open Road data; historical base map web service by the National Library of Scotland; historical map data from OS 1:1million–1:10K, the 1900s.).

5. Discussion and Conclusions

In this paper, we present our experience of geo-referencing historical addresses from the 1901 Census in England and Wales at a range of scales from the national to the local. We develop a two-stage address matching method, employing the geocoded contemporary and historical address corpora. We achieve an overall match rate of 85% (4,132,117 addresses) at varying levels covering 79.7% (25,882,022 people) of the entire population of England and Wales in 1901, which shows it is feasible to geo-reference a large share of historical census addresses with the proposed method.

The vast majority of the matched addresses (63%) are successfully linked to the contemporary AddressBase at the first stage. While this is by no means proof of linkage of the same built structure, it does demonstrate success in linking locations with high apparent spatial precision and provides a basis for evaluation of possible anomalies, such as streets that today have different number ranges or nomenclatures to those observed in 1901. The highest match rates to the contemporary AddressBase are concentrated in the historical cores of towns and cities. In absolute terms, this, of course, reflects the high concentrations of addresses in urban areas, but it is also clear that it was in urban areas that standard address referencing first emerged. The difference between urban and rural match rates also has been reflected in terms of the occupational distribution. We find the group of people conducting agricultural and forestry tasks is slightly underrepresented in the population with geocoded addresses with respect to other occupational groups. Apart from this, there is no serious bias among other groups.

Comparison of the mapped matches also provides broad brush indications of the nature and extent of redevelopment processes over the last 100+ years, as well as the effects of national and local planning systems in guiding or restricting development. At national levels, comparison of the periods also indicates the ravages of World War Two and subsequent redevelopment processes. Similar practices of linking historical addresses to the present address registers have been developed both in Scotland and England [19,20].

Linkage of historical census data to the current OS AddressBase misses some pieces of the historical address jigsaw. The GB1900 Gazetteer of Great Britain complements the OS AddressBase by geo-referencing an additional 22% of census addresses, drawn principally from rural parishes. Moreover, the Gazetteer also succeeds in matching a large number of addresses in the Welsh language by converting address strings into Unicode. Based on the observations, we find that modern address registers and historical gazetteers together appear to be two useful and complementary data sources for geo-referencing historical addresses. We currently have around 15% unmatched historical addresses, most of which are in rural parishes. Some of these could be linked through labour-intensively manual intervention, although our impression is that some Census records are simply too ambiguous to be linked with a high level of confidence.

One limitation of this work is the lack of validation of the matched results. Further spatial analysis of the geocoded censuses could be influenced by geocoding quality such as positional errors [25,26]. To date, we have only checked the outcome of matching in a few areas, but there is clear scope to use manual, semi-automated and automated methods in order to highlight and ultimately to accommodate anomalies in the matching process. This is clearly a fertile topic for future research.

In the context of other historical GIS and spatial humanities projects, geo-referencing is only a starting point of introducing further spatial analysis into geo-temporal demographic analysis. In addition to addresses, the historical censuses in England, Wales and Scotland provide a set of informative characteristics of residents, including population counts, demographics, household structure, occupation, fertility and disabilities. Geo-referencing of the historical census records spatially enables an intriguing range of topics in digital humanities and creates a framework for geo-temporal analysis where data from successive censuses can be linked. In our future research, we propose to apply automated techniques [21] developed for linking recent national consumer registers [27] to the similar linkage of historical census data. This agenda encompasses developing new geo-temporal demographics, migration analysis, segregation studies and an over-arching analysis of processes and patterns of social and spatial mobility in Great Britain.

Author Contributions: Conceptualization, Tian Lan and Paul Longley; Methodology, Tian Lan; Coding, Tian Lan; Formal Analysis, Tian Lan and Paul Longley; Writing-Original Draft Preparation, Tian Lan; Writing-Review & Editing, Paul Longley; Project Administration, Paul Longley.

Funding: This research was funded by the Engineering and Physical Sciences Research Council (EPSRC), grant number EP/M023583/1 (UK Regions Digital Research Facility) and the Economic and Social Research Council (ESRC), grant number ES/L011840/1 (Consumer Data Research Centre, CDRC).

Acknowledgments: The authors would like to thank the Great Britain Historical GIS project for making the GB1900 Gazetteer accessible. The authors would also like to thank the Cambridge Group for the History of Population and Social Structure for sharing the consistent historical parish boundaries.

Conflicts of Interest: The authors declare no conflicts of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

References

1. Higgs, E.; Schurer, K. *Integrated Census Microdata (I-CeM), 1851–1911*; UK Data Service: Colchester, UK, 2014.
2. Ruggles, S. Big Microdata for Population Research. *Demography* **2014**, *51*, 287–297. [CrossRef] [PubMed]
3. St-Hilaire, M.; Moldofsky, B.; Richard, L.; Beaudry, M. Geocoding and mapping historical census data: The geographical component of the Canadian Century Research Infrastructure. *Hist. Methods: J. Quant. Interdiscip. Hist.* **2007**, *40*, 76–91. [CrossRef]
4. Logan, J.R.; Jindrich, J.; Shin, H.; Zhang, W. Mapping America in 1880: The urban transition historical GIS project. *Hist. Methods* **2011**, *44*, 49–60. [CrossRef] [PubMed]
5. Schürer, K.; Garrett, E.M.; Jaadla, H.; Reid, A. Household and family structure in England and Wales (1851–1911): Continuities and change. *Contin. Chang.* **2018**, *33*, 365–411. [CrossRef]
6. Baskerville, P.; Dillon, L.; Inwood, K.; Roberts, E.; Ruggles, S.; Schürer, K.; Warren, J.R. Mining microdata: Economic opportunity and spatial mobility in Britain and the United States, 1850–1881. In Proceedings of the 2014 IEEE International Conference on Big Data (Big Data), Washington, DC, USA, 27–30 October 2014; pp. 5–13.
7. Congdon, P.; Campos, R.M.; Curtis, S.E.; Southall, H.R.; Gregory, I.N.; Jones, I.R. Quantifying and explaining changes in geographical inequality of infant mortality in England and Wales since the 1890s. *Int. J. Popul. Geogr.* **2001**, *7*, 35–51. [CrossRef]
8. Gregory, I.; Bennett, C.; Gilham, V.L.; Southall, H.R. The Great Britain Historical GIS Project: From Maps to Changing Human Geography. *Cartogr. J.* **2002**, *39*, 37–49. [CrossRef]
9. Longley, P.A.; Singleton, A.D. Linking social deprivation and digital exclusion in England. *Urban Studies* **2009**, *46*, 1275–1298. [CrossRef]
10. Singleton, A.D.; Longley, P.A. Geodemographics, visualisation, and social networks in applied geography. *Applied Geogr.* **2009**, *29*, 289–298. [CrossRef]
11. Goldberg, D.W.; Wilson, J.P.; Knoblock, C.A. From text to geographic coordinates: The current state of geocoding. *J. Urban Reg. Inf. Syst. Assoc.* **2007**, *19*, 33–46.
12. Gregory, I.; Donaldson, C.; Murrieta-Flores, P.; Rayson, P. Geoparsing, GIS, and Textual Analysis: Current Developments in Spatial Humanities Research. *Int. J. Humanit. Arts Comput.* **2015**, *9*, 1–14. [CrossRef]
13. Gregory, I.N.; Healey, R.G. Historical GIS: Structuring, mapping and analysing geographies of the past. *Prog. Hum. Geogr.* **2007**, *31*, 638–653. [CrossRef]
14. Orford, S.; Dorling, D.; Mitchell, R.; Shaw, M.; Smith, G.D. Life and death of the people of London: A historical GIS of Charles Booth's inquiry. *Health Place* **2002**, *8*, 25–35. [CrossRef]
15. Southall, H.; Aucott, P.; Fleet, C.; Pert, T.; Stoner, M. GB1900: Engaging the public in very large scale gazetteer construction from the ordnance survey "County series" 1: 10,560 mapping of Great Britain. *J. Map Geogr. Libr.* **2017**, *13*, 7–28. [CrossRef]
16. Cura, R.; Dumenieu, B.; Abadie, N.; Costes, B.; Perret, J.; Gribaudi, M. Historical Collaborative Geocoding. *ISPRS Int. J. Geo-Inf.* **2018**, *7*, 262. [CrossRef]
17. Carrion, D.; Migliaccio, F.; Minini, G.; Zambrano, C. From historical documents to GIS: A spatial database for medieval fiscal data in Southern Italy. *Hist. Methods J. Quant. Interdiscip. Hist.* **2016**, *49*, 1–10. [CrossRef]
18. Clough, P.; Tang, J.; Hall, M.M.; Warner, A. Linking archival data to location: A case study at the UK National Archives. *Aslib Proc.* **2011**, *63*, 127–147. [CrossRef]

19. Daras, K.; Feng, Z.; Dibben, C. HAG-GIS: A spatial framework for geocoding historical addresses. In Proceedings of the 23rd GIS Research UK Conference, Leeds, UK, 3–6 November 2015; pp. 15–17.
20. Walford, N.S. 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. [CrossRef]
21. Lansley, G.; Li, W.; Longley, P. Linked Consumer Registers for Granular Demographic Analysis. *Accept. J. R. Stat. Soc. Ser. (Stat. Soc.)* **2019**. [CrossRef]
22. Aucott, P.; Southall, H.; Ekinsmyth, C. Citizen science through old maps: Volunteer motivations in the GB1900 gazetteer-building project. *Hist. Methods: J. Quant. Interdiscip. Hist.* **2019**, 1–14. [CrossRef]
23. Law, C.M. The Growth of Urban Population in England and Wales, 1801–1911. *Trans. Inst. Br. Geogr.* **1967**, *41*, 125–143. [CrossRef]
24. Masucci, A.P.; Stanilov, K.; Batty, M. Limited urban growth: London's street network dynamics since the 18th century. *PLoS ONE* **2013**, *8*, e69469. [CrossRef] [PubMed]
25. Zandbergen, P.A. Influence of geocoding quality on environmental exposure assessment of children living near high traffic roads. *BMC Public Health* **2007**, *7*, 37. [CrossRef] [PubMed]
26. Zandbergen, P.A. Geocoding Quality and Implications for Spatial Analysis. *Geogr. Compass* **2009**, *3*, 647–680. [CrossRef]
27. Longley, P.; Cheshire, J.; Singleton, A. *Consumer Data Research*; UCL Press: London, UK, 2018. [CrossRef]