

Geo-linkages scoping review

By Nicolás Libuy and David Church





Economic and Social Research Council

Contents

Contents	1
Summary	4
1. Introduction	5
1.1 Literature Review	5
1.1.1. Geo-linkage Methods for Longitudinal Studies	5
1.1.2. Academic Research	7
2. Findings	11
2.1. Geographical Data	11
2.1.1. Review of Geographical Datasets	11
Table 2.1. Summary of the datasets reviewed	11
Figure 2.1. Linkage Potential Index	17
Table 2.2. Availability and Quality (AQ) and Temporal Extent (TE) scores	17
3. Conclusions	19
4. References	20
Appendix A	23
A.1. Individual 'Dataset Profile' documents	23
Table A.1. Data Profile: MasterMap Topography Layer	24
Table A.2. Relevant Documents: MasterMap Topography Layer	25
Figure A.1. Unified Modelling Language (UML) diagram – OS MasterMap Sites Laye	ər 26
Table A.3. Relevant Documents: MasterMap Sites Layer	27
Table A.4. Data Profile MasterMap: Highways Network	28
Table A.5. Relevant Documents: MasterMap Highways Network	29
Table A.6. Data Profile: MasterMap Topography Layer	31
Table A.7. Relevant Documents: MasterMap Topography Layer	32
Table A.8. Data Profile: MasterMap Imagery Layer	33
Table A.9. Georeferencing files	33
Table A.10. Relevant Documents: MasterMap Imagery Layer	34
Table A.11. Data Profile: MasterMap Water Network	35
Table A.12. Relevant Documents: MasterMap Water Network	36
Table A.13. Data Profile: MasterMap Greenspace Layer	37
Table A.14. Relevant Documents: MasterMap Greenspace Layer	39
Table A.15. Data Profile: AddressBase Premium	40
Table A.16. Explanation of AddressBase products	41
Table A.17. Relevant Documents: AddressBase Premium	42

Table A.18	. Data Profile: MasterMap Points of Interest	43
Table A.19	. Summary of MasterMap Points of Interest group categories	43
Table A.20	. Relevant Documents: MasterMap Points of Interest	44
Table A.21	. Data Profile: HM Land Registry Open Data	45
Table A.22	. Description of variables: 'Price Paid' data	46
Table A.23	. Description of variables: 'Transaction' data	47
Table A.24	. Description of variables: 'UK House Price Index' data	48
Table A.25	. Relevant Documents: HM Land Registry Open Data – House Price Index	49
Table A.26	. Relevant Documents: HM Land Registry Open Data – Paid Price Data	50
Table A.27	. Relevant Documents: HM Land Registry Open Data – Transaction Data	50
Table A.28	. Data Profile: Energy Performance Certificate (EPC)	52
Table A.29	. Description of variables: Domestic Energy Performance Certificate	53
Table A.30	. Description of variables: Non Domestic Energy Performance Certificate	59
Table A.31	. Description of variables: Display Energy Certificates	61
Table A.32	. Relevant Documents: Energy Performance Certificate (EPC)	64
Table A.33	. Data Profile: National Register for Social Housing	66
Table A.34	. Relevant Documents: National Register for Social Housing	67
Table A.35	. Data Profile: Historic Ordnance Survey	68
Table A.36	. Summary of Maps Available – Historic Ordnance Survey	69
Table A.37	. Relevant Documents: Historic Ordnance Survey	72
Table A.38	. Data Profile: Met Office data	73
Table A.39	. Summary of UKCP18 marine projections data	74
Table A.40	. Summary of UKCP18 climate models and scenarios for projections over lar	nd 75
Table A.41	. Available UKCP18 variables for the marine and projections over land	76
Table A.42	. Data Profile: Met Office data	.77
Table A.43	. Relevant Documents	78
Table A.44	. Data Profile: Air Pollution	79
Table A.45	. Available background mapping data for local authorities	81
Table A.46	. Relevant Documents: Air Pollution - AURN	81
Table A.47	. Relevant Documents: Air Pollution – Background Map Local Authorities	81
Table A.48	. Relevant Documents: Air Pollution - Modelled	81
Table A.49	. Data Profile: Crime Data	83
Table A.50	. Description of Variables: Crime Data	83
Table A.51	. Data Profile: Crime Data	84
Table A.52	. Relevant Documents: Crime Data	85
Table A.53	. Data Profile: Technology Data	87
Table A.54	. Data Profile: Technology Data	88

Table A.55. Description of Variables: Technology Data	88
Table A.56. Relevant Documents: Technology Data	90
Table A.57. Data Profile: Political Data	91
Table A.58. Summary of British Local Elections – 1889 - 1972	91
Table A.59. Summary of British Local Elections – 1973 - 2003	92
Table A.60. Relevant Documents: Political Data	92
Table A.61. Data Profile: Health Data	93
Table A.62. Description of Variables: Health Data - Practice prescribing data	94
Table A.63. Description of Variables: Health Data - Practice codes, names and addres	s 94
Table A.64. Description of Variables: Health Data - prescribing chemical and bnf codes	s. 94
Table A.65. Relevant Documents: Health Data	95
Table A.66. Data Profile: Educational Data	96
Table A.67. Overview of the standard extracts - NPD	96
Table A.68. Relevant Documents: Educational Data	97
Table A.69. Data Profile: Geo demographic segmentation Data	98
Table A.70. Description of Variables: Geo demographic categories	99
Table A.71. Relevant Documents: Geo demographic segmentation Data	100
Table A.72. Data Profile: Census Data 2001 and 2011	102
Table A.73. Relevant Documents: Census Data 2001 and 2011	103
Appendix B	103
B.1. Linkage Potential Index	103
Table B.1. Linkage Potential Index (LPI) scores	103

Summary

- We describe and provide examples of geographical methods that can be used to successfully link individual's location through their life with a wide spectrum of exposures, such as environmental and neighbourhood characteristics, physical resources, among others.
- We systematically analysed the linkage potential of twenty-one geographical datasets, according to their duration, and quality, and conclude there is great potential for making valuable geographical linkages with the CLS birth cohort studies
- A more detailed review of the scientific value of the specific variables within each geographical dataset is needed to prioritize academic research.

1. Introduction

The CLS birth cohort longitudinal studies provide a unique opportunity to study geographical exposure across the life course. Detailed information about cohort members' location at different moments in life allows researchers to link a wide number of geographical variables with excellent accuracy. However, the feasibility of such novel linkages is restricted by the availability, quality, and geographical and temporal extent of geographical datasets, as well as the availability for research of local identifiers derived from the addresses of study members (or other locations they spend time in, e.g. work, childcare, school).

There are three objectives in this scoping review. First, to provide an introduction to geographical linkages methods for Longitudinal Studies by showing how these methods are used in different pieces of literature. Second, to evaluate the quality and research potential of twenty-one geographical datasets. And third, to present potential challenges and recommendations regarding the linkage between the CLS birth cohort longitudinal studies and the geographical datasets.

This report is organized as follow. Section 1 discusses the geographical linkage methods and literature review. Section 2 shows the findings of the review of geographical datasets. Finally, Section 3 presents the recommendations and concludes. Appendix A includes a individual 'Dataset Profile' for each of the datasets reviewed.

1.1 Literature Review

This section describes the types of spatial data available in geographical data and provides an introductory overview of methods that can be used to link spatial data with birth cohort longitudinal studies. We use the classification created by H. Auchincloss et al. (2012) to show how proximity and aggregated methods can be used to link individual spatial trajectories with geographical variables, such as exposure to pollution, access to green spaces or fast-food restaurants, measures of local deprivation, or aggregated measures of educational segregation.

1.1.1. Geo-linkage Methods for Longitudinal Studies

Real world information is usually represented using coordinate values (e.g. latitude and longitude) and a system of reference for these coordinates (e.g. The World Geodetic System (WGS84)). While coordinates in a reference system allow us positioning real objects in a map, five types of spatial models allow us to accurately characterize the distinctive features of spatial objects:

- Points: a single point location, such as geocoded address or a store location
- Lines: a set of ordered points, connected by straight line segments, such as streets, rivers or a rail line.
- Polygons: an area, marked by one or more enclosing lines, possibly containing holes, such as lakes, regions, or postcodes.

- Grid or Raster: a collection of points or rectangular cells organised in a regular lattice, such as modelled background pollution maps at 1x1 km resolution or density of green vegetation.
- Networks: a complex line based structures such as the transport network.

Attributes, such as street names, stores types or maximum highway speed, can be associated with points, lines, polygons or grids, to extend the content of spatial data. Typically, Geographical Information Systems (GIS) are used to analyse, manage and store spatial data. GIS organizes layers of information and integrate many types of data, making them uniquely suited to perform extensive spatial operations¹ (Laurini and Thompson 1992). A variety of software², such as R, GRASS, QGIS, SAGA, GDAL or ArgGis are available to work with spatial data. While most of the GIS software work by 'pointing and clicking', R and QGIS allow the researcher to work with codes and command line instructions facilitating the transparency and reproducibility of the scientific practice.

In longitudinal birth cohort studies, the cohort member location (e.g. the location of individual's residence, schools or workplaces) can be used to provide insights about how changes in their environment affect individual outcomes and choices through the course of their lives. Several spatial methods can be used to link environmental features to the individual at a particular point in their life. According to H. Auchincloss et al. (2012), who reviewed spatial methods in peer-reviewed journals in Epidemiology, the most common spatial methods employed can be categorized in proximity methods, aggregation methods, and various forms of multivariable methods, such as clustering, spatial smoothing and interpolation methods, and spatial regression.

Spatial Proximity

The most common practice for measuring spatial proximity between a location and a spatial object varies by scientific fields, but a frequent method used is the distance to the nearest point, line or polygon centroid³. While the simplest way of calculating the distance between two points is a straight line or Euclidean distance, it usually does not reflect real world features such as the geography or the city structure. Others, more accurate, measures of proximity are Geodetic distances, Block distances, and Network distances. Geodetic distances use a mathematical model of the earth to calculate the length of the shortest curve between two points, Block distances use the shortest angular route between two points, and Network distances calculate the length of the shortest path between two points along the network. While Euclidean, Geodetic and Block distances can be easily calculated with standard statistical software, such as Stata or R, Network distances require specific tools available only on GIS software such as QGIS or ArcMap.

¹ Some of the spatial operation that can be used in a GIS are: clip, erase, split, or intersect spatial objects, among many others

² Detailed information can be found in the following webpages. GRASS: <u>https://grass.osgeo.org/;</u> QGIS: <u>https://www.qgis.org/en/site/;</u> SAGA: <u>http://www.saga-gis.org/;</u> GDAL: <u>https://gdal.org/;</u> ArgGis: <u>https://www.esri.com/</u>

³ A weighted centroid of a polygon can be also used as a point of reference when calculating the distance between two points. An example is a weighted population centroid where the 'centroid' is pulled towards areas with higher population density.

Aggregation Methods

In aggregation methods, a relevant area or buffer size must be defined to summarize geographical variables⁴. Relevant areas could be defined by administrative boundaries⁵ but buffers could also be created using a predefined fix radius around the individual's residence or using adaptive buffers techniques where the size varies depending on population density. Once the area or buffer is defined, several methods could be used to aggregate geographical variables within the predefined area. Simple aggregation operations such as counting, averaging or other statistics can be used to summarize information. More complex operations, that combine proximity and aggregation methods such as inverse distance weighting averaging methods, can be used to assign more weight to nearby spatial objects.

Multivariable methods: Cluster Detection Techniques, Spatial Interpolation and Smoothing Methods and Multivariable Spatial Regression

Cluster detection techniques (i.e. *'the process of grouping object with certain dimensions into groups such that objects within a group exhibit similar characteristics when compared to those which are in the other groups'* (Surendran 2013)) are typically divided into global, local and focused methods (Fritz et al. 2013). Local clustering methods focus on specific small-scale clusters, focused methods use a prefixed fixed point, and global methods do not use a specific location (H. Auchincloss et al. 2012). These methods are commonly used in Epidemiology to investigate suspicious geographical groupings of diseases or in hot spots analysis of criminal activity.

Spatial interpolation consists on using a sample of data points to fits a spatial surface where data are unobserved, and Smoothing methods consist on filtering information to reveal underlying spatial structures (Genebes, Renaud, and Sémécurbe 2018). Both methods are frequently used to improve the accuracy estimation around a point or small areas. Some spatial interpolation techniques are Inverse Distance Weighted Interpolation, Linear Regression, Kriging, among others.

Multivariable Spatial Regression Techniques are methods that account for the spatial autocorrelation and/or spatial heterogeneity observed in geographical data. The most common methods utilized in Epidemiology can be groped on Spatial autoregressive models and Bayesian regression models (H. Auchincloss et al. 2012).

1.1.2. Academic Research

Although scholars from different scientific fields have been interested in uncovering paths through which the proximity to local environments or differential exposure to real-world facilities could be associated with health, behaviour and individual's outcomes through life,

⁴ Geographical variables are defined as 'variables that provide information about the characteristics of a place, instead of just the location of that place' (Dibben and Feng 2003)

⁵ Some administrative area in the UK are: Government Office Regions, Standard Statistical Regions (SSRs), Statistical Wards, Census Area Statistics wards, Constituencies, Lower Super Output Areas, Output Areas, Enumeration Districts (EDs), Electoral Wards/Divisions.

the methods used to link individuals to exposures are mixed, leading to great variability across scientific fields. To illustrate the type of questions that could be investigated using spatial proximity, aggregation and/or multivariable regression methods, we present examples from a wide spectrum of scientific fields.

Air Pollution and Environmental Exposure

Air pollution and climate data are often recorded hourly using a national network of monitoring stations or satellites. Environmental researchers and official environmental agencies, such as the Department for Environment Food & Rural Affairs (DEFRA), use this data to produce modelled gridded time-series maps that researchers usually link with individual's location through life.

When the individual location is available researchers use it when linking exposure. Coneus and Spiess (2012) linked members of the German Socio-Economic Panel (SOEP) between 2002–2007 to 500 monitors in 11 states across Germany using the distance to the nearest monitor. Allen et al. (2009) use residential proximity to major roads to evaluate the association between traffic pollutants exposure and systemic atherosclerosis in the US. Houston et al. (2006) assessed child care facilities' proximity to heavily travelled roadways to evaluate the exposure of young children to vehicle-related pollution. Huss et al. (2009) study whether living in the proximity of 220–380 kV power lines is associated with a higher risk of death from neurodegenerative diseases.

However, researchers often do not have the individual's address but only the zip code of residence, and therefore assign pollution to the zip code centroid using an inverse distance weighted average of monitors near the zip code. For example, Moretti and Neidell (2011) use the shortest distance between the centroid of each zip code in an area in Los Angeles to the port in an attempt to measure exposure to ozone. Currie et al. (2009) assign air pollution to the zip code centroid where the mother lives using an inverse distance weighted average of monitors near the zip code of a large sample of infants born in New Jersey to evaluate air pollution effects on infant health. Foster et al. (2009) use satellite-based measures of aerosol optical depth to obtain estimates of air quality for the whole Mexican territory and then use inverse distance weighted average of monitors at the county level to evaluate the association between air quality and infant health. In an early study, Douglas and Waller (1966) used aggregated domestic coal consumption data to create local measure of pollution. They linked city level pollution measures to 3,866 children form the 1946 National Birth Cohort (Douglas and Blomfield, 1958) to explore the effect of air pollution on respiratory infection in children.

Similar techniques have been used to assess the effect of air pollution on cognitive development, productivity, mental health and road safety. Midouhas et al. (2018) linked 8,198 children of the Millennium Cohort Study (MCS) with mean annual estimates of nitrogen dioxide (NO2) levels within ward areas in the UK. They use 1 km grids, modelled from National Atmospheric Emissions Inventory data, to study the effect of outdoor and indoor air quality and cognitive ability in young children. Bishop et al. (2018) is maybe one of the biggest attempt to link longitudinal data with air pollution in the US. To track the evolution of individuals' health, the onset of dementia, financial decisions, and cumulative residential exposure to air pollution, they linked fifteen years of Medicare records for 6.9 million adults age 65 and older to the Environmental Protection Agency (EPA) air quality monitoring network. They combined the surface distance from each zip code centroid to each monitor and inverse weighting methods to assign air pollution to construct individual-specific

exposure histories. Chang et al. (2016) and Chang et al. (2016) studied the effect of air pollution on productivity of pear packers and Call Centres workers in China linking the nearest monitoring station to the individual's workplace. Sager (2018) use distance weighted average methods and atmospheric satellite data to study the effect of air pollution on road safety.

Proximity to Resources and Healthy Behaviours

Researchers interested in evaluating how accessibility to local resources and healthy environments could affect unhealthy behaviours, use a variety of proximity and aggregation methods (buffer definitions and sizes) to link the location of fast-food restaurants, grocery stores, public and green spaces.

Davis and Carpenter (2009) evaluate the association between fast-food restaurants within one half mile of the student's school and obesity among middle and high school students in California. Zenk et al. (2005) use block distance, i.e. the shortest angular route, to the nearest supermarket to study the spatial accessibility among the impoverished neighbourhoods in Detroit. Michael et al. (2010) used residential addresses to assess proximity to parks, trails, and recreational facilities to evaluate the influence of physical activity resources on walking behaviours among older men in Portland. They used one eighth, one quarter, and one half mile to aggregate the number of parks, trails, and recreational facilities.

Richardson et al. (2017) studied the relationship between green spaces and children's social, emotional and behavioural development. They linked 2,909 urban-dwelling children from the Growing Up in Scotland (GUS) survey to geocoded information of public parks and natural space within 500 meters (Euclidean distance) of each child's postcode. Flouri et al. (2019) use the Greenspace components of the Multiple Environmental Deprivation Index (MEDIx)⁶ to create ward level greenspace exposure. They linked 4,758 children from the Millennium Cohort Study (MCS) to study whether children living in greener neighbourhoods have better spatial working memory.

In a recent literature review, Carlin et al. (2017) argue that backyard space, neighbourhood features such as pedestrian and cyclist safety, or lack of sidewalks and streetlights could be associated physical activity behaviour. In this literature, researchers use alternative methods to characterize neighbourhood features. For example, Hobbsab et al. (2019) use the Point of Interest (POI) dataset form the Ordnance Survey (OS) and defined a 2 km radial buffer centred on the home postcode to create density measures of physical activity facilities and parks. They linked 8,864 individuals, between 18-68 years old, from the Yorkshire Health Study to examine the longitudinal associations between the recreational physical activity environment, change in body mass index, and obesity. In a related study, Cervero and Duncan (2003), use aggregation and proximity methods to evaluate the relationship between urban environments and walking and cycling behaviours in San Francisco. They first create analytical measures of built environment using factor analysis methods⁷ and then linked them to individual's trips records. Li et al. (2005) use neighbourhoods density measures of residential households, places of employment, street intersections, green and public spaces

⁶ See MEDIx; http://cresh.org.uk/cresh-themes/environmental-deprivation/medix-and-medclass/

⁷ Some of the variables included are square meters per block within 1 mile, three/four/five-way intersections as proportion of total intersections within 1 mile, measures of entropy that assess the type of residency variability (single-family housing units, multifamily housing units, retail/service employment, office employment, manufacturing/trade/other employment) within 1 mile.

to evaluate the walkability of an older population in Portland. Lovasi et al. (2009) use a buffer of 1 km radius around the participant's home address to study the association between walkable environments and lower body mass index (BMI) in New York City.

Local socio-economic environment and Neighbourhood

Researchers who study area effects across the life course must rely on historical measures of socio-economic variables. However, consistent area level information at multiple points of time is usually difficult to find, leading to different methodologies to overcome this challenge. For example, Murray et al. (2012) and Murray et al. (2013) study the association between area deprivation and physical capability across the life course using the National Survey of Health and Development (the 1946 British Birth Cohort). They use a wide variety of deprivation variables available⁸ in historical census data to create consistent area measure of 'social class' across time. In a similar study, Murray et al. (2019) link local authority unemployment through the life course of individuals to evaluate its association with health employment status, occupational class and educational achievement.

⁸ The authors use the following variables: low social class, unemployment, overcrowding, renters, no car, lacking household amenities and higher education.

2. Findings

2.1. Geographical Data

Twenty-one independent datasets, which potentially contained geographical variables across a wide variety of topics, were analysed to identify novel linkages opportunities using the CLS birth cohort longitudinal studies (National Child Development Study 1958, 1970 British Cohort Study, Next Steps and Millennium Cohort Study). A review of each dataset and an assessment of its potential for linkages opportunities are included in this report. A set of comparable 'Dataset Profile' documents are set out in an Appendix, where the potential for academic purposes of the data was judged according to its quality and the content of individual variables within each dataset.

2.1.1. Review of Geographical Datasets

A summary of the datasets reviewed is shown in Table 1. Each 'Dataset Profile' includes an ID number to uniquely identify the data; a description of the data; its main structure (e.g. key variables, sub-databases, etc.); temporal and geographical extent; analytical units (points, lines, polygons); format, approximate number of observations; supplier or data owner; the relevant document used; and an assessment of the data quality (See Appendix A for more details of each 'Dataset Profile').

ID	Name	Description	Temporal Extent	Supplier	Analytical Units	Format	Geo Extent	Access
1	Master Map Topogr aphy Layer	Most detailed and accurate view of GB's landscape – i.e. all buildings, green/blue spaces – contains UPRN/TOID keys for linking features across time. A vector map of Great Britain – can be used for calculating areas (e.g. of green space, buildings).	1999 - 2019	Ordnanc e Survey	Points, Lines, Polygons	Geodata bases (GDB), Shapefil e (SHP)	Great Britain	Open access for Educati onal Purpuse s
2	Master Map Highwa ys Networ k	GB road network, used for calculating drive times/distances and for creating zones of analysis around e.g. cohort members. A separate dataset contains urban pathways networks.	1997 – 2019	Ordnanc e Survey	Points, Lines, Polygons	Geodata bases (GDB), Shapefil e (SHP)	Great Britain	Open access for Educati onal Purpose s

Table 2.1. Summary of the datasets reviewed

3	Master Map Buildin g Heights	Adds a third dimension to analysis, by including building heights. Can be used to enhance accuracy of air pollution modelling and visualising urban density.	2014 – 2019	Ordnanc e Survey	Points, Lines, Polygons	Geodata bases (GDB), Shapefil e (SHP)	Great Britain	Open access for Educati onal Purpose s
4	Master Map Imager y Layer	High resolution (25cm) aerial photography. Used for spatial analysis of land use (e.g. green/blue space analysis).	1998 – 2019	Ordnanc e Survey	Images	TIFF, JPEG, and ECW	Great Britain	Open access for Educati onal Purpose s
5	Master Map Water Networ k Layer	Dataset showing flow and precise course of every river, stream, lake and canal in Great Britain. Used for blue space analysis.	2016 – 2019	Ordnanc e Survey	Points, Lines	Geodata bases (GDB), Shapefil e (SHP)	Great Britain	Open access for Educati onal Purpose s
6	Master Map Green Space Layer	Accessible and non-accessible green spaces, from domestic gardens up to National Parks. When used in conjunction with Points of Interest data, can permit highly detailed/accurate analysis of access to green spaces.	1999 – 2019	Ordnanc e Survey	Points, Lines, Polygons	Geodata bases (GDB), Shapefil e (SHP)	Great Britain	Open access for Educati onal Purpose s
7	Addres sBase Premiu m	Continually updated, canonical database of every postal/geographic address in Great Britain (circa 43 million features), which follows the lifecycle of each address, from construction to demolition. Contains UPRN/TOID keys. Used for geocoding (e.g. earliest sweeps of NCDS/BCS70, and for more accurate (e.g. environmental/pollu tion) at the address level.	1970 – 2019	Ordnanc e Survey	Points	Geodata bases (GDB), Shapefil e (SHP)	Great Britain	Open access for Educati onal Purpose s

8	Points of Interest	Dataset containing over 4 million commercial/non- residential features, classified by over 600 categories (e.g. bookmakers, supermarkets, fast food outlets). Also contains UPRN/TOID.	2005 – 2019	Ordnanc e Survey	Points	Geodata bases (GDB), Shapefil e (SHP)	Great Britain	Open access for Educati onal Purpose s
9	Land Registr y Price Paid dataset	Property-level dataset (circa 25 million records) containing all property sales (excluding some categories, e.g. Right to Buy) from January 1995 onwards.	1995 – 2019	Land Registry	Points (Postcode s)	CSV	Great Britain	Open access
10	Energy Perfor mance Certific ate data	Property-level dataset (circa 14 million records) containing many fields on energy use, insulation, central heating etc. Importantly, it also contains accurate an accurate measurement of each property's floor space (m ²).	2008 – 2016	Commun ities, Housing and Local Governm ent	Points, Polygons (Postcode s)	CSV	Great Britain	Once registre d on the webpag e, all data fields other than the address and postcod e data are Open Access
11	Nation al Registe r of Social Housin g (NROS H)	Property-level dataset (circa 3.3 million records) containing details on social housing (including nominal value, rent payable, year built, number of rooms, whether RTB etc.)	2014	National Register of Social Housing (NROSH)	Points, (Address)	CSV	Great Britain	Open Govern ment Licence
12	Historic Ordnan ce Survey data	Georeferenced scans of original OS maps of Great Britain, at 1:1250 scale (urban), 1:2500 scale (suburban) and 1:10560 scale (rural). Currently in raster format; functionality greatly enhanced if/when these data can be converted to vector format. Can be used in current format to assist with geocoding	1946 – 1992	Ordnanc e Survey	Points, Lines, Polygons	TIFF	Great Britain	Open access for Educati onal Purpose s

		earliest sweeps of NCDS/BCS70.						
13	Met Office data	A wide range of weather data available on a daily/monthly basis from (mostly) 1961 onwards. Includes max/min temperature, cloud cover, precipitation and sunshine.	1961 – 2016	Met Office	Points, Grids	Geodata bases (GDB), Shapefil e (SHP), CVS	Great Britain	Open access
14	Pollutio n data	Annual air pollution for six pollutants, interpolated on a 1km x 1km grid. It's possible that historic pollution data exists, going back to 1961. There is also proprietary monthly air pollution data at 100m x 100m available through KCL/Imperial College.	1999 – 2016	DEFRA	Points, Grids	Geodata bases (GDB), Shapefil e (SHP), CVS	Great Britain	Open access
15	Crime data	Street level dataset of circa 40 million crime events, categorised by 14 types. I constructed a crime index, based on the seriousness of the crime.	2010 – 2019	Police UK	Points (Streets)	Geodata bases (GDB), Shapefil e (SHP), CVS	Great Britain	Open access
16	Techno logy data	Postcode-level dataset containing data on broadband availability/usage.	2013 – 2019	OFCOM	Polygons (Local- and- unitary- authority; Output- area; Parliament ary- constituen cy;Postcod e)	CSV	Great Britain	Open access
17	Politica I data	Westminster constituency level results of every general election since the Second War. Ward level data on the 2016 EU referendum. Colin Rallings and Michael Thrasher's ward-level database of local elections results since 1945.	1951 – 2017	Various	Polygons (Westmins ter constituen cy)	CSV	Great Britain	Open access

18	Health data	Quarterly, GP level list of all medicines prescribed in England.	2010 - 2019	NHS	Points	CVS	Great Britain	Open Govern ment Licence
19	Educati on	School/pupil level performance metrics (e.g. NPD).	2007-2018	Ofsted/D fE	Points, Polygons	CVS	Great Britain	Restrict ed (Depart ment for Educati on)
20	Geode mograp hic data	Commercial, postcode level geodemographic classification (based on incomes/household expenditure/subscri ptions etc.)	2003, 2015, 2017	CACI	Points, Polygons (Postcode s and output areas)	CVS	Great Britain	Availabl e to Register ed or Authoris ed Users, UK Data Service
21	Census data	Output-area level data on a wide range of socio- economic variables. An interesting, and under-used aspect of 2001 and 2011 Census data is the interaction data (i.e. origin-destination matrices for migration and commuting).	1971, 1981, 1991, 2001, 2011	ONS	Polygons (Output areas)	Geodata bases (GDB), Shapefil e (SHP), CVS	Great Britain	Open access

Using the information collected in the Review, we evaluate the linkage potential of each geographical dataset using a simple rubric based on two main areas: Temporal Extent (TE), and Availability and Quality of supplementary materials (AQ). We create a Linkage Potential Index (LPI) that range from 0-7, where higher scores could be translated to a higher linkage potential. The objective of this rubric is to identify, among the twenty-one datasets in Table 1, which data has the higher potential to be linked with all CLS birth cohort studies. The TE area has a range of 0-4 points and the AQ area has a range of 0-3 points, where each point represents a specific item. The LPI is the sum of TE and AQ areas. At the beginning of the scoring process, each dataset was scored with the maximum in each item and points were subtracted when the evaluated criterion was not present.

The TE area evaluates the feasibility of linking each geographical data with the CLS birth cohort studies included in this scoping review (i.e. National Child Development Study, British Cohort Study, Next Steps, and Millennium Cohort Study). To construct the TE score we followed three steps. 1) We calculated the temporal extent of each CLS birth cohort study in years (e.g. MCS started in 2000; therefore, it has a temporal extension of nineteen years). 2) Using the starting year of each geographical data and for each CLS birth cohort study, we calculated the temporal extension in years not covered by datasets in Table 1 (e.g. 'MasterMap Points of Interest' started in 2005, meaning that during the first six years of the MCS, from 2000 to 2005, this data cannot be linked to the MCS. For the MCS, there are six

years that are not covered by the 'MasterMap Points of Interest'). And finally, we divided 1) by 2) to create the percentage of each CLS birth cohort study not covered by a dataset in Table 1 (e.g. the 'MasterMap Points of Interest' does not cover a 31.6 (=6/16) percentage of the MCS temporal extension). Four score items, one per CLS birth cohort, were created as one minus the percentage of a birth cohort study not covered by a dataset (e.g. following the previous example, the TE score for the 'MasterMap Points of Interest' in the MCS item is equal to 0.684(=1-0.316)). For example, using this scoring scheme, geographical data that covers the complete period of the MCS scored 1 point.

The AQ area includes the following three binary items: 1) The dataset does not includes variables that help to identify errors or assumptions during the data process; 2) The ownership of the data flags the overall quality of the datasets within the supplementary documents; and 3) It was not possible to find enough documentation to identify the variables included in the data. If an item is true for specific data, one point was subtracted from the total AQ score.

Figure 2.1 shows the distribution of the Linkage Potential Score and Table 2.2 shows results for both, the Availability and Quality (AQ) and Temporal Extent (TE) scores. Appendix B includes detailed individual scores.

Overall, we find great potential for linkage opportunities with the Ordnance Survey products, 'The British Local Elections Database', Meteorological and Air Pollution datasets. These results are mainly driven by the long temporal extension of these datasets. When we split the Linkage Potential Score by sub-area we find a high heterogeneity for TE score within each AQ score (see Table 2.2).

A description of all variables within each geographical data is shown in the 'Data profile' documents in the Appendix A.

Figure 2.1. Linkage Potential Index



Table 2.2. Availability and Quality (AQ) and Temporal Extent (TE) scores

ID	Availability and Quality of supplementary materials score	Temporal Extent Score
	AQ score =3	
12	Historic Ordnance Survey data	4.00
17	Political data	4.00
13	Met Office data	3.93
7	AddressBase Premium	3.77
21	Census data	3.73
2	MasterMap Highways Network	2.47
4	MasterMap Imagery Layer	2.40
6	MasterMap Green Space Layer	2.33
1	MasterMap Topography Layer	2.33
14	Pollution data	2.33
3	MasterMap Building Heights	1.72
8	Points of Interest	1.60
5	MasterMap Water Network Layer	0.25

	AQ score =2	
20	Geodemographic data	1.84
19	Education	1.35
15	Crime data	0.98
18	Health data	0.98
16	Technology data	0.61
	AQ score =1	
9	Land Registry Price Paid dataset	2.61
10	Energy Performance Certificate data	1.23
11	National Register of Social Housing	
	(NROSH)	0.49

3. Conclusions

Linking CLS birth cohort studies to geographical data can enhance longitudinal research studies in many ways and its feasibility usually depends, among other issues, on the availability of high-quality geographical data. In this review, we reviewed spatial linkage methods and a set of available geographical datasets that can be successfully linked to the CLS birth longitudinal cohorts. In particular, we ranked available geographical data using a simple rubric to assess the feasibility of the linkage process.

Two main recommendations can be extracted from the present work. First, while the geographical/temporal extent and the quality of geographical datasets are useful to hierarchize linkage efforts, a more detailed review of the variables within each geographical dataset is needed to prioritize academic research. For example, by using proximity or aggregation spatial methods to count the number of Fast Food restaurants near to individuals' locations, the 'MasterMap Point of Interest' dataset could be used to create a more accurate individual exposure to obesogenic environments, which would add strong scientific value to obesity research within the cohorts. A more comprehensive mapping of variables within each geographical dataset would enable us to identify other specific topics that would be of high scientific value within the context of the CLS birth cohort studies, and to identify gaps in previous research in collaboration with, or advice from academics in specific fields.

Second, a more detailed review would enable important technical issues to be considered, such as secular changes in area measures, changes in geographical units over time (Murray et al., 2012), and ethical and legal issues regarding consents and disclosure risk (Calderwood and Lessof, 2009).

4. References

Allen, Ryan W. et al. 2009. "Fine Particulate Matter Air Pollution, Proximity to Traffic, and Aortic Atherosclerosis." *Epidemiology* 20(2): 254–64.

Bishop, Kelly C, Jonathan D Ketcham, and Nicolai V Kuminoff. 2018. *Hazed and Confused: The Effect of Air Pollution on Dementia*. National Bureau of Economic Research. Working Paper. http://www.nber.org/papers/w24970 (June 7, 2019).

Calderwood, Lisa, and Carli Lessof. 2009. "Enhancing Longitudinal Surveys by Linking to Administrative Data." In *Methodology of Longitudinal Surveys*, John Wiley & Sons.

Carlin, Angela et al. 2017. "A Life Course Examination of the Physical Environmental Determinants of Physical Activity Behaviour: A 'Determinants of Diet and Physical Activity' (DEDIPAC) Umbrella Systematic Literature Review." *PLoS ONE* 12(8). https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5546676/ (June 7, 2019).

Cervero, Robert, and Michael Duncan. 2003. "Walking, Bicycling, and Urban Landscapes: Evidence From the San Francisco Bay Area." *American Journal of Public Health* 93(9): 1478–83.

Chang, Tom, Joshua Graff Zivin, Tal Gross, and Matthew Neidell. 2016. "Particulate Pollution and the Productivity of Pear Packers." *American Economic Journal: Economic Policy* 8(3): 141–69.

Chang, Tom, Joshua Graff Zivin, Tal Gross, and Matthew Neidell. 2016. *The Effect of Pollution on Worker Productivity: Evidence from Call-Center Workers in China*. National Bureau of Economic Research. Working Paper. http://www.nber.org/papers/w22328 (June 7, 2019).

Coneus, Katja, and C. Katharina Spiess. 2012. "Pollution Exposure and Child Health: Evidence for Infants and Toddlers in Germany." *Journal of Health Economics* 31(1): 180–96.

Currie, Janet, Matthew Neidell, and Johannes F. Schmieder. 2009. "Air Pollution and Infant Health: Lessons from New Jersey." *Journal of Health Economics* 28(3): 688–703.

Davis, Brennan, and Christopher Carpenter. 2009. "Proximity of Fast-Food Restaurants to Schools and Adolescent Obesity." *American Journal of Public Health* 99(3): 505–10.

Dibben, Chris, and Zhiqiang Feng. 2003. *A Review of Resources for Geographic Variables*. London, UK: CLOSER. Resource report.

Douglas, J. W. B., and J. M. Blomfield. 1958. "Children under Five. The Results of a National Survey Made by a Joint Committee of the Institute of Child Health (University of London), the Society of Medical Officers of Health and the Population Investigation Committee." *Children under Five. The Results of a National Survey made by a Joint Committee of the Institute of Child Health (University of London), the*

Society of Medical Officers of Health and the Population Investigation Committee. https://www.cabdirect.org/cabdirect/abstract/19592701913 (June 28, 2019).

Douglas, J. W. B., and R. E. Waller. 1966. "Air Pollution and Respiratory Infection in Children." *British Journal of Preventive & Social Medicine* 20(1): 1–8.

Flouri, Eirini, Efstathios Papachristou, and Emily Midouhas. 2019. "The Role of Neighbourhood Greenspace in Children's Spatial Working Memory." *British Journal of Educational Psychology* 89(2): 359–73.

Foster, Andrew, Emilio Gutierrez, and Naresh Kumar. 2009. "Voluntary Compliance, Pollution Levels, and Infant Mortality in Mexico." *The American Economic Review* 99(2): 191–97.

Fritz, E. Charles, Nadine Schuurman, Colin Robertson, and Scott Lear. 2013. "A Scoping Review of Spatial Cluster Analysis Techniques for Point-Event Data | Geospatial Health." *Geospatial Health* 7(2): 183-198.

Genebes, Laure, Auriane Renaud, and Francois Sémécurbe. 2018. "Ch. 8. Spatial Smoothing." In *Handbook of Spatial Analysis, Theory and Application with R*, Insee - Eurostat.

H. Auchincloss, Amy, Samson Y. Gebreab, Christina Mair, and Ana V. Diez Roux. 2012. "A Review of Spatial Methods in Epidemiology, 2000–2010 | Annual Review of Public Health." *Annual Review of Public Health* 33: 107–22.

Hobbsab, M. et al. 2019. "Examining Longitudinal Associations between the Recreational Physical Activity Environment, Change in Body Mass Index, and Obesity by Age in 8864 Yorkshire Health Study Participants." *Social Science & Medicine* 227: 76–83.

Houston, Douglas, Paul Ong, Jun Wu, and Arthur Winer. 2006. "Proximity of Licensed Child Care Facilities to Near-Roadway Vehicle Pollution." *American Journal of Public Health* 96(9): 1611–17.

Huss, Anke, Adrian Spoerri, Matthias Egger, and Martin Röösli. 2009. "Residence Near Power Lines and Mortality From Neurodegenerative Diseases: Longitudinal Study of the Swiss Population." *American Journal of Epidemiology* 169(2): 167–75.

Laurini, Robert, and Derek Thompson. 1992. *Fundamentals of Spatial Information Systems - 1st Edition*. 1st ed. Academic Press.

Li, F., K Fisher, R. Brownson, and M. Bosworth. 2005. "Multilevel Modelling of Built Environment Characteristics Related to Neighbourhood Walking Activity in Older Adults." *Journal of Epidemiology and Community Health* 59(7): 558–64.

Lovasi, Gina S. et al. 2009. "Effect of Individual or Neighborhood Disadvantage on the Association Between Neighborhood Walkability and Body Mass Index." *American Journal of Public Health* 99(2): 279–84.

Michael, Yvonne L. et al. 2010. "Physical Activity Resources and Changes in Walking in a Cohort of Older Men." *American Journal of Public Health* 100(4): 654–

60.

Midouhas, Emily, Theodora Kokosi, and Eirini Flouri. 2018. "Outdoor and Indoor Air Quality and Cognitive Ability in Young Children." *Environmental Research* 161: 321–28.

Moretti, Enrico, and Matthew Neidell. 2011. "Pollution, Health, and Avoidance Behavior Evidence from the Ports of Los Angeles." *Journal of Human Resources* 46(1): 154–75.

Murray, Emily T. et al. 2012. "Challenges in Examining Area Effects across the Life Course on Physical Capability in Mid-Life: Findings from the 1946 British Birth Cohort." *Health & Place* 18(2): 366–74.

———. 2013. "Area Deprivation Across the Life Course and Physical Capability in Midlife: Findings From the 1946 British Birth Cohort." *American Journal of Epidemiology* 178(3): 441–50.

———. 2019. "Linking Local Labour Market Conditions across the Life Course to Retirement Age: Pathways of Health, Employment Status, Occupational Class and Educational Achievement, Using 60 Years of the 1946 British Birth Cohort." *Social Science & Medicine (1982)* 226: 113–22.

Richardson, Elizabeth A., Jamie Pearce, Niamh K. Shortt, and Richard Mitchell. 2017. "The Role of Public and Private Natural Space in Children's Social, Emotional and Behavioural Development in Scotland: A Longitudinal Study." *Environmental Research* 158: 729–36.

Sager, Lutz. 2018. *Estimating the Effect of Air Pollution on Road Safety Using Atmospheric Temperature Inversions*. Grantham Research Institute on Climate Change and the Environment.

Surendran, Subu. 2013. "Review of Spatial Clustering Methods." *International Journal of Information Technology Infrastructure* 2(3). http://warse.org/pdfs/2013/ijiti01232013.pdf.

Zenk, Shannon N. et al. 2005. "Neighborhood Racial Composition, Neighborhood Poverty, and the Spatial Accessibility of Supermarkets in Metropolitan Detroit." *American Journal of Public Health* 95(4): 660–67.

Appendix A

This Appendix includes all 'Dataset Profile' documents created as part of the review of datasets. Some of the individual 'Dataset Profiles' included texts that were copied from datasets manuals, webpages, and technical notes.

A.1. Individual 'Dataset Profile' documents

Data Profile 1: MasterMap Topography Layer and Sites Layer

Name of the Data: MasterMap Topography Layer9

This data profile describes a dataset held by the Ordnance Survey that has been publicly available for educational or research purposes through the © Crown copyright.

Description of the Data

The MasterMap Topography Layer is a maintained framework for referencing geographic information in Great Britain. A unique reference number called TOID identify individual real-world topographic features represented by points, lines, and polygons. Each feature is uniquely referenced and has attributes that record the feature's life cycle. The feature's life cycle is linked to the life cycle of the real-world object it represents. The life cycle records certain types of changes to the feature that occur over time¹⁰.

Data Profile

Field	Description
Dataset ID	1
Dataset Name	MasterMap Topography Layer
Temporal Extent	1999 - 2019
Supplier/data owner	Ordnance Survey
Analytical Units	Points, Lines, Polygons
Data Format	Geodatabases (GDB), Shapefile (SHP)
Geographical Extent	Great Britain
Access	Open access for Educational Purposes
Cost	Free
Observations	Circa 500 million
Web page	https://www.ordnancesurvey.co.uk/business-and- government/products/topography-layer.html

Table A.1. Data Profile: MasterMap Topography Layer

Structure of the Data

The Topography Layers is grouped into feature themes. Within each theme there are features deemed to belong to that theme. A feature can belong to more than one theme.

Each feature has one or more versions of itself, which can be tracked as a means of demonstrating change over time. New versions of features replace existing features through change only updates (COU) and features that no longer exist can be deleted as a result.

⁹ This document is a summary of relevant elements contained in the "OS MasterMap Topography Layer Product Guide v2 08/2017"

¹⁰ Ordnance Survey updates OS MasterMap Topography Layer on a regular basis, and these changes are passed on to the users through a dataset called change-only update (COU).

Themes: A theme is a set of features that have been grouped together to provide a highlevel means of dividing the data on the layer in a logical and user-friendly fashion. A feature can be a member of any number of themes but must belong to at least one theme as a minimum requirement. OS MasterMap Topography Layer comprises nine themes. These are:

- Administrative boundaries
- Buildings
- Heritage and antiquities
- Land
- Rail
- Roads, tracks, and paths
- Structures
- Terrain and height
- Water

Features: A real-world object is represented by a feature in OS MasterMap data. Each feature has one of three geometrical structures – a point, a line, or a polygon. The list of all real-world objects is in the OS MasterMap[™] real-world object catalogue.

Attributes: Each feature comes with an extensive set of attributes that provide information about the feature, for example, its identity, its relationship to other features, geometry, and the kind of real-world object it purports to represent.

Relevant Documents

The following documents are included in the folder attached to this document.

Document Name	Name file
Real World Object Catalogue	os-mastermap-real-world-object-catalogue.pdf
Topography Layer Product Guide	os-mastermap-topography-layer-product-guide.pdf
Topography Layer Technical Specification	os-mastermap-topography-layer-technical- specification.pdf

Table A.2. Relevant Documents: MasterMap Topography Layer

Name of the Data: MasterMap Sites Layer¹¹

Description of the Data

The MasterMap Sites Layer is a view of the detailed extents of important locations such as airports, schools, hospitals, utility and infrastructure sites, and more. The points of access and routing points, to navigate in and out of the sites, are also provided. The access and routing points, reference OS MasterMap Integrated Transport Network[™] (ITN) Layer features. OS MasterMap Sites Layer features are a representation of what people see in the real world.

Structure of the Data

It contains three feature types or spatial object types:

 $^{^{11}}$ This document is a summary of relevant elements contained in the "OS MasterMap Sites Layer user guide chapter 1 v1.2 – 06/2016"

- Functional site polygons (FunctionalSite);
- Functional site access points (AccessPoint);
- Functional site routing points (RoutingPoint)

The Sites Layers is grouped into seven feature themes:

- Air transport
- Education
- Medical care
- Rail transport
- Road transport
- Water transport
- Utility or industrial

Unique identifiers

OS MasterMap Sites Layer contains two unique identifiers; both are populated with the TOID. The TOID is a unique identifier that is associated with every feature in many of Ordnance Survey's large-scale products. It is a unique reference consisting of the prefix 'osgb' and up to 16 digits, which will stay the same during a features lifecycle. GML. It also requires a gml:id attribute which must be unique. It was decided to populate this attribute with the TOID to ensure only one set of unique identifiers must be managed.

The next figure shows the attributes for each feature type.





Source: 'OS MasterMap Sites Layer user guide chapter 1 v1.2 - 06/2016'

Relevant Documents

The following documents are included in the folder attached to this document.

Document Name	File Name
Sites Getting Started guide	os-mastermap-sites-getting-started-guide.pdf
Sites Layer Release Notes	os-mastermap-sites-layer-release-notes-oct18.pdf
Sites Layer Technical Specification	os-mastermap-sites-layer-technical-specification.pdf
Sites Layer User Guide	os-mastermap-sites-layer-user-guide.pdf

 Table A.3. Relevant Documents: MasterMap Sites Layer

Quality

There are several features of this data that makes it an invaluable resource for crosssectional and longitudinal geospatial analysis. By using a unique identifier when representing real-world features across all Great Britain it allows researchers to consistently analyse cross-sectional and longitudinal changes of real-world objects. Besides, several steps are performed by the Ordnance Survey to continually update the data and assess its quality. The Ordnance Survey also provides all the technical documents and Metadata to evaluate the quality of the data. For example, the quality of each data element is characterized according to the following criteria: completeness, logical consistency, positional accuracy, temporal accuracy, and thematic accuracy (attribute accuracy).

Data Profile 2: MasterMap Highways Network

Name of the Data: MasterMap Highways Network¹²

Description of the Data: OS MasterMap Highways Network brings together Ordnance Surveys detailed road and path information together with the National Street Gazetteer (NSG) and the Trunk Road Street Gazetteer (TRSG). The NSG and TRSG contains the definitive information provided by the Local and National Highways authorities. Currently, OS MasterMap Highways Network product family only contains the information from Local and National Highway Authorities from across England and Wales.

To bring this information together a spatial match has been carried out between the geometry of the Ordnance Survey's Road Links and the geometry of the NSG and TRSG Elementary Street Units (ESUs). The ESUs make up the Unique Street Reference Number (USRN) which is the identifier used within the NSG and TRSG and within the Ordnance Survey's AddressBase® products. The Ordnance Survey geometry is the base geometry used for the Highways network and where this has been spatially matched to an ESU this has enabled the bringing together of the NSG and TRSG with Ordnance Survey data.

Additionally, OS MasterMap Highways Network is enriched with third party information on speed data which is connected to the Ordnance Survey road network and will provide detailed information on average speed and speed limits across Great Britain.

Data	Profile
------	---------

Field	Description	
Dataset ID	2	
Dataset Name	MasterMap Highways Network	
Temporal Extent	1997 – 2019	
Supplier/data owner	Ordnance Survey	
Analytical Units	Points, Lines, Polygons	
Data Format	Geodatabases (GDB), Shapefile (SHP)	
Geographical Extent	Great Britain	
Access	Open access for Educational Purposes	
Cost	Free	
Observations	-	
Web page	https://www.ordnancesurvey.co.uk/business-and- government/products/os-mastermap-highways-network- products.html	

Table A.4. Data Profile MasterMap: Highways Network

Structure of the Data:

The OS MasterMap Highways Network product family includes three core products:

- OS MasterMap Highways Network Roads
- OS MasterMap Highways Network Roads and Routing and Asset Management

 $^{^{12}}$ This document is a summary of relevant elements contained in the "OS MasterMap Highways Network product guide Chapter 3 v2.3 – _07/2018"

• OS MasterMap Highways Network – Paths.

In addition to the three core products, the OS MasterMap Highways Network family is enhanced with 3 speed data products:

- OS Master Highways Network with Routing and Asset Management Information and Average Speed
- OS Master Highways Network with Routing and Asset Management Information and Speed Limits
- OS Master Highways Network with Routing and Asset Management Information and Average Speed and Speed Limits

OS MasterMap Highways Network product contains a number features. These include;

- Unique Street Reference Number (USRN)
- Road names from the naming and numbering authority
- DfT road classifications
- Road maintenance authority
- Motorway Junction to junction information
- Routing information
- Height, weight, width and length restrictions information
- Special designations
- Road reinstatement information
- Connected network across GB including all islands through the Ferry Network
- Average Speed information broken down into six time periods for each day
- Speed Limits

Relevant Documents

The following documents are included in the folder attached to this document.

Data Profile

Document Name	Name file
Highways playbook	highways-playbook-v1.pdf
ITN highways comparison	itn-highways-comparison.pdf
Highways network guide to	os-mastermap-highways-network-guide-to-routing.pdf
routing	
Highways network paths	os-mastermap-highways-network-paths-technical-
technical specification	specification.pdf
Highways network product	os-mastermap-highways-network-product-guide.pdf
guide	
Highways network roads	os-mastermap-highways-network-roads-technical-
technical specification	specification.pdf
Highways network routing	os-mastermap-highways-network-routing-and-asset-
and asset management	management-technical-specification.pdf
technical specification	
Highways network speed	osmastermap-highways-network-speed-data-technical-
data technical specification	specification-v1.pdf
Osmm highways network	osmm-highways-network-getting-started-guide.pdf
getting started guide	

Use of NSG and ordnance	use-of-nsg-and-ordnance-survey-data.pdf
survey data	

Quality

The OS MasterMap Highways Network, as well as others OS products, is reviewed frequently to update and correct information in the data. One of the unique characteristic of this particular set of data is that allow to perform Network Analysis using ESRI products as ArcMap across all Great Britain. Another of the key strengths of this product is the collection of street information at the local highway authority level. The benefit of this is that the data capture is at the earliest point of creation within the local highway authority and there is detailed local knowledge driven by statutory requirements.

Data Profile 3: MasterMap Topography Layer

Name of the Data: MasterMap Topography Layer – Building Height Attribute¹³

Description of the Data

OS MasterMap Topography Layer – Building Height Attribute is an enhancement to, and forms part of, OS MasterMap Topography Layer. It provides a set of height attributes for Topographic Area features with a buildings theme within OS MasterMap Topography Layer.

Data Profile

Field	Description
Dataset ID	3
Dataset Name	MasterMap Building Heights
Temporal Extent	2014 – 2019
Supplier/data owner	Ordnance Survey
Analytical Units	Point, Lines, Polygons
Data Format	Geodatabases (GDB), Shapefile (SHP)
Geographical Extent	Great Britain
Access	Open access for Educational Purposes
Observations	Circa 41 million features
Web page	https://www.ordnancesurvey.co.uk/business and- government/products/os-mastermap-building-height-attribute.html
Data Description	Adds a third dimension to analysis, by including building heights. Can be used to enhance accuracy of air pollution modelling and visualising urban density.

Table A.6. Data Profile: MasterMap Topography Layer

Structure of the Data

The OS MasterMap® Topography Layer – Building Height Attribute is generated from photogrammetrically derived Digital Terrain Models (DTM) and Digital Surface Models (DSM). That is, information gathered from aerial survey as part of our cyclical revision programmes. It includes absolute a relative hight values for structures.

¹³ This document is a summary of relevant elements contained in the "OS MasterMap

[®] Topography Layer-Building Height Attribute Product Guide v1.2"

Relevant Documents

The following documents are included in the folder attached to this document.

Document Name	Name file
Building height attribute getting started guide	building-height-attribute-getting-started-guide.pdf
Building height attribute release note	building-height-attribute-release-note-april19.pdf
Osmm topography layer bha product guide	osmm-topo-layer-bha-product-guide.pdf
Osmm topography layer bha technical specification	osmm-topo-layer-bha-technical-specification.pdf

Quality

There are several features of this data that makes it an invaluable resource for crosssectional and longitudinal geospatial analysis. By using a unique identifier when representing real-world features across all Great Britain it allows researchers to consistently analyse cross-sectional and longitudinal changes of real-world objects. Besides, several steps are performed by the Ordnance Survey to continually update the data and assess its quality. The Ordnance Survey also provides all the technical documents and Metadata to evaluate the quality of the data. It includes attributes that allow to the user to quantify different levels of the accuracy and missing values of the Building Height.

Data Profile 4: MasterMap Imagery Layer

Name of the Data: MasterMap Imagery Layer¹⁴

Description of the Data

The OS MasterMap Imagery Layer adds a visual and contextual capability to the other OS MasterMap Layers and OS data. Ordnance Survey has set the specification and quality levels for the OS MasterMap Imagery Layer with the aim of providing a reliable and consistent source of orthorectified aerial photography for general use in Great Britain.

Data Profile

Field	Description
Dataset ID	4
Dataset Name	MasterMap Imagery Layer
Temporal Extent	1998 – 2019
Supplier/data owner	Ordnance Survey
Analytical Units	Images
Data Format	TIFF, JPEG, and ECW
Geographical Extent	Great Britain
Access	Open access for Educational Purposes
Cost	Free
Observations	-
Web page	https://www.ordnancesurvey.co.uk/business-and- government/products/imagery-layer.html

Table A.8. Data Profile: MasterMap Imagery Layer

Structure of the Data

It contains 1km by 1km tiles imagery. The following Table shows files format included in this data.

Table A.9. Georeferencing files

Image Format	Georeferencing file	Description	Example
TIF	*.TFW	TIFF world file states the X,Y coordinate of the centre of the north-west pixel	0.250000 0.000000 0 -0.250000 297000 125000 570999.875000

¹⁴ This document is a summary of relevant elements contained in the "OS MasterMap Imagery Layer product guidev2.1 07/2017"

JPEG	*.JGW	JPEG world file states the X,Y coordinate of the centre of the north-west pixel	0.250000 0 -0.250000 467000.125000 989999.875000
ECW	*.EWW	ECW world file states the X,Y coordinate of the centre of the north-west pixel	0.250000 0 0 -0.250000 176000.125000 827999.875000

Relevant Documents

The following documents are included in the folder attached to this document.

rabio / iror resolution indetermap inagery Layer	
Document Name	File Name
Imagery layer product guide	os-mastermap-imagery-layer-product-guide.pdf
Imagery-layer-technical-	os-mastermap-imagery-layer-technical-specification.pdf
specification	

Table A.10. Relevant Documents: MasterMap Imagery Layer

Quality

There are several features of this data that makes it an invaluable resource for crosssectional and longitudinal geospatial analysis. By using a unique identifier when representing real-world features across all Great Britain it allows researchers to consistently analyse cross-sectional and longitudinal changes of real-world objects. Besides, several steps are performed by the Ordnance Survey to continually update the data and assess its quality. The Ordnance Survey also provides all the technical documents and Metadata to evaluate the quality of the data.

Data Profile 5: MasterMap Water Network

Name of the Data: MasterMap Water Network¹⁵

Description of the Data

OS MasterMap Water Network is a three-dimensional digital representation of the watercourses in Great Britain. It includes rivers, streams, lakes, lochs and canals as a series of watercourse network lines. The network lines (links) are attributed to provide a range of information about the section of watercourse they depict.

Structure of the Data

Data Profile

OS MasterMap Water Network features are classified into feature types. Each feature type has associated attribution and further detail can be found in the Technical Specification.

The OS MasterMap Water Network as currently published consist of the following core features:

- WatercourseLink: A WatercourseLink feature represents the alignment of a watercourse; the majority have been derived from Ordnance Survey detailed topographic data, made available to customers as the OS MasterMap Topography Layer.
- HydroNode: A HydroNode feature explicitly represents the source, sink, or junction of a watercourse. In addition, it recodes the location where any real world related attribution change.

Table A.T. Data Frome. Mastermap water Network	
Field	Description
Dataset ID	5
Dataset Name	MasterMap Water Network Layer
Temporal Extent	2016 – 2019
Supplier/data owner	Ordnance Survey
Analytical Units	Points, Lines
Data Format	Geodatabases (GDB), Shapefile (SHP)
Geographical Extent	Great Britain
Access	Open access for Educational Purposes
Cost	Free
Observations	-
Web page	https://www.ordnancesurvey.co.uk/business-and- government/products/os-mastermap-water-network.html

Table A.11. Data Profile: MasterMap Water Network

Relevant Documents

 $^{^{15}}$ This document is a summary of relevant elements contained in the "OS MasterMap® Water Network product guide os mastermap water network v1.1 – 06/2015
The following documents are included in the folder attached to this document.

Document Name	Name file
Water network getting started guide	os-mastermap-water-network-getting-started-guide.pdf
Water network product guide	os-mastermap-water-network-product-guide.pdf
Water network technical specification	os-mastermap-water-network-technical-specification.pdf
Water network release note	osmm-water-network-release-note-apr19.pdf

Table A.12. Relevant Documents: MasterMap Water Network

Quality

There are several features of this data that makes it an invaluable resource for crosssectional and longitudinal geospatial analysis. By using a unique identifier when representing real-world features across all Great Britain it allows researchers to consistently analyse cross-sectional and longitudinal changes of real-world objects. Besides, several steps are performed by the Ordnance Survey to continually update the data and assess its quality. The Ordnance Survey also provides all the technical documents and Metadata to evaluate the quality of the data.

Data Profile 6: MasterMap Greenspace Layer

Name of the Data: MasterMap Greenspace Layer¹⁶

Description of the Data

OS MasterMap Greenspace Layer gives a comprehensive view of the greenspaces within an urban area. The dataset comprises of topographic areas published in OS MasterMap Topography Layer, with additional greenspace-specific attribution to describe their function. It includes both publicly accessible and private green spaces, sports facilities and natural environment features, to give a comprehensive view of spaces that are important for applications such as, environmental management, health, planning or habitat mappings.

Data Profile

Field	Description
Dataset ID	6
Dataset Name	MasterMap Green Space Layer
Temporal Extent	1999 – 2019
Supplier/data	Ordnance Survey
owner	
Analytical Units	Points, Lines, Polygons
Data Format	Geodatabases (GDB), Shapefile (SHP)
Geographical Extent	Great Britain
Access	Open access for Educational Purposes
Cost	Free
Observations	-
Web page	https://www.ordnancesurvey.co.uk/business-and-
	government/products/os-mastermap-greenspace.html

Table A.13. Data Profile: MasterMap Greenspace Layer

Structure of the Data

Features

OS MasterMap Greenspace contains one feature type called GreenspaceArea.

The MasterMap Greenspace Layer is comprised of a subset of the TopographicArea polygons from OSMM Topography Layer (which represent topographic objects that have a polygon-based geometry). Only polygons which have been classified as a form of greenspace are supplied. This subset of polygons have the Topographic Identifier (TOID) and Version number from Topography Layer (to enable the datasets to be used in conjunction) plus additional attributes providing the greenspace specific information.

¹⁶ This document is a summary of relevant elements contained in the "OS MasterMap Greenspace Layer Technical Specification V1.0 May"

The greenspace specific attribution within this product consists of the form and function of the greenspace area. The functions included within OS MasterMap Greenspace Layer describe the purpose of the Greenspace areas.

Attributes

- TOID: The unique reference number of the feature in OSMM Topography Layer. Combined with the version number, this enables joins with the Topography Layer dataset.
- Version: The version number of the feature in OSMM Topography Layer. This identifies the specific version of Topography Layer feature this dataset was created from. Combined with the TOID, this enables joins with the Topography Layer dataset.
- Primary Function: The main function of the greenspace area. Functions are determined from a specific greenspace list, and only features with a greenspace function have been included in this product. A list of the possible functions for features within this dataset can be found below.
- Secondary Function: Where a polygon has more than one greenspace function present, this field will be populated with the second function. These functions come from the same list as the Primary Functions. The order of the Primary and Secondary Functions is predetermined by a hierarchy (shown in brackets in the below list) and does not indicate an order of importance or scale.
- Primary Form: The main form of the greenspace area. Forms are determined from a specific greenspace list.
- Secondary Form: Where a polygon has more than one greenspace form present, this field will be populated with the second form. These forms come from the same list as the Primary Forms. The order of the Primary and Secondary Forms is predetermined by a hierarchy (shown in brackets in the below list) and does not indicate an order of importance or scale.

OSMM Greenspace Layer includes the following functions (the population of Primary Function and Secondary Function attributes is driven by a hierarchy detailed on brackets):

- Allotments or Community Growing Spaces (14)
- Amenity Residential or Business (5)
- Amenity Transport (6)
- Land Use Changing (17)
- Bowling Green (16)
- Camping or Caravan Park (7)
- Cemetery (9)
- Golf Course (4)
- Institutional Grounds (3)
- Natural (18)
- Other Sports Facility (12)
- Religious Grounds (8)
- Play Space (15)
- Playing Field (11)
- Private Garden (10)
- Public Park or Garden (1)
- School Grounds (2)
- Tennis Court (13)

The forms included within OS MasterMap Greenspace Layer are below (form hierarchy detailed on brackets);

- Woodland (3)
- Open Semi-Natural (1)
- Inland Water (2)
- Beach or Foreshore (4)
- Manmade Surface (5)
- Multi Surface (6)

Relevant Documents

The following documents are included in the folder attached to this document.

	······································
Document Name	Name file
Greenspace layer release	os-mastermap-greenspace-layer-release-note-oct-
note	2018.pdf
Greenspace getting started	osmm-greenspace-getting-started-guide.pdf
guide	
Greenspace product guide	osmm-greenspace-product-guide.pdf
Greenspace technical	osmm-greenspace-technical-specification.pdf
specification	

Table A.14. Relevant Documents: MasterMap Greenspace Layer

Quality

There are several features of this data that makes it an invaluable resource for crosssectional and longitudinal geospatial analysis. By using a unique identifier when representing real-world features across all Great Britain it allows researchers to consistently analyse cross-sectional and longitudinal changes of real-world objects. Besides, several steps are performed by the Ordnance Survey to continually update the data and assess its quality. The Ordnance Survey also provides all the technical documents and Metadata to evaluate the quality of the data.

Ordnance Survey is committed to maintaining its products to the highest levels of accuracy and currency. The initial capture of data for MasterMap Greenspace will use OS existing topographic databases and aerial imagery. As such the quality of the data will be constrained to what can be achieved with this approach. For example, where an access into a site is obscured (e.g. under trees) it will not be captured. In addition, the use of our existing databases to identify the location of sites of interest means that we cannot guarantee that all relevant sites will be included in the data. However, where we are informed and can verify that a feature is missing or inaccurately depicted in the dataset, we will make the necessary amendments to the dataset within 12 months of such verification.

The Greenspace Feedback Tool is used to report errors and omissions in MasterMap Greenspace. It can also be used to track the progress of error reports. Access to the Greenspace Feedback Tool is available for public sector organisations only.

Data Profile 7: AddressBase Premium

Name of the Data: AddressBase Premium¹⁷

Description of the Data

The AddressBase product family is made up of three products:

- AddressBase,
- AddressBase Plus,
- AddressBase Premium

AddressBase products are created by bringing together different address sources:

- Local Authority Gazetteers for England, Wales and Scotland,
- Royal Mail® PAF® data,
- References to Valuation Office Agency (VOA) data,
- Additional addresses and coordinates from Ordnance Survey.

This information is managed by GeoPlace under a joint venture partnership between Local Government Agency and Ordnance Survey. Scottish Local Authority address information is supplied to GeoPlace under licence between Ordnance Survey and the Improvement Service supported by the Scottish Government. Ordnance Survey is responsible for customer management, sales, marketing and distribution of the AddressBase products.

Data Profile

Field	Description
Dataset ID	7
Dataset Name	AddressBase Premium
Temporal Extent	1970 – 2019
Supplier/data owner	Ordnance Survey
Analytical Units	Points
Data Format	Geodatabases (GDB), Shapefile (SHP)
Geographical Extent	Great Britain
Access	Open access for Educational Purposes
Cost	Free
Observations	Circa 43 million features
Web page	https://www.ordnancesurvey.co.uk/business-and- government/products/addressbase-premium.html

Table A.15. Data Profile: AddressBase Premium

¹⁷ This document is a summary of relevant elements contained in the "AddressBase Product Guide – Version 2.0 – 03/2016"

Structure of the Data

Product name	AddressBase Premium	AddressBase Plus	AddressBase
Features	UPRN	UPRN	UPRN
	Current / Live addresses	Current / Live	Royal Mail Postcode
		addresses	Address File
	Alternative addresses	Royal Mail Postcode	National Grid
		Address File	coordinates
	Provisional addresses	National Grid	Latitude and
		coordinates	Longitude coordinates
	Historical addresses	Latitude and Longitude	Primary level of classification
		coordinates	
	Royal Mail Postcode	Four levels of	
	Address File	classification	
	National Grid coordinates	Feature life cycle dates	
	Latitude and Longitude coordinates	Local authority addresses	
	Four levels of classification	OWPA records	
	Valuation Office Agency	Multiple occupancy	
	classification scheme	addresses	
	Feature life cycle dates	Local authority street information	
	Local authority addresses	USRN (Unique Street Reference Number)	
	OWPA records	References to OS MasterMap	
	Multiple occupancy addresses	Topography Layer TOIDs	
	Local authority street information	References to	
	USRN (Unique Street Reference Number)	OS MasterMap Integrated Transport Network Layer TOIDs	
	Street start and end coordinates	Parent/child relationships	
	References to OS	Valuation Office	
	MasterMap	Agency cross references	
	Topography Laver TOIDs		
	References to		
	OS MasterMap Integrated		
	TOIDS		
	Parent/child relationships		
	Valuation Office Agency cross references		

Table A.16. Explanation of AddressBase products

Relevant Documents

The following documents are included in the folder attached to this document.

Document Name	Name file
Addressbase premium	addressbase-premium-technical-specification.pdf
technical specification	
Addressbase product guide	addressbase-product-guide.pdf
Addressbase products	addressbase-products-getting-started-guide.pdf
getting started guide	
Addressing products attribute	os-addressing-products-attribute-mapping.pdf
mapping	

Table	A.17	Relevant	Documents:	AddressBase	Premium
IUDIC	~	Itele vant	Dooumento.	Addicoobdoc	1 I CHIIGHI

Quality

There are several features of this data that makes it an invaluable resource for crosssectional and longitudinal geospatial analysis. By using a unique identifier when representing real-world features across all Great Britain it allows researchers to consistently analyse cross-sectional and longitudinal changes of real-world objects. Besides, several steps are performed by the Ordnance Survey to continually update the data and assess its quality. The Ordnance Survey also provides all the technical documents and Metadata to evaluate the quality of the data.

The products AddressBase, AddressBase Plus, AddressBase Premium have been designed to meet distinct customer requirements. The source data is collated, verified and quality assured by GeoPlace. GeoPlace is a limited liability partnership jointly owned by the Local Government Agency and Ordnance Survey and is the centre of excellence for spatial address and street information in Great Britain.

Data Profile 8: MasterMap Points of Interest

Name of the Data: MasterMap Points of Interest¹⁸

Description of the Data

Points of Interest (POI) is a dataset containing around 4 million different geographic features. All features are supplied with location, functional information and addresses where possible. It has been available continuously since 2002. The product covers all of Great Britain.

Data Profile

Field	Description
Dataset ID	8
Dataset Name	Points of Interest
Temporal Extent	2005 – 2019
Supplier/data owner	Ordnance Survey
Analytical Units	Points
Data Format	Geodatabases (GDB), Shapefile (SHP)
Geographical Extent	Great Britain
Access	Open access for Educational Purposes
Cost	Free
Observations	Circa 4 million commercial/non-residential features
Web page	https://www.ordnancesurvey.co.uk/business-and- government/products/points-of-interest.html

Table A.18. Data Profile: MasterMap Points of Interest

Structure of the Data

POI has a three-level classification to assist customers in identifying the features or sets of features they require. The first level of classification, comprising nine groups, is given below.

Table A.19. Summary of MasterMap Points of Interest group categories

Group Cod	Group description	Count	%
1	Accommodation, eating and drinking	195,779	4,8
2	Commercial services	768,859	18,8
3	Attractions	689,037	16,8
4	Sport and entertainment	123,469	3,0
5	Education and health	190,803	4,7
6	Public infrastructure	670,624	16,4
7	Manufacturing and production	418,168	10,2
9	Retail	341,155	8,3

¹⁸ This document is a summary of relevant elements contained in the "Points of Interest user guide chapter 1 v3.3 - 12/2014"

10	Transport	694,240	17,0
	Total general	4,092,134	100,0

Relevant Documents

The following documents are included in the folder attached to this document.

Table A.20. Relevant Documents: MasterMap Points of Interest

Document Name	Name file
Points of Interest frequencies	D8.OS Points of Interest.xls
Points of interest class count	points-of-interest-class-count.pdf
Points of interest	points-of-interest-classification-scheme.pdf
classification scheme	
Points of interest user guide	points-of-interest-user-guide.pdf

Quality

There are several features of this data that makes it an invaluable resource for crosssectional and longitudinal geospatial analysis. By using a unique identifier when representing real-world features across all Great Britain it allows researchers to consistent analyse crosssectional and longitudinal changes of real-world objects. Besides, several steps are performed by the Ordnance Survey to continually update the data and assess its quality. The Ordnance Survey also provides all the technical documents and Metadata to evaluate the quality of the data.

Points of Interest is created and maintained by PointX, an independent company in the joint ownership of Ordnance Survey and Landmark® Information Group. Ordnance Survey is the sole 'data only' distributor for Points of Interest.

PointX collects data from around 150 suppliers, including Ordnance Survey. It receives updates from these suppliers on an ongoing basis. The suppliers are chosen for being the most authoritative source or sources for the particular type of feature they supply and for the quality and completeness of the data they supply. It should be noted, however, that for many types of feature there is no absolute, definitive source, nor can any list of the types of features found in Points of Interest be said to be complete.

Data Profile 9: HM Land Registry Open Data

Name of the Data: HM Land Registry Open Data¹⁹

Description of the Data

HM Land Registry publishes the following public datasets:

- Price paid data: updated monthly, data available from 1995.
- Transaction data: updated monthly, data available from December 2011.
- UK House Price Index data: updated monthly, data available from January 1995. HM Land Registry publishes the UK House Price Index on behalf of Office for National Statistics, Registers of Scotland and Land and Property Services Northern Ireland.

Data Profile

Table A.21. Data Profile: HM Land Registry Open Data

Field	Description
Dataset ID	9
Dataset Name	Land Registry Price Paid dataset
Temporal Extent	1995 – 2019
Supplier/data owner	Land Registry
Analytical Units	Points (Postcodes)
Data Format	CSV
Geographical Extent	Great Britain
Access	Open access
Cost	Free
Observations	Circa 25 million records
Web page	http://landregistry.data.gov.uk/

Structure of the Data

1. Price paid data

Price Paid Data tracks property sales in England and Wales submitted to HM Land Registry for registration. Price Paid Data is based on the raw data released each month. The amount of time between the sale of a property and the registration of this information with HM Land Registry varies. It typically ranges between 2 weeks and 2 months. This excludes some data for example, all commercial transactions, sales under a court order, gifts.

¹⁹ This document is a summary of relevant elements contained in the link https://www.gov.uk/government/organisations/land-registry/

Data item	Explanation (where appropriate)
Transaction unique identifier	A reference number, which is, generated automatically
	recording each published sale. The number is unique
	and will change each time a sale is recorded.
Price	Sale price stated on the transfer deed.
Date of Transfer	Date when the sale was completed, as stated on the
	transfer deed.
Postcode	This is the postcode used at the time of the original
	transaction. Note that postcodes can be reallocated
	and these changes are not reflected in the Price Paid
	Dataset.
Property Type	D = Detached, S = Semi-Detached, T = Terraced, F =
	Flats/Maisonettes, O = Other Note that:
	 we only record the above categories to describe
	property type, we do not separately identify
	bungalows.
	- end-of-terrace properties are included in the
	l erraced category above.
	- Other is only valid where the transaction relates to a
	property type that is not covered by existing values.
Old/INEW	nuccates the age of the property and applies to all
	price paid transactions, residential and non-residential. V = a nowly built property $N = an established$
	residential building
Duration	Polotoo to the tenuro: E Freehold L Lessehold etc.
Duration	Relates to the tenure. $F = Freehold$, $L = Leasehold etc.$
	7 years or less in the Price Paid Dataset
BAON	Primary Addressable Object Name, Typically the
FAON	house number or name
SAON	Secondary Addressable Object Name Where a
SACIN	property has been divided into separate units (for
	example flats) the PAON (above) will identify the
	building and a SAON will be specified that identifies
	the separate unit/flat.
Street	Street
Locality	Locality
Town/City	Town/City
District	District
County	County
PPD	Indicates the type of Price Paid transaction.
Category Type	A = Standard Price Paid entry. includes single
	residential property sold for full market value. B =
	Additional Price Paid entry including transfers under a
	power of sale/repossessions, buy-to-lets (where they
	can be identified by a Mortgage) and transfers to non-
	private individuals.
	Note that category B does not separately identify the
	transaction types stated.
	HM Land Registry has been collecting information on
	Category A transactions from January 1995. Category
	B transactions were identified from October 2013.

Table A.22. Description of variables: 'Price Paid' data

Record Status - monthly file only	Indicates additions, changes and deletions to the records.(see guide below). A = Addition C = Change D = Delete. Note that where a transaction changes category type due to misallocation (as above) it will be deleted from the original category type and added to the correct category with a new transaction unique identifier.

2. Transaction data

Transaction Data includes information about the number and types of applications that have been completed. The data reflects the volume of applications lodged by customers using an HM Land Registry account number on their application form.

Structure of the Data

Acronym	Title	Description
DFL	Dispositionary first lease	An application for the registration of a new lease granted by the proprietor of registered land
DLG	Dealing	An application in respect of registered land. This includes transfers of title, charges and notices
FR	First registration	An application for a first registration of land both freehold and leasehold. For leasehold this applies when the landlord's title is not registered
ТР	Transfer of part	An application to register the transfer of part of a registered title
OS(W)	Search of whole	An application to protect a transaction for value, such as purchase, lease or charge for the whole of a title
OS(P)	Search of part	An application to protect a transaction for value, such as purchase, lease or charge for part of a title
OS(NPW)	Non-priority search of whole	An application to search the whole of the register without getting priority
OS(NPP)	Non-priority search of part	An application to search a part of the register without getting priority
OC1	Official copy	An application to obtain an official copy of a register or title plan represents a true record of entries in the register and extent of the registered title at a specific date and time. The data includes historical editions of the register and title plan where they are kept by the registrar in electronic form
OC2	Official copy of a deed or document	An application to obtain a copy of a document referred to in the register or relates to an application. This includes correspondence, surveys, application forms and emails relating to applications that are pending, cancelled or completed
SIM	Search of the index map	An application to find out whether or not land is registered and, if so, to obtain the title number

Table A.23. Description of variables: 'Transaction' data

3. UK House Price Index

The UK House Price Index (UK HPI) captures changes in the value of residential properties. The UK HPI uses sales data collected on residential housing transactions, whether for cash or with a mortgage. Properties have been included in:

- England and Wales since January 1995
- Scotland since January 2004
- Northern Ireland since January 2005

Data is available at a national and regional level, as well as counties, local authorities and London boroughs.

Structure of the Data:

Column Header	Explanation
Date	The year and month to which the monthly statistics apply
RegionName	Name of geography (Country, Regional, County/Unitary/District Authority and London Borough)
AreaCode	Code of geography (Country, Regional, County/Unitary/District Authority and London Borough)
Average Price	Average house price for a geography in a particular period
Index	House price index for a geography in a particular period (January 2015=100).
IndexSA	Seasonally adjusted house price for a geography in a particular period (January 2015=100).
1m%change	The percentage change in the Average Price compared to the previous month
12m%change	The percentage change in the Average Price compared to the same period twelve months earlier.
AveragePricesSA	Seasonally adjusted Average Price for a geography in a particular period
Sales Volume	Number of registered transactions for a geography in a particular period
[Property Type]Price	Average house price for a particular property type (such as detached houses), for a geography in a particular period.
[Property Type]Index	House price index for a particular property type (such as detached houses), for a geography in a particular period (January 2015=100).
[Property Type]1m%change	The percentage change in the [Property Type] Price (such as detached houses) compared to the previous month
[Property Type]12m%change	The percentage change in the [Property Type] Price (such as detached houses) compared to the same period twelve months earlier.
[Cash/Mortgage]Price	Average house price by funding status (such as cash), for a geography in a particular period.

Table A.24. Description of variables: 'UK House Price Index' data

[Cash/Mortgage]Index	House price index by funding status (such as cash) for
[Cash/Mongaye]index	a geography in a particular period (January 2015=100).
[Cash/Mortgage]1m%change	The percentage change in the [Cash/Mortgage]Price compared to the previous month
[Cash/Mortgage]12m%change	The percentage change in the [Cash/Mortgage]Price compared to the same period twelve months earlier.
[Cash/Mortgage] Sales Volume	Number of registered transactions [Cash/Mortgage] for a geography in a particular period
[FTB/FOO]Price	Average house price by buyer status (such as first time buyer/former owner occupier), for a geography in a particular period.
[FTB/FOO]Index	House price index by buyer status (such as first time buyer/former owner occupier), for a geography in a particular period. (January 2015=100).
[FTB/FOO]1m%change	The percentage change in the [<u>FTB/FOO]</u> Price compared to the previous month
[FTB/FOO]12m%change	The percentage change in the [<u>FTB/FOO</u>]Price compared to the same period twelve months earlier.
[New/Old]Price	Average house price by property status (such as new or existing property), for a geography in a particular period.
[New/Old]Index	House price index by property status (such as new or existing property), for a geography in a particular period. (January 2015=100).
[New/Old]1m%change	The percentage change in the [New/Old]Price compared to the previous month
[New/Old]12m%change	The percentage change in the [New/Old]Price compared to the same period twelve months earlier.
[New/Old] Sales Volume	Number of registered transactions [New/Old] for a geography in a particular period

Relevant Documents

1. House Price Index

The following documents are included in the folder attached to this document.

Table A.25. Relevant Documents: HM Land Registry Open Data – House Price Index

Document Name	Name file
About the UK House Price	About the UK House Price Index - GOV.UK.pdf
Index	
UK House Price Index	UK House Price Index_ reports - GOV.UK.pdf
reports	

2. Price paid data

The following documents are included in the folder attached to this document.

Document Name	Name file
How to access HM Land	How to access HM Land Registry Price Paid Data -
Registry Price Paid Data	GOV.UK.pdf
Privacy Impact Assessment	PIA_Report_6_13.pdf
Review	
Privacy Impact Assessment	ppd_pia.pdf
Report	
Price Paid Data	Price Paid Data - GOV.UK.pdf
Price Paid Data Standard	Price Paid Data_ Standard reports - GOV.UK.pdf
reports	
Privacy Impact Assessment	Privacy Impact Assessment review_ Price Paid Data -
review: Price Paid Data	GOV.UK.pdf
Public data	Public data - GOV.UK

Table A.26. Relevant Documents: HM Land Registry Open Data – Paid Price Data

3. Transaction Data

The following documents are included in the folder attached to this document.

Table A.27	7. Relevant	Documents: HN	l Land	d Registry Open	Data –	Transaction	Data

Document Name	Name file
HM Land Registry	HM Land Registry_ Transaction Data - GOV.UK.pdf
Transaction Data	

Quality

1. House Price Index²⁰

The UK HPI is not as timely in publishing as other house price index measures published in the UK because it is based on completed sales at the end of the conveyancing process, rather than advertised or approved prices.

The UK HPI has wide coverage of both cash and mortgage transactions and a large data source (land registrations such as that maintained by HM Land Registry) allowing data to be published down to a local authority level with further breakdowns available by property type, buyer status, funding status and property status.

Estimates for the most recent months are provisional and are likely to be updated as more data is incorporated into the index. While changes to estimates are small at the headline level, these can be larger at lower geographies due to the fewer transactions used. Further details on why estimates change can be found in the monthly revisions policy documents (see https://www.gov.uk/government/publications/about-the-uk-house-price-index/about-the-uk-house-price-in

2. Price paid data²¹

²⁰ This section is extracted from the document "About the UK House Price Index"

²¹ This section is extracted from the document "Public data - GOV.UK.pdf"

The data published has been collected from a variety of sources as part of the land registration process. The HM Land Registry have developed internal quality processes to help ensure the data correctly reflects the information provided to them on registration.

The HM Land Registry use reasonable skill and care in the provision of the data, but cannot guarantee the data is fit for particular intended purposes or uses.

Price Paid Data provides information on single residential property sales in England and Wales, sold for full market value, that are lodged with the HM Land Registry for registration. The data is based on information provided to the HM Land Registry and may not reflect the true value of a property at the time of sale. HM Land Registry does not guarantee or hold that the data is fit for particular purposes or uses.

A title number is HM Land Registry's reference given to an individual piece of registered land. The title provides an up-to-date record of the legal ownership of the land or property in that title. The ownership of land can be either freehold or leasehold. A single title may comprise several individual plots in different ownership and may contain multiple individual properties. These can consist of freehold, leasehold and under leases, all or some of which, may be registered. Not all property ownership needs to be registered, and this includes leases for a term of seven years or less. A title number will therefore not necessarily show all the ownerships associated with the land or properties within that title record.

3. Transaction Data²²

HM Land Registry use reasonable skill and care in the provision of the data and strive to ensure our data is as accurate as possible but cannot guarantee it is free from error. HM Land Registry cannot guarantee our data is fit for particular intended purposes or uses.

²² This section is extracted from the document "HM Land Registry_ Transaction Data - GOV.UK"

Data Profile 10: Energy Performance Certificate (EPC)

Name of the Data: Energy Performance Certificate (EPC)²³

Description of the Data

The data is drawn from Energy Performance Certificates (EPCs) issued for domestic and non-domestic buildings constructed, sold or let since 2008. Data from Display Energy Certificates (DECs) issued for buildings occupied by public authorities also dates back to 2008. It provides information on the energy efficiency ratings of domestic and non-domestic buildings during the energy assessment process. This data release includes data for buildings with multiple EPCs or DECs as well as for buildings where only a single energy certificate has been issued. The registers do not hold data for every domestic and non-domestic building or every building occupied by public authorities in England and Wales. This data should, therefore, not be interpreted as a true representation of the whole of the building stock in England and Wales, but viewed as part of a wider package of Government's provision of information on the energy efficiency of buildings.

Data Profile

Field	Description
Dataset ID	10
Dataset Name	Energy Performance Certificate data
Temporal Extent	2008 – 2016
Supplier/data owner	Communities, Housing and Local Government
Analytical Units	Points, Polygons (Postcodes)
Data Format	CSV
Geographical Extent	Great Britain
Access	Once registred on the webpage, all data fields other than the address and postcode data are Open Access
Cost	Free
Observations	Circa 14 million records
Web page	https://epc.opendatacommunities.org/

Table A.28. Data Profile: Energy Performance Certificate (EPC)

Structure of the Data

The data is divided in three main categories:

- 1. Domestic EPCs
- 2. Non Domestic EPCs
- 3. Display Energy Certificates (DECs)

²³ This document is a summary of relevant elements contained in the link https://epc.opendatacommunities.org/

1. Domestic EPCs

Table A.29. Description o	f variables: Domestic Energy Performance Certificate
Name V/entelale	Description

Name Variable	Description
LMK_KEY	Individual lodgement identifier. Not
	guaranteed to be unique. See also
	Certificate Hash. We're making some
	changes soon, so that this will be unique
	Read more.
ADDRESS1	First line of the address
ADDRESS2	Second line of the address
ADDRESS3	Third line of the address
POSTCODE	The postcode of the property
BUILDING_REFERENCE_NUMBER	Unique identifier for the building
CURRENT_ENERGY_RATING	Current energy rating converted into a
	linear 'A to G' rating (where A is the most
	energy e client and G is the least energy e
DOTENTIAL ENERGY DATING	client)
POTENTIAL_ENERGY_RATING	Estimated potential energy rating converted
	Into a linear A to G rating (where A is the
	anoray o client)
CURRENT ENERGY EFFICIENCY	Based on cost of energy i.e. energy
	required for space heating, water heating
	and lighting [in kWh/year] multiplied by fuel
	costs. $(f/m^2/vear where cost is derived from$
	kWh).
POTENTIAL_ENERGY_EFFICIENCY	Summary of the potential environmental
	efficiency of the property feature
PROPERTY_TYPE	Describes the type of property such as
	House, Flat, Mansion, Maisonette etc. This
	is actually the type differentiator for
	Property but only a limited number of
	property types, notably Apartment and
	Apartment Block, have any specific
	characteristics and warrant their own
	definition.
	Detached Semi-Detached Terrace etc.
	Together with the Property Type, the Build
	Form produces a structured description of
	the property
INSPECTION DATE	The date that the inspection was actually
	carried out by the energy assessor
LOCAL AUTHORITY	Office for National Statistics (ONS) code.
_	Local authority area in which the building is
	located.
CONSTITUENCY	Office for National Statistics (ONS) code.
	Parliamentary constituency in which the
	building is located.
COUNTY	County in which the building is located
	(where applicable)

LODGEMENT_DATE	Date lodged on the Energy Performance of Buildings Register
TRANSACTION_TYPE	Type of transaction that triggered EPC. For example, one of: marketed sale; non- marketed sale; rental; not sale or rental; assessment for Green Deal; following Green Deal; FIT application; none of the above; RHI application; ECO assessment. Where the reason for the assessment is
	unknown by the energy assessor the transaction type will be recorded as 'none of the above'. Transaction types may be changed over time.
ENVIRONMENT_IMPACT_CURRENT	The Environmental Impact Rating. A measure of the property's current impact on the environment in terms of carbon dioxide
	(CO_2) emissions. The higher the rating the
	lower the CO_2 emissions. (CO_2 emissions in toppes (year)
ENVIRONMENT_IMPACT_POTENTIAL	The potential Environmental Impact Rating. A measure of the property's potential impact on the environment in terms of
	carbon dioxide (CO ₂) emissions after
	improvements have been carried out. The
	higher the rating the lower the CO_2
	emissions. (CO_2 emissions in tonnes / year)
ENERGY_CONSUMPTION_CURRENT	Estimated total energy consumption for the Property in a 12 month period. Value is Kilowatt Hours per Square Metre (kWh/m ²)
ENERGY_CONSUMPTION_POTENTIAL	Estimated potential total energy consumption for the Property in a 12 month period. Value is Kilowatt Hours per Square Metre (kWh/m ²)
CO2_EMISSIONS_CURRENT	CO ₂ emissions per year in tonnes/year.
CO2_EMISS_CURR_PER_FLOOR_AREA	CO ₂ emissions per square metre floor area per year in kg/m ²
CO2_EMISSIONS_POTENTIAL	Estimated value in Tonnes per Year of the
	total CO ₂ emissions produced by the
	Property in 12 month period.
LIGHTING_COST_CURRENT	GBP. Current estimated annual energy costs for lighting the property.
LIGHTING_COST_POTENTIAL	GBP. Potential estimated annual energy costs for lighting the property after improvements have been made.
HEATING_COST_CURRENT	GBP. Current estimated annual energy costs for heating the property.
HEATING_COST_POTENTIAL	GBP. Potential annual energy costs for lighting the property after improvements have been made.

HOT_WATER_COST_CURRENT	GBP. Current estimated annual energy
HOT WATER COST DOTENTIAL	CPR Detential estimated appual aperav
HOT_WATER_COST_POTENTIAL	GBP. Polential estimated annual energy
TOTAL FLOOD ADEA	The total wasful flags area is the total of all
TOTAL_FLOOR_AREA	The total useful noor area is the total of all
	enclosed spaces measured to the internal
	fleer erec of measured in coordenes with
	the guideness included from time to time by
	the Boyel Institute of Chartered Surveyore
	the Royal institute of Chartered Surveyors
	Turne of electricity to riff for the property of a
	single.
MAINS_GAS_FLAG	Whether mains gas is available. Yes means
	that there is a gas meter or a gas- burning
	appliance in the dwelling. A closed-o" gas
	pipe does not count.
FLOOR_LEVEL	Flats and maisonettes only. Floor level
	relative to the lowest level of the property (0
	for ground floor). If there is a basement, the
	basement is level 0 and the other floors are
	from 1 upwards
FLAT_TOP_STOREY	Whether the flat is on the top storey
FLAT_STOREY_COUNT	The number of Storeys in the Apartment
	Block.
MAIN_HEATING_CONTROLS	Type of main heating controls. Includes
	both main heating systems if there are two.
MULTI_GLAZE_PROPORTION	The estimated banded range (e.g. 0% -
	10%) of the total glazed area of the
	Property that is multiple glazed.
GLAZED_TYPE	The type of glazing. From British
	Fenestration Rating Council or
	manufacturer declaration, give as one of;
	single; double; triple.
GLAZED_AREA	Ranged estimate of the total glazed area of
	the Habitable Area.
EXTENSION_COUNT	I ne number of extensions added to the property. Between 0 and 4
NUMBER HABITABLE ROOMS	Habitable rooms include any living room
	sitting room dining room bedroom study
	and similar: and also a non-separated
	conservatory. A kitchen/diner having a
	discrete seating area (with space for a table
	and four chairs) also counts as a habitable
	room. A non-separated conservatory adds
	to the habitable room count if it has an
	internal quality door between it and the
	dwelling. Excluded from the room count are
	any room used solely as a kitchen. utility
	room, bathroom. cloakroom. en-suite
	accommodation and similar: any hallway
	stairs or landing; and also any room not
	having a window.

NUMBER_HEATED_ROOMS	The number of heated rooms in the property if more than half of the habitable rooms are not heated.
LOW_ENERGY_LIGHTING	The percentage of low energy lighting present in the property as a percentage of the total fixed lights in the property. 0% indicates that no low-energy lighting is present.
NUMBER_OPEN_FIREPLACES	The number of Open Fireplaces in the Property. An Open Fireplace is a fireplace that still allows air to pass between the inside of the Property and the outside.
HOTWATER_DESCRIPTION	Overall description of the property feature
HOT_WATER_ENERGY_EFF	Energy efficiency rating. One of: very good; good; average; poor; very poor. On actual energy certificate shown as one to five star rating.
HOT_WATER_ENV_EFF	Environmental efficiency rating. One of: very good; good; average; poor; very poor. On actual energy certificate shown as one to five star rating.
FLOOR_DESCRIPTION	Overall description of the property feature
FLOOR_ENERGY_EFF	Energy efficiency rating. One of: very good; good; average; poor; very poor. On actual energy certificate shown as one to five star rating.
FLOOR_ENV_EFF	Environmental efficiency rating. One of: very good; good; average; poor; very poor. On actual energy certificate shown as one to five star rating.
WINDOWS_DESCRIPTION	Overall description of the property feature
WINDOWS_ENERGY_EFF	Energy efficiency rating. One of: very good; good; average; poor; very poor. On actual energy certificate shown as one to five star rating.
WINDOWS_ENV_EFF	WINDOWS. Environmental efficiency rating. One of: very good; good; average; poor; very poor. On actual energy certificate shown as one to five star rating.
WALLS_DESCRIPTION	Overall description of the property feature
WALLS_ENERGY_EFF	Energy efficiency rating. One of: very good; good; average; poor; very poor. On actual energy certificate shown as one to five star rating.
WALLS_ENV_EFF	Environmental efficiency rating. One of: very good; good; average; poor; very poor. On actual energy certificate shown as one to five star rating.
SECONDHEAT_DESCRIPTION	Overall description of the property feature
SHEATING_ENERGY_EFF	Energy efficiency rating. One of: very good; good; average; poor; very poor. On actual energy certificate shown as one to five star rating.

SHEATING_ENV_EFF	Environmental efficiency rating. One of: very good; good; average; poor; very poor. On actual energy certificate shown as one to five star rating.
ROOF_DESCRIPTION	Overall description of the property feature
ROOF_ENERGY_EFF	Energy efficiency rating. One of: very good; good; average; poor; very poor. On actual energy certificate shown as one to five star rating.
ROOF_ENV_EFF	Environmental efficiency rating. One of: very good; good; average; poor; very poor. On actual energy certificate shown as one to five star rating.
MAINHEAT_DESCRIPTION	Overall description of the property feature
MAINHEAT_ENERGY_EFF	Energy efficiency rating. One of: very good; good; average; poor; very poor. On actual energy certificate shown as one to five star rating.
MAINHEAT_ENV_EFF	Environmental efficiency rating. One of: very good; good; average; poor; very poor. On actual energy certificate shown as one to five star rating.
MAINHEATCONT_DESCRIPTION	Overall description of the property feature
MAINHEATC_ENERGY_EFF	Energy efficiency rating. One of: very good; good; average; poor; very poor. On actual energy certificate shown as one to five star rating.
MAINHEATC_ENV_EFF	Environmental efficiency rating. One of: very good; good; average; poor; very poor. On actual energy certificate shown as one to five star rating.
LIGHTING_DESCRIPTION	Overall description of property feature. Total number of fixed lighting outlets and total number of low-energy fixed lighting outlets
LIGHTING_ENERGY_EFF	Energy efficiency rating. One of: very good; good; average; poor; very poor. On actual energy certificate shown as one to five star rating.
LIGHTING_ENV_EFF	Environmental efficiency rating. One of: very good; good; average; poor; very poor. On actual energy certificate shown as one to five star rating.
MAIN_FUEL	The type of fuel used to power the central heating e.g. Gas, Electricity
WIND_TURBINE_COUNT	Number of wind turbines; 0 if none.
HEAT_LOSS_CORRIDOOR	Flats and maisonettes only. Indicates that the flat contains a corridor through which heat is lost. Heat loss corridor, one of: no corridor; heated corridor; unheated corridor
UNHEATED_CORRIDOR_LENGTH	The total length of unheated corridor in the flat. Only populated if flat or maisonette contains unheated corridor. If unheated corridor, length of sheltered wall (m ²).

FLOOR_HEIGHT	Average height of the storey in metres.
PHOTO_SUPPLY	Percentage of photovoltaic area as a
	percentage of total roof area. 0% indicates
	that a Photovoltaic Supply is not present in
	the property.
SOLAR_WATER_HEATING_FLAG	Indicates whether the heating in the
	Property is solar powered.
MECHANICAL_VENTILATION	Identifies the type of mechanical ventilation
	the property has. This is required for the
	RdSAP calculation.
ADDRESS	Field containing the concatenation of
	address1, address2 and address3. Note
	that post code is recorded separately.
LOCAL_AUTHORITY_LABEL	The name of the local authority area in
	which the building is located. This field is
	for additional information only and should
	Not be relied upon: please refer to the Local
	Authority UNS Code.
CONSTITUENCY_LABEL	The name of the panlamentary
	constituency in which the building is
	information only and should not be relied
	upon: ploase refer to the Constituency ONS
	Code
CERTIFICATE HASH	Unique cryptographic bash of a sot of
	fields from the certificate. Can be used to
	uniquely identify a certificate in the
	downloads and the API. Note that these
	may change in the future, but will be
	consistent within a single session. Note:
	this will be removed soon in favour of just
	using I MK KEY Read more

2. Non Domestic EPCs

Table A 30 Descri	ntion of variables. No	n Domestic Energy	Performance Certificate
		n Donicouo Energy	

Name Variable	Description
LMK_KEY	Individual lodgement identifier. Not guaranteed to
	be unique. See also: Certificate Hash. We're
	making some changes soon, so that this will be
	unique Read more.
ADDRESS1	Address Line 1
ADDRESS2	Address Line 2
ADDRESS3	Address Line 3
POSTCODE	Postcode for the building address
BUILDING_REFERENCE_NUMBER	Unique identifier for the building
ASSET_RATING	Energy Performance Asset Rating. The CO ₂
	emissions from the actual building in comparison
	to a Standard Emission Rate. (kg CO ₂ /m ²)
ASSET_RATING_BAND	Energy Performance Asset Rating converted into an energy band/grade into a linear 'A+ to G' scale (where A+ is the most energy e client and G the least energy e client)
PROPERTY_TYPE	Describes the type of building that is being inspected. Based on planning use class.
INSPECTION_DATE	The date that the inspection was actually carried out by the energy assessor.
LOCAL_AUTHORITY	Office for National Statistics (ONS) code. Local authority area in which the building is located.
CONSTITUENCY	Office for National Statistics (ONS) code. Parliamentary constituency in which the building is located.
COUNTY	County in which the building is located (where applicable)
LODGEMENT_DATE	Date lodged on the Energy Performance of Buildings Register
	Type of transaction that triggered EPC. For example, one of: marketed sale; non- marketed sale; rental; not sale or rental; assessment for Green Deal; following Green Deal; FIT application; none of the above; RHI application; ECO assessment. Where the reason for the assessment is unknown by the energy assessor the transaction type will be recorded as 'none of the above'. Transaction types may be changed over time.
	built.
EXISTING_STOCK_BENCHMARK	I he Benchmark value of existing stock for this type of building
BUILDING_LEVEL	Building Complexity Level based on Energy Assessor National Occupation Standards.

MAIN_HEATING_FUEL	Main Heating fuel for the building is taken as the fuel which delivers the greatest total thermal output for space or water heating.
OTHER_FUEL_DESC	Text description of unspecified fuel type if 'Other' is selected for Main Heating Fuel.
SPECIAL_ENERGY_USES	Special energy uses discounted. This only appears on the Recommendations Report.
RENEWABLE_SOURCES	On-site renewable energy sources. This only appears on the Advisory Report.
FLOOR_AREA	The total useful floor area is the total of all enclosed spaces measured to the internal face of the external walls, i.e. the gross floor area as measured in accordance with the guidance issued from time to time by the Royal Institute of Chartered Surveyors or by a body replacing that institution. (m2)
STANDARD_EMISSIONS	Standard Emission Rate is determined by applying a fixed improvement factor to the emissions from a reference building. (kg CO ₂ /m ² /year).
TARGET_EMISSIONS	The target emission rate is the minimum energy performance requirement (required by Building Regulation) for a new non- domestic building (kg CO ₂ /m ² /year).
TYPICAL EMISSIONS	Typical Emission Rate.
BUILDING_EMISSIONS	Building Emissions Rate, Annual CO ₂ emissions
BUILDING_EMISSIONS	Building Emissions Rate. Annual CO_2 emissions from the building. Decimal (kg CO_2/m^2)
BUILDING_EMISSIONS AIRCON_PRESENT	Building Emissions Rate. Annual CO ₂ emissions from the building. Decimal (kg CO ₂ /m ²) Air Conditioning System. Does the building have an air conditioning system?
BUILDING_EMISSIONS AIRCON_PRESENT AIRCON_KW_RATING	Building Emissions Rate. Annual CO ₂ emissions from the building. Decimal (kg CO ₂ /m ²) Air Conditioning System. Does the building have an air conditioning system? Air conditioning System. Rating in kW
BUILDING_EMISSIONS AIRCON_PRESENT AIRCON_KW_RATING ESTIMATED_AIRCON_KW_RATING	 Building Emissions Rate. Annual CO₂ emissions from the building. Decimal (kg CO₂/m²) Air Conditioning System. Does the building have an air conditioning system? Air conditioning System. Rating in kW Air Conditioning System. If exact rating unknown, what is the estimated total e"ective output rating of the air conditioning system
BUILDING_EMISSIONS AIRCON_PRESENT AIRCON_KW_RATING ESTIMATED_AIRCON_KW_RATING AC_INSPECTION_COMMISSIONED	 Building Emissions Rate. Annual CO₂ emissions from the building. Decimal (kg CO₂/m²) Air Conditioning System. Does the building have an air conditioning system? Air conditioning System. Rating in kW Air Conditioning System. If exact rating unknown, what is the estimated total e"ective output rating of the air conditioning system One of:1=Yes, inspection completed; 2=Yes, inspection commissioned; 3=No inspection completed or commissioned; 4=Not relevant; 5=Don't know
BUILDING_EMISSIONS AIRCON_PRESENT AIRCON_KW_RATING ESTIMATED_AIRCON_KW_RATING AC_INSPECTION_COMMISSIONED BUILDING_ENVIRONMENT	 Building Emissions Rate. Annual CO₂ emissions from the building. Decimal (kg CO₂/m²) Air Conditioning System. Does the building have an air conditioning system? Air conditioning System. Rating in kW Air Conditioning System. If exact rating unknown, what is the estimated total e"ective output rating of the air conditioning system One of:1=Yes, inspection completed; 2=Yes, inspection commissioned; 3=No inspection completed or commissioned; 4=Not relevant; 5=Don't know Building environment which is taken as the servicing strategy that contributes the largest proportion of the building's CO₂ emissions.
BUILDING_EMISSIONS AIRCON_PRESENT AIRCON_KW_RATING ESTIMATED_AIRCON_KW_RATING AC_INSPECTION_COMMISSIONED BUILDING_ENVIRONMENT ADDRESS	 Building Emissions Rate. Annual CO₂ emissions from the building. Decimal (kg CO₂/m²) Air Conditioning System. Does the building have an air conditioning system? Air conditioning System. Rating in kW Air Conditioning System. If exact rating unknown, what is the estimated total e"ective output rating of the air conditioning system One of:1=Yes, inspection completed; 2=Yes, inspection commissioned; 3=No inspection completed or commissioned; 4=Not relevant; 5=Don't know Building environment which is taken as the servicing strategy that contributes the largest proportion of the building's CO₂ emissions. Field containing the concatenation of address1, address2 and address3. Note that post code is recorded separately.
BUILDING_EMISSIONS AIRCON_PRESENT AIRCON_KW_RATING ESTIMATED_AIRCON_KW_RATING AC_INSPECTION_COMMISSIONED BUILDING_ENVIRONMENT ADDRESS LOCAL_AUTHORITY_LABEL	 Building Emissions Rate. Annual CO₂ emissions from the building. Decimal (kg CO₂/m²) Air Conditioning System. Does the building have an air conditioning system? Air conditioning System. Rating in kW Air Conditioning System. If exact rating unknown, what is the estimated total e"ective output rating of the air conditioning system One of:1=Yes, inspection completed; 2=Yes, inspection commissioned; 3=No inspection completed or commissioned; 4=Not relevant; 5=Don't know Building environment which is taken as the servicing strategy that contributes the largest proportion of the building's CO₂ emissions. Field containing the concatenation of address1, address2 and address3. Note that post code is recorded separately. The name of the local authority area in which the building is located. This field is for additional information only and should not be relied upon: please refer to the Local Authority ONS Code.

	relied upon: please refer to the Constituency ONS Code.
CERTIFICATE_HASH	Unique, cryptographic hash of a set of fields from the certificate. Can be used to uniquely identify a certificate in the downloads and the API. Note that these may

3. Display Energy Certificates (DECs)

Name Variable	Description
LMK_KEY	Individual lodgement identifier. Not guaranteed to be unique. See also: Certificate Hash. We're making some changes soon, so that this will be
	Address Line 1
ADDRESS2	Address Line 2
ADDRESS3	Address Line 3
POSTCODE	Postcode for the building address
BUILDING_REFERENCE_NUMBER	Unique identifier for the building
CURRENT_OPERATIONAL_RATING	Current Operational Rating (OR) for this building. A numeric indicator of the amount of energy consumed during the occupation of the building over a period of 12 months. An OR is
	a measure of the annual (CO ₂) emission per unit of area of the building caused by its consumption of energy, compared to a value that would be considered typical for the particular type of building. The numbers do not represent actual units of energy consumed; they represent comparative energy efficiency.
YR1_OPERATIONAL_RATING	Operational Ratings from previous years
	(CO ₂).
YR2_OPERATIONAL_RATING	Operational Ratings from previous years (CO ₂).
OPERATIONAL_RATING_BAND	Current Operational Rating converted into an energy band/grade into a linear 'A to G' scale (where A is the most energy e client and G the least energy e client).
ELECTRIC_CO2	Total CO ₂ emissions from electricity. The energy used by the building is converted into
	an amount of carbon dioxide (CO_2) . Di"erent
	types of fuel emit di"erent amounts of CO ₂ .
	Total CO_2 emissions in tonnes per year of CO_2 .

 Table A.31. Description of variables: Display Energy Certificates

HEATING_CO2	Total CO_2 emissions from heating. The energy used by the building is converted into an
	amount of carbon dioxide (CO ₂). Di"erent types
	of fuel emit di"erent amounts of CO_2 . Total CO_2
	emissions in tonnes per year of CO ₂ .
RENEWABLES_CO2	Total CO ₂ emissions from Renewable sources.
	On-Site Renewables (OSR) include
	technologies that generate heat or electricity
	from ambient sources and have zero (or near
	zero) CO ₂ emissions. The energy they deliver
	reduces CO_2 emissions from the building
PROPERTY TYPE	Describes the type of building that is being
	inspected.
INSPECTION_DATE	The date that the inspection was actually
	carried out by the energy assessor
LOCAL_AUTHORITY	Office for National Statistics (ONS) code. Local
	authority area in which the building is located.
CONSTITUENCY	Office for National Statistics (ONS) code.
	Parliamentary constituency in which the
	Dullaing is located.
COUNTY	County in which the building is located (where
LODGEMENT DATE	applicable)
	Date lodged on the Energy Performance of Ruildings Register
	The bonchmark is the average operav
	performance for a building of this type under a
	number of standardised conditions for
	temperature, occupancy and proportion of non-
	electrical energy used. Under certain
	circumstances, these benchmarks may be
	adjusted according to location, occupancy and
	the ratio of non-electrical energy used.
MAIN_HEATING_FUEL	This indicates the main type of fuel used to
	heat the building.
OTHER_FUEL	Text description of unspecified fuel type if
	'Other' is selected for Main-Heating-Fuel.
SPECIAL_ENERGY_USES	Separable energy uses. The aim of the
	Operational Rating is to compare the annual
	energy consumption of the building with that of
	a building typical of its type. In some cases the
	building may include activities that consume
	that building type. It may be reasonable to
	subtract these senarable energy uses in
	certain circumstances. In order to be able to
	isolate and remove the annual separable
	energy consumption from the total any
	separable energy uses must be separately
	metered. This only appears on the Advisory
	Report.

RENEWABLE_SOURCES	On-site renewable energy sources. This only appears on the Advisory Report
TOTAL FLOOR AREA	Total Useful Floor Area (m ²)
ANNUAL THERMAL FUEL USAGE	Annual Energy Use (kWh/m²/year) for heating
TYPICAL_THERMAL_FUEL_USAGE	Typical Energy Use (kWh/m²/year) for heating
ANNUAL_ELECTRICAL_FUEL_USAGE	Annual Energy Use (kWh/m²/year) for electricity
TYPICAL_ELECTRICAL_FUEL_USAGE	Typical Energy Use (kWh/m²/year) for electricity
RENEWABLES_FUEL_THERMAL	Percentage of energy obtained from on-site renewable sources for heating (if any)
RENEWABLES_ELECTRICAL	Percentage of energy obtained from on-site renewable sources for electricity (if any).
YR1_ELECTRICITY_CO2	CO2 emissions from electricity in previous
	reporting year (if any). Total CO ₂ emissions in
	tonnes per year of CO ₂ .
YR2_ELECTRICITY_CO2	CO2 emissions from electricity in previous
	reporting year (if any). Total CO_2 emissions in
	tonnes per year of CO ₂ .
YR1_HEATING_CO2	CO2 emissions from heating in previous
	reporting year (if any). Total CO ₂ emissions in
	tonnes per year of CO ₂ .
YR2_HEATING_CO2	CO2 emissions from heating in previous
	reporting year (if any). Total CO_2 emissions in
	tonnes per year of CO_2 .
YR1_RENEWABLES_CO2	CO_2 emissions from renewable sources in
	previous reporting year (if any). Total CO_2
	emissions in tonnes per year of CO ₂ .
YR2_RENEWABLES_CO2	CO_2 emissions from renewable sources in
	previous reporting year (if any). Total CO_2
	emissions in tonnes per year of CO ₂ .
AIRCON_PRESENT	Air Conditioning System. Does the building have an air conditioning system?
AIRCON_KW_RATING	Air conditioning System. Rating in kW
ESTIMATED_AIRCON_KW_RATING	Air Conditioning System. If exact rating unknown, what is the estimated total e"ective output rating of the air conditioning system in kW.
AC_INSPECTION_COMMISSIONED	One of:1=Yes, inspection completed; 2=Yes, inspection commissioned; 3=No inspection completed or commissioned; 4=Not relevant; 5=Don't know

BUILDING_ENVIRONMENT	Building environment which is taken as the servicing strategy that contributes the largest
	proportion of the building's CO2 emissions.
BUILDING_CATEGORY	Building category codes (described below). This data field may contain multiple benchmark categories. Where a building has a mix of uses that would place parts of the building in a di"erent benchmark category, it is possible to construct a composite benchmark, e.g. a school with a swimming pool.
ADDRESS	Field containing the concatenation of address1, address2 and address3. Note that post code is recorded separately.
LOCAL_AUTHORITY_LABEL	The name of the local authority area in which the building is located. This field is for additional information only and should not be relied upon: please refer to the Local Authority ONS Code.
CONSTITUENCY_LABEL	The name of the parliamentary constituency in which the building is located. This field is for additional information only and should not be relied upon: please refer to the Constituency ONS Code.
CERTIFICATE_HASH	Unique, cryptographic hash of a set of fields from the certificate. Can be used to uniquely identify a certificate in the downloads and the API. Note that these may change in the future, but will be consistent within a single session. Note: this will be removed soon in favour of just using LMK KEY. Read more.

Relevant Documents

The following documents are included in the folder attached to this document.

Table A.52. Relevant Documents. Lifergy Ferrormance Certificate (LFC)				
Document Name	Name file			
Energy Performance Certificate data	D10. Energy Performance Certificate data			
Energy Performance of Buildings Documentation	Energy Performance of Buildings Documentation.pdf			

Table / tole tall becamenter Energy i enermanee eertineate (Er e)

Quality²⁴

Primary responsibility for the robustness of the data in relation to individual buildings lies with the energy assessor who carried out the energy assessment. Additional validation checks are undertaken as the data is lodged on the registers and improved quality assurance procedures for energy assessor accreditation schemes have been introduced.

²⁴ This section is extracted from the document "Energy Performance of Buildings Documentation"

The data cannot be altered once it has been lodged on the registers, although it can be marked at 'not for issue' or 'cancelled' for a range of reasons. The data has, therefore, been published without any changes or alterations, leading to recognise that there may be anomalies in the published data.

Data Profile 11: National Register for Social Housing

Name of the Data: National Register for Social Housing²⁵

Description of the Data

Development of the National Register for Social Housing (NROSH) was started by the Department for Communities and Local Government (DCLG) in 2004. NROSH aimed to be a database of all social housing properties in England, with a range of details captured on each property. NROSH was transferred to the Tenant Services Authority, the social housing regulator, in April 2010 and was discontinued in May 2011. Ownership of the latest NROSH dataset passed from the TSA to the Homes and Communities Agency (HCA) when responsibility for social housing regulation passed to the Regulation Committee of the HCA in April 2012.

Data Profile

Field	Description
Dataset ID	11
Dataset Name	National Register of Social Housing (NROSH)
Temporal Extent	2014
Supplier/data owner	National Register of Social Housing (NROSH)
Analytical Units	Points, (Address)
Data Format	CSV
Geographical Extent	Great Britain
Access	Open Government Licence
Cost	Free
Observations	Circa 3.3 million records
Web page	https://data.gov.uk/dataset/cf3aac3e-f3f3-4564-9df4- 099346246251/national-register-of-social-housing-nrosh

Table A.33. Data Profile: National Register for Social Housing

Structure of the Data

Two datasets are available. One is the latest NROSH database held by the HCA as at May 2011. This release contains a large subset of the full NROSH dataset (48 from 201 fields in total; for 4,826,417 unique property records). The data in this release does not include those fields where data could enable specific identification of vulnerable people or other sensitive personal data. It also excludes fields where a minimum completion threshold is not met (generally fields where less than 25% of records have data).

Relevant Documents

²⁵ This document is a summary of relevant elements contained in the link https://data.gov.uk/dataset/cf3aac3e-f3f3-4564-9df4-099346246251/national-register-of-socialhousing-nrosh

The following documents are included in the folder attached to this document.

	· · · · · · · · · · · · · · · · · · ·
Document Name	Name file
Data field description for DCLG NROSH	dclg_data_field_description.doc
Data Field Summary: HCA release of NROSH data	HCA_Data_Field_Summary_03112014_REVISED.doc
National Register of Social Housing	National Register of Social Housing (NROSH) - data.gov.uk.pdf

Quality

In addition to being out of date, the records submitted by social landlords to NROSH are of varying quantity and quality with many incomplete, inaccurate or missing records. The database may also contain a number of duplicate entries. Overall, there are still issues of quality, incomplete data, and potential duplication of records in the data that HCA is not able to resolve.

Data Profile 12: Historic Ordnance Survey

Name of the Data: Data Historic Ordnance Survey²⁶

Description of the Data

Historic Digimap delivers Landmark historic Ordnance Survey maps of Great Britain to UK Higher and Further Education. The historic maps can be viewed on-line, printed and downloaded for use in Geographical Information Systems (GIS) and other software.

They contains geo-referenced scans of original Ordnance Survey maps of Great Britain, at 1:1250 scale (urban), 1:2500 scale (suburban) and 1:10560 scale (rural).

Data Profile

Table A.35. Data Profile: Historic Ordnance Survey

Field	Description
Dataset ID	13
Dataset Name	Historic Ordnance Survey data
Temporal Extent	1946 – 1992
Supplier/data owner	Ordnance Survey
Analytical Units	Points, Lines, Polygons
Data Format	TIFF
Geographical Extent	Great Britain
Access	Open access for Educational Purposes
Cost	Free
Observations	-
Web page	https://digimap.edina.ac.uk/historic

Structure of the Data

The table below provides a summary of the maps available through the Historic Digimap interface. It includes the following fields:

Landmark Epoch

The name assigned by Landmark Information Group to each "layer" they have created, grouped in most cases by map series, edition and/or scale. Landmark refer to each group of maps as an "Epoch".

- Map Series / Category Name
 A descriptive name for each of the Landmark "Epochs"
- Scale/s The scale or scales of maps grouped in each Landmark "Epoch"

²⁶ This document is a summary of relevant elements contained in the link https://digimap.edina.ac.uk/historic

• Publishing Date Range

The earliest and latest publishing date within each "Epoch"

• Notes and Coverage

Additional information and links to Coverage Maps showing the geographic availability of maps in each "Epoch"

Landmark Epoch	Map Series/ Category Name	Scale/s	Publishing Date Range	Notes and Coverage
Epoch 1	County Series Survey	1:2,500	1854-1901	First Edition County Series maps. These are available as original County Series map sheets (rectangular) and as National Grid square tiles (each tile sheet may contain parts of up to 6 original County Series map sheets).
		1:10,560	1846-1899	1:2,500 coverage: high, but coverage poor in North East England, Dumfries and Galaway and North West Scotland (no high land areas mapped). View coverage map
		("6 inch to the mile")		1:10,560 coverage: very high, currently no maps for Perthshire and parts of Isles of Lewis/Harris. View coverage map
Epoch 2	County Series 1st Revision	1:2,500	1893-1915	First revision of County Series maps. These are available as original County Series map sheets (rectangular) and as National Grid square tiles (each tile sheet may contain parts of up to 6 original County Series map sheets).
		1:10,560	1888-1914	1:2,500 coverage: high, but coverage poor in North Yorkshire, parts of South East England and North West Scotland (no high land areas mapped). View coverage map
		("6 inch to the mile")		1:10,560 coverage: very high, currently no maps for Shetland Isles. View coverage map
Epoch 3	County Series 2nd Revision	1:2,500	1906-1939	Second revision of County Series maps. These are available as original County Series map sheets (rectangular) and as National Grid square tiles (each tile sheet may contain parts of up to 6 original County Series map sheets).
		1:10,560	1903-1949	1:2,500 coverage: medium. View coverage map
		("6 inch to the mile")		1:10,560 coverage: high, but poor coverage in Cornwall, Wales, North West Scotland, Scottish/English border, Lincolnshire. View coverage map

Table A.36. Summary of Maps Available – Historic Ordnance Survey

Epoch 4	County Series 3rd Revision	1:2,500	1924-1949	Third revision of County Series maps. These are available as original County Series map sheets (rectangular) and as National Grid square tiles (each tile sheet may contain parts of up to 6 original
		1:10,560	1922-1969	1:2,500 coverage: low, concentrated in urban areas. View coverage map
		("6 inch to the mile")		1:10,560 coverage: medium. View coverage map
Epoch A5	National Grid overhaul	1:1,250	1943-1993	Represents "Edition A" maps - i.e. the first post World War Two National Grid large scale maps.
	and re- survey (Edition A)	1:2,500	1943-1995	1:1,250 in urban areas only.
				Some areas mapped at 1:1,250 in later editions are mapped at 1:2,500 in Edition A.
				1:1,250 coverage: v. high - urban only View coverage map
				1:2,500 coverage: v. high - not highland areas View coverage map
Epoch B6	National Grid resurvey	1:1,250	1944-1993	Represents "Edition B" maps. Some areas mapped at 1:1,250 in later editions are mapped at 1:2,500 in Edition B.
	(Edition B)	1:2,500	1949-1992	1:1,250 coverage: high, but urban areas only. View coverage map
				1:2,500 coverage: medium - areas of significant change; not highland areas View coverage map
Epoch C7	National Grid resurvey	1:1,250	1946-1993	Represents "Edition C" maps. 1:1,250 in urban areas only.
	(Edition C)	1:2,500	1954-1993	Some areas mapped at 1:1,250 in Edition C are only mapped at 1:2,500 in Editions A and B.
				1:1,250 coverage: medium, urban areas only. View coverage map
				1:2,500 coverage: low. View coverage map
Epoch D8	National Grid resurvey	1:1,250	1951-1992	Represents "Edition D" maps. 1:1,250 in urban areas only.
	(Edition D)	1:2,500	1960-1992	Some areas mapped at 1:1,250 in Edition D are only mapped at 1:2,500 in Editions A and B.
				1:1,250 coverage: low, urban areas only. View coverage map
				1:2,500 coverage: very low. View coverage map

Epoch E9	National Grid resurvey	1:1,250	1953-1992	Represents "Edition E" maps. 1:1,250 in urban areas only.
	(Edition E)	1:2,500	1967-1992	Some areas mapped at 1:1,250 in Edition E are only mapped at 1:2,500 in Editions A and B.
				1:1,250 coverage: very low, urban areas only. View coverage map
				1:2,500 coverage: very low. View coverage map
Epoch i5	National Grid Imperial	1:10,560	1948-1977	Includes both "Provisional" and "Regular" First Editions.
	"6 inches to the mile" First Editions			These maps use the new metric National Grid referencing system but units of measurement are imperial.
				Coverage: high - near to full View coverage map
Epoch i6	National Grid Imperial	1:10,560	1949-1981	These maps use the new metric National Grid referencing system
	"6 inches to the mile" 1st revision			but units of measurement are imperial.
				Coverage: medium - revisions only View coverage map
Epoch i7	National Grid Imperial	1:10,560	1959-1982	These maps use the new metric National Grid referencing system
	"6 inches to the mile" 2nd revision			but units of measurement are imperial.
				Coverage: low - revisions only View coverage map
Epoch i8	National Grid Imperial	1:10,560	1967-1976	These maps use the new metric National Grid referencing system
	"6 inches to the mile" 3rd revision			but units of measurement are imperial.
				Coverage: very low - revisions only View coverage map
Epoch m5	National Grid 1:10,000 series	1:10,000	1969-1996	First fully metric 1:10,000 scale edition, replacing imperial "6 inch to the mile" National Grid maps.
	Fully metric			Coverage: Medium to high - some gaps View coverage map
	First Edition.			
-------------	---	---------------------------------------	-----------	---
Epoch m6	National Grid 1:10,000 series	1:10,000	1973-1996	Revision of 1:10,000 scale metric edition.
	metric First Revision			Coverage: low View coverage map
Epoch m7	National Grid 1:10,000 metric and 1:10,560 imperial latest editions	1:10,560 and 1:10,000 scale.	1958-1996	This map "layer" contains the latest map available for each mapped location. The latest edition for any given location could be either a 1:10,560 scale imperial National Grid map (pre 1970) or a 1:10,000 scale metric National Grid map, so this "layer" is necessarily a mix of both imperial 1:10,560 and metric 1:10,000 maps. Coverage: Very high - there is a small number of "missing" maps, which show up as small white dots on the Coverage map

Relevant Documents

The following documents are included in the folder attached to this document.

Document Name	Name file
About Historic Digimap	About Historic Digimap.pdf
Summary of Maps, Scales and Dates	Summary of Maps, Scales and Dates.pdf
Map Series Timelines	Map Series Timelines.pdf

Table A.37. Relevant Documents: Historic Ordnance Survey

Quality

Historic Digimap products include high quality maps that allow researchers to consistently analyse historical changes of real-world objects. Besides, several steps are performed by Digimap to continually update the data and assess its quality.

Data Profile 13: Met Office data

Name of the Data: Met Office data²⁷

Description of the Data

There are five main UK climate datasets

- 1. **UK and regional series:** Download time-series of monthly, seasonal and annual values for the UK, countries and regions in rank or date order.
- 2. **HadUK-Grid:** A new high-resolution gridded land surface observational climate data set for the UK.
- 3. **UKCP09 gridded observation datasets:** Gridded climate data for the UK interpolated from station observations.
- 4. **Historic station data**²⁸: Long running monthly records for a selection of stations.
- 5. **Met Office DataPoint:** DataPoint is an unsupported service to access freely available Met Office data feeds in a format that is suitable for application developers. It is aimed at anyone looking to re-use Met Office data within their own innovative applications, for example professionals, the scientific community, student and amateur developers. Whilst unsupported, there is a Google user group where questions and answers can be shared with other DataPoint users; details can be found on the Getting started page.

Data Profile

Field	Description
Dataset ID	14
Dataset Name	Met Office data
Temporal Extent	1961 – 2016
Supplier/data owner	Met Office
Analytical Units	Points, Grids
Data Format	Geodatabases (GDB), Shapefile (SHP), CVS
Geographical Extent	Great Britain
Access	Open access
Cost	Free
Observations	-
Web page	https://www.metoffice.gov.uk/climate/uk/data

Table A.38. Data Profile: Met Office data

Structure of the Data

²⁷ This document is a summary of relevant elements contained in the link https://www.metoffice.gov.uk/climate/uk/data

²⁸ <u>https://www.metoffice.gov.uk/public/weather/climate-historic/#?tab=climateHistoric</u>

1. UK and regional series

The following variables and categories are available in this data:

- Variables: Max Temp, Min Temp, Mean Temp, Sunshine, Rainfall, Rain days ≥1.0mm, Days of Air frost.
- Region categories: UK, England, Wales, Scotland, Northern Ireland, England & Wales, England N, England S, Scotland N, Scotland E, Scotland W, England E & NE, England NW/Wales N, Midlands, East Anglia, England SW/Wales S, England SE/Central S.
 - 2. HadUK-Grid:

HadUK-Grid is a new collection of gridded climate variables derived from the network of UK land surface observations. The data have been interpolated from meteorological station data onto a uniform grid to provide complete and consistent coverage across the UK. The data sets cover the UK up to 1km x 1km resolution and a range of other resolutions to allow for comparison to data from climate projections and across a country, administrative regions and river basins. The dataset spans the period from 1862 to present, but the start time is dependent on climate variable and temporal resolution. The grids are produced for daily, monthly, seasonal and annual timescales, as well as long-term averages for a set of climatological reference periods. Variables include air temperature (maximum, minimum and mean), precipitation, sunshine, mean sea level pressure, wind speed, relative humidity, vapour pressure, days of snow lying, and days of ground frost. This data set will supersede the UKCP09 gridded observation. The primary purpose of these data is to facilitate monitoring of UK climate and research into climate change, impacts and adaptation. The datasets have been created by the Met Office with financial support from the Department for Business, Energy and Industrial Strategy (BEIS) and Department for Environment, Food and Rural Affairs (DEFRA) in order to support the Public Weather Service Customer Group (PWSCG), the Hadley Centre Climate Programme, and the UK Climate Projections (UKCP18) project. The data recovery activity has also been funded by the Natural Environment Research Council (NERC grant ref: NE/L01016X/1) project "Analysis of historic drought and water scarcity in the UK".

There are two main group of data: UKCP18 marine projections and UKCP18 climate models.

Dataset	Description	Emissions scenarios	Time Period	Domain
Time mean sea level at 12 km	Projections of future changes in sea water level	RCP2.6 RCP4.5 RCP8.5	2007-2100	UK coastline
Storm surge trend at 12km	Projections of storm surge trend around the UK coastline	RCP8.5	2007-2100	UK coastline
Storm surge simulations	Time series of gridded historical and future simulations of sea water level	RCP8.5	1970-2099	UK

Table A.39. Summary of UKCP18 marine projections data

Short event case studies	Time series of gridded historical and future simulations of sea water level for three events (6 Dec 2013, 3 Feb 2014, 11 Jan 2015)	N/A	N/A	UK
Time mean sea level at 12km	Exploratory projections of future changes in sea water level around the UK coastline	RCP2.6 RCP4.5 RCP8.5	2007-2300	UK

Source: 'UKCP18 Guidance: Data availability, access and formats' Met Office

Table A.40. Summary of UKC	P18 climate models	and scenarios for	projections over
land			

Dataset	Description	Emissions scenarios	Time Period	Geographical Domain
Probabilistic projections at 25km	Probabilistic projections over land	RCP2.6 RCP4.5 RCP6.0 RCP8.5 SRESA1B	1961-2100	UK
Global projections at 60km	Global climate model projections including 15 from the Met Office Hadley Centre (PPE-15) and 13 models from other international climate modelling centres (CMIP5-13)	RCP8.5	1900-2100	Global UK
Regional projections at 12km	A set of 12 regional high resolution climate projections driven by the global climate model projections	RCP8.5	1981-2080	Europe UK
Derived projections at 60km	Projections for RCP2.6 and worlds with 2°C and 4°C warming produced using statistical methods and based on the global climate model projections	RCP2.6 2°C world 4°C world	1900-2100	UK

Source: 'UKCP18 Guidance: Data availability, access and formats' Met Office

Variable at the surface (short name in CEDA catalogue)	Units	Marine	Probabilistic projections	†Global projections	Derived projections	Regional projections
Cloud cover (clt)	%		\checkmark	\checkmark		\checkmark
Precipitation (pr)	mm/day		\checkmark	\checkmark	√+	\checkmark
Radiation, total downward short wave flux (rsds)	Wm-2		\checkmark			
Radiation, net long wave (rls)	Wm-2		\checkmark	\checkmark		\checkmark
Radiation, net short wave (rss)	Wm-2		\checkmark	\checkmark		\checkmark
Relative humidity (hurs)	%			\checkmark	\checkmark	\checkmark
Sea level pressure (psl)	hPa		\checkmark	\checkmark		\checkmark
Sea water level	m	\checkmark				
Specific humidity (huss)			\checkmark	\checkmark		\checkmark
Temperature, maximum (tasmax)	°C		\checkmark	\checkmark		\checkmark
Temperature, Mean (tas)	°C		\checkmark	\checkmark	√+	\checkmark
Temperature, minimum (tasmin)	°C		\checkmark	\checkmark		\checkmark
Wind speed (sfcWind)	m/s			\checkmark	\checkmark	\checkmark
Wind speed eastwards (uas)	m/s			\checkmark	\checkmark	\checkmark
Wind speed northwards (vas)	m/s			\checkmark	\checkmark	\checkmark
Time steps		Dataset- dependent	Monthly Seasonal Annual 20/30-year means	Daily Monthly Seasonal* Annual* 20/30-year means*	Daily+ Monthly	Daily Monthly Seasonal* Annual* 20/30-year means*

Table A.41. Available UKCP18 variables for the marine and projections over land

Source: 'UKCP18 Guidance: Data availability, access and formats' Met Office; *these are available over the UK only +only daily precipitation and temperature are available for the derived projections †Not all variables are available for CMIP-13.

3. UKCP09 gridded observation datasets

Gridded data sets based on surface observations have been generated for a range of climatic variables and are available under open licence. The primary purpose of this data resource is to encourage and facilitate research into climate change impacts and adaptation. The data sets have been created with financial support from the Department for Environment, Food and Rural Affairs (Defra) and are promoted within the UK Climate Projections (UKCP09). The UKCP09 report "The climate of the UK and recent trends" uses these gridded data sets to describe UK climatologies and recent regional trends. These datasets are being revised as part of the UKCP18 Project and support an annual State of the UK Climate publication.

The data sets cover the UK at 5 x 5 km resolution and span the period 1910–2016. They are available for daily, monthly and annual timescales, as well as long-term averages for the 1961–1990 climate baseline period. Baseline averages are also available.

The following datasets are available:

- Daily temperature and precipitation at 5km resolution
- Monthly climate variables at 5km resolution
- Precipitation and temperature indices at 5km resolution
- Long term averages at 5km resolution
- Long term averages at 25km resolution
- Long term averages for administrative regions and river basins
- 4. Historic station data

The data consists of:

- Mean daily maximum temperature
- Mean daily minimum temperature
- Days of air frost
- Total rainfall
- Total sunshine duration

Monthly data are available for a selection of long-running historic stations. The series typically range from 50 to more than 100 years in length.

Relevant Documents

1. UK and regional series

The following documents are included in the folder attached to this document.

Table A.42. Data Profile: Met Office data

Document Name	Name file
UK and regional series	UK and regional series_Met Office.pdf

2. HadUK-Grid

The following documents are included in the folder attached to this document.

Document Name	Name file
Data formats	Data formats - Met Office.pdf
Data availability access and	data-availability-access-and-formatspdf
formats	
Datasets	Datasets - Met Office.pdf
HadUK Grid	HadUK-Grid - Met Office.pdf
HadUK Grid Frequently	HadUK-Grid Frequently Asked Questions - Met Office.pdf
Asked Questions	
HadUK Grid Methods	HadUK-Grid Methods - Met Office.pdf
HadUK Grid Overview	HadUK-Grid Overview - Met Office.pdf
The Generation of Daily	cm24_generation_of_daily_gridded_datasets.pdf
Gridded Datasets of	
Temperature and Rainfall for	
the UK	

3. UKCP09 gridded observation datasets

The following documents are included in the folder attached to this document.

Document Name	Name file
Data formats	Data formats - Met Office.pdf
Datasets	Datasets - Met Office.pdf
Methods	Methods - Met Office.pdf
UKCP09 Frequently Asked	UKCP09 Frequently Asked Questions - Met Office.pdf
Questions	
Questions - Met Office.pdf	UKCP09 gridded observation datasets - Met Office.pdf
UKCP09	

Table A.43. Relevant Documents

Quality

The station observations used to produce meteorological data have undergone a quality control process, which corrects or removes erroneous data. See document "The Generation of Daily Gridded Datasets of Temperature and Rainfall for the UK" for more details.

Data Profile 14: Air Pollution

Name of the Data: Air Pollution²⁹

Description of the Data

There are three main UK Air Pollution datasets:

- Automatic Urban and Rural Network (AURN): The AURN is the UK's largest automatic monitoring network and is the main network used for compliance reporting against the Ambient Air Quality Directives. It includes automatic air quality monitoring stations measuring oxides of nitrogen (NO2), sulphur dioxide (SO2), ozone (O2), carbon monoxide (CO) and particles (PM10, PM2.5). These sites provide high resolution hourly information which is communicated rapidly to the public, using a wide range of electronic, media and web platforms.
- 2. **Modelled background pollution data:** Background pollution maps at 1x1 km resolution are modelled each year under Defra's Modelling of Ambient Air Quality (MAAQ) contract.
- 3. Background Mapping data for local authorities³⁰: These concentrations maps are provided to assist local authorities in support of Review and Assessment of local air quality. To assist local authorities undertaking their air quality Review and Assessments, 2017-based background maps for years 2017 to 2030 for NO_x, NO₂, PM₁₀ and PM_{2.5} are available below. Users should note that 2017-based maps now replace the 2015-based maps. The 2017 reference year background maps are based on monitoring and meteorological data for 2017.

Field	Description
Dataset ID	15
Dataset Name	Pollution data
Temporal Extent	1999 – 2016
Supplier/data owner	DEFRA
Analytical Units	Points, Grids
Data Format	Geodatabases (GDB), Shapefile (SHP), CVS
Geographical Extent	Great Britain
Access	Open access
Cost	Free
Observations	-

Table A.44. Data Profile: Air Pollution

²⁹ This document is a summary of relevant elements contained in the link https://uk-air.defra.gov.uk/data/

³⁰ <u>https://laqm.defra.gov.uk/review-and-assessment/tools/background-maps.html</u>

and https://uk-air.defra.gov.uk/data/laqm-background-home

Structure of the Data

1. Automatic Urban and Rural Network (AURN):

Data availability: from 22/02/1973 to 09/12/2018

Variables:

- Ambient Temperature
- Barometric pressure
- Carbon monoxide
- Daily measured PM (uncorrected)
- Daily measured PM (uncorrected)
- Modelled Temperature
- Modelled Wind Direction
- Modelled Wind Speed
- Nitric oxide
- Nitrogen dioxide
- Nitrogen oxides as nitrogen dioxide
- Non-volatile PM (Hourly measured)
- Non-volatile PM (Hourly measured)
- Ozone
- PM10 Ambient pressure measured
- PM10 Ambient Temperature
- PM2.5 Ambient Preasure
- PM2.5 Ambient Temperature
- PM particulate matter (Daily measured)
- PM particulate matter (Hourly measured)
- PM particulate matter (Hourly measured)
- PM particulate matter (Daily measured)
- PM particulate matter (Hourly measured)
- Rainfall
- Relative Humidity
- Sulphur dioxide
- Total Particulates
- Volatile PM (Hourly measured)
- Volatile PM (Hourly measured)
- Wind Direction
- Wind Speed
- 2. Modelled background pollution data

The modelled data include annual mean of the following pollutants: PM10, PM2.5, NOX and NO2, CO, SO2, Ozone, Benzene.

3. Background Mapping data for local authorities

The background mapping data for local authorities include data of the following pollutants and years.

Table A.45. Available background mapping data for local authorities

Air Pollutants	Years
NOx, NO2, PM10 and PM2.5.	2004, 2006,2008, 2011, 2013, 2015, 2017
SO2, CO, benzene and 1,3-butadiene	2001

Relevant Documents

1. AURN

The following documents are included in the folder attached to this document.

Table A.46. Relevant Documents: Air Pollution - AURN

Document Name	Name file
Automatic Urban and	Automatic Urban and Rural Network (AURN) - Defra, UK.pdf
Rural Network	
Data Validation and	Data_Validation_and_Ratification_Process_Apr_2017.pdf
Ratification Process	
Site environment types	Site environment types - Defra, UK.pdf

2. Background Map Local Authorities

The following documents are included in the folder attached to this document.

Table A.47. Relevant Documents: Air Pollution – Background Map Local Authorities

Document Name	Name file
Based background maps	2017-based-background-maps-user-guide-v1.0.pdf
user guide	
Background Mapping data	Background Mapping data for local authorities - Defra,
for local authorities	UK.pdf

3. Modelled

The following documents are included in the folder attached to this document.

Table A.48. Relevant Documents: Air Pollution - Modelled

Document Name	Name file
Modelled background pollution data	Modelled background pollution data - Defra, UK.pdf
Technical report on UK supplementary assessment	Ricardo Energy and Environment 2014 - Technical report on UK supplementary assessment.pdf

Quality

Defra conduct Quality Assurance and Quality Control (QA/QC) procedures periodically. Provisional data and statistics are clearly marked with a flag in the database to indicate their status. Data Validation is carried out on an ongoing basis and is nominally a process to "cleanup" the initial provisional data. Any corrections to the data made during the validation process are automatically uploaded (still as provisional at this stage) to UK-AIR for end users to access. The process includes:

- Further manual review of the data to exclude any data from instrument malfunctions or faulty calibrations.
- Incorporation of any data, which were initially missing due to communications failure with a monitoring station.
- Updates to data scaling following application of the most recent calibration factors.

ground Mapping data for local authorities

Data Profile 15: Crime Data

Name of the Data: Crime Data³¹

Description of the Data

This data includes individual crime and anti-social behaviour (ASB) incidents, including street-level location information and subsequent police and court outcomes associated with the crime.

Every month each police force generates a Crime and ASB file and a Police Outcomes file in a set format. The forces upload these to a private server managed by the Home Office in the Government network, where the files undergo quality assurance. Copies of the data from police forces is then sent to the Ministry of Justice (MoJ), where they try to match the crimes with any court results contained in their own records. The MoJ send any matching court results back to the Home Office, where they are integrated with the existing data. All data is then anonymised before being published to this site.

Data Profile

Field	Description
Dataset ID	16
Dataset Name	Crime data
Temporal Extent	2010 – 2019
Supplier/data owner	Police UK
Analytical Units	Points (Streets)
Data Format	Geodatabases (GDB), Shapefile (SHP), CVS
Geographical Extent	Great Britain
Access	Open access
Cost	Free
Observations	Circa 40 million crime event
Web page	https://data.police.uk/data/

Table A.49. Data Profile: Crime Data

Structure of the Data

The following Table describes the variables included in the datasets.

Table A.Su. Description of Variables: Crime Data			
Variables	Description		
Crime ID	Crime ID		
Month	Month		

Table A.50. Description of Variables: Crime Data

³¹ This document is a summary of relevant elements contained in the link https://data.police.uk/data/

Reported by	The force that provided the data about the crime.
Falls within	At present, also the force that provided the data about the crime. This is currently being looked into and is likely to change in the near future.
Longitude	The anonymised coordinates of the crime
Latitude	The anonymised coordinates of the crime
Location	
LSOA code	References to the Lower Layer Super Output Area that the anonymised point falls into, according to the LSOA boundaries (http://data.gov.uk/dataset/lower_layer_super_output_area_lsoa_boundarie
LSOA name	s) rovided by the Office for National Statistics.
Outcome type	A reference to whichever of the outcomes associated with the crime occurred most recently. For
	example, this crime (https://www.police.uk/leicestershire/NC04/crime/2017- 05/allcrime/
	883235/590d68b69228a9ff95b675bb4af591b38de561aa03129dc09a03ef34 f537588c/)'s
	'Last outcome category' would be 'Formal action is not in the public interest'.
Crime Type	One of the crime types listed in the Police.UK FAQ
	faqs/#what-do-the-crime-categories-mean).

The following Table describes the main Crime categories.

Table	A.51.	Data	Profile:	Crime	Data
IUNIC	/	Dutu		0	Dutu

Crime categories	Description
Anti-social behaviour	Includes personal, environmental and nuisance anti-social behaviour.
Bicycle theft	Includes the taking without consent or theft of a pedal cycle.
Burglary	Includes offences where a person enters a house or other building with the intention of stealing.
Criminal damage and arson	Includes damage to buildings and vehicles and deliberate damage by fire.
Drugs	Includes offences related to possession, supply and production.
Other crime	Includes forgery, perjury and other miscellaneous crime.
Other theft	Includes theft by an employee, blackmail and making off without payment.
Possession of weapons	Includes possession of a weapon, such as a firearm or knife.

Public order	Includes offences which cause fear, alarm or distress.
Robbery	Includes offences where a person uses force or threat of force to steal.
Shoplifting	Includes theft from shops or stalls.
Theft from the person	Includes crimes that involve theft directly from the victim (including handbag, wallet, cash, mobile phones) but without the use or threat of physical force.
Vehicle crime	Includes theft from or of a vehicle or interference with a vehicle.
Violence and sexual offences	Includes offences against the person such as common assaults, Grievous Bodily Harm and sexual offences.

Location Anonymisation: The latitude and longitude locations of Crime and ASB incidents published on this site always represent the approximate location of a crime - not the exact place that it happened. See FAQs - Police.uk.pdf for more details.

Relevant Documents

The following documents are included in the folder attached to this document.

Table A.52. Relevant Documents. Chine Data			
Document Name	Name file		
About data police	About _ data.police.uk.pdf		
FAQs	FAQs - Police.uk.pdf		

Table A 52 Polovant Documents: Crime Data

Quality³²

The data that police forces provide to the Home Office and Ministry of Justice goes through a rigorous quality control process, involving format validation, automated testing, and manual verification and approval.

Format Validation

At the point of upload the following checks are carried out on the Crime & ASB and Police Outcomes files. The upload is rejected if any of the checks fail. All fields are in the correct format (includes length validation and checks against regular expressions). All dates are valid and fall within the correct time periods. No required fields are blank. All locations are within England, Wales and Northern Ireland.

Manual Verification

After the upload has been completed, a summary of the data and potential issues are provided to the force. The force must review the data and approve it as suitable for publication.

This manual check includes assessing:

- The percentage changes since last month, overall and for each individual category of crime.
- The number of crimes that have been uploaded with no location.

³² This section is extracted from the document "About data.police.uk"

- The number of crimes that have not been successfully anonymised to a map point.
- The number of crimes that have been anonymised to a map point outside of the originating police force boundary.
- All free-text context content for potential privacy breaches.
- The overall number of outcomes compared to the number of crimes.
- The percentage split across the outcome categories, and comparison to previous months.
- The number of outcomes that were uploaded that couldn't be joined up to an original crime record.

Automated Testing

After the automated tests have been completed, a final round of automated testing is carried out. If any anomalies are spotted, these are flagged to the team at the Home Office to investigate.

The automated tests run include such checks as:

- All CSVs have more than one line.
- All CSVs are unique.
- With four specific exceptions, no one region's crime count doubles or halves from one month to the next.
- All force boundaries contain points falling within one degree of the other points in that boundary.
- All force boundary points fall between within the rectangle with edges at 8.2°W, 1.8°E, 49.8°N and 55.9°N.

Data Profile 16: Technology Data

Name of the Data: Technology Data³³

Description of the Data

Ofcom collected and analysed data from fixed telecoms operators. This data contains broadband coverage availability and performance data from operators. The availability data also includes coverage information provided by alternative network providers. Further information can be found in the Connected Nations Methodology (see Connected Nations 2018 Methodology.pdf).

Data Profile

Field	Description
Dataset ID	17
Dataset Name	Technology data
Temporal Extent	2013 – 2019
Supplier/data owner	OFCOM
Analytical Units	Polygons (Local-and-unitary-authority; Output-area; Parliamentary- constituency;Postcode)
Data Format	CSV
Geographical Extent	Great Britain
Access	Open access
Cost	Free
Observations	1.6 million postocdes per year
Web page	https://www.ofcom.org.uk/research-and-data/multi-sector- research/infrastructure-research/connected-nations-2018/data- downloads

Table A.53. Data Profile: Technology Data

Structure of the Data

There are four type of data aggregated at different geographical levels.

- Local-and-unitary-authority
- Output-area
- Parliamentary-constituency
- Postcode

Each data is then divided by coverage and performance:

Coverage variables

³³ This document is a summary of relevant elements contained in the link

https://www.ofcom.org.uk/research-and-data/multi-sector-research/infrastructure-research/connected-nations-2018/data-downloads

Current headers	Note
postcode	Postcode with spaces removed, such as SE19HA
postcode space	Postcode with single space, such as SE1 9HA
postcode area	Postcode area. such as SE
SFBB availability (% premises)	Percentage of premises that have Superfast Broadband (30Mbit/s to less than 300Mbit/s) coverage
UFBB availability (% premises)	Percentage of premises that have Ultrafast Broadband (300Mbit/s or greater) coverage
Full Fibre availability (% premises)	Percentage of premises that have coverage from a full fibre service
% of premises unable to receive 2Mbit/s	Percentage of premises that do not have access to services above 2Mbit/s
% of premises unable to receive 5Mbit/s	Percentage of premises that do not have access to services above 5Mbit/s
% of premises unable to receive 10Mbit/s	Percentage of premises that do not have access to services above 10Mbit/s
% of premises unable to receive 30Mbit/s	Percentage of premises that do not have access to services above 30Mbit/s
% of premises below the USO	Percentage of premises that do not have access to download speeds at or above 10Mbit/s and upload speeds at or above 1Mbit/s including non-matched records and zero predicted speeds
% of premises able to receive decent broadband from FWA	Percentage of premises with Decent Broadband from Fixed Wireless Access provision
% of premises able to receive SFBB from FWA	Percentage of premises with Superfast Broadband (30Mbit/s or above) from Fixed Wireless Access provision
% of premises with NGA	Percentage of premises with Next Generation Access

Table A.54. Data Profile: Technology Data

Performance variables

Table A.55. Description of Variables: Technology Data

Current headers	Note
postcode	Postcode with spaces removed, such as SE19HA
postcode_spaces	Postcode with single space, such as SE1 9HA
postcode area	Postcode area, such as SE
Median download speed (Mbit/s)	Median download speed (in Mbit/s) of all connections
Average download speed (Mbit/s)	Average download speed (in Mbit/s) of all connections
Minimum download speed (Mbit/s)	Minimum download speed (in Mbit/s) of all connections To Be Removed in 2019

Maximum download speed (Mbit/s)	Maximum download speed (in Mbit/s) of all connections
Average download speed (Mbit/s) for lines < 10Mbit/s	Average download speed (in Mbit/s) of all connections that deliver speed less than 10Mbit/s
Average download speed (Mbit/s) for lines 10<30Mbit/s	Average download speed (in Mbit/s) of all connections that deliver speeds of 10Mbit/s or higher and less than 30Mbit/s
Average download speed (Mbit/s) for SFBB lines	Average download speed (in Mbit/s) of all connections that deliver speeds of 30Mbit/s to less than 300Mbit/s
Average download speed (Mbit/s) for UFBB lines	Average download speed (in Mbit/s) of all connections that deliver speeds of 300Mbit/s or greater
Median upload speed (Mbit/s)	Median upload speed (in Mbit/s) of all connections
Average upload speed (Mbit/s)	Average upload speed (in Mbit/s) of all connections
Minimum upload speed (Mbit/s)	Minimum upload speed (in Mbit/s) of all connections To Be Removed in 2019
Maximum upload speed (Mbit/s)	Maximum upload speed (in Mbit/s) of all connections
Average upload speed (Mbit/s) for lines < 10Mbit/s	Average upload speed (in Mbit/s) of all connections that deliver speed less than 10Mbit/s
Average upload speed (Mbit/s) for lines 10<30Mbit/s	Average upload speed (in Mbit/s) of all connections that deliver speed less than 30Mbit/s
Average upload speed (Mbit/s) for SFBB lines	Average upload speed (in Mbit/s) of all connections that deliver speeds of 30Mbit/s to less than 300Mbit/s
Average upload speed (Mbit/s) for UFBB lines	Average upload speed (in Mbit/s) of all connections that deliver speeds of 300Mbit/s or greater
Number of connections < 2 Mbit/s (number of lines)	Count of connections that have average measured speeds below 2Mbit/s
Number of connections 2<5 Mbit/s (number of lines)	Count of connections that have average measured speeds of 2Mbit/s or greater and below 5Mbit/s
Number of connections 5<10 Mbit/s (number of lines)	Count of connections that have average measured speeds of 5Mbit/s or greater and below 10Mbit/s
Number of connections 10<30 Mbit/s (number of lines)	Count of connections that have average measured speeds of 10Mbit/s or greater and below 30Mbit/s
Number of connections 30<300 Mbit/s (number of lines)	Count of connections that have average measured speeds of 30Mbit/s or greater and below 300Mbit/s

Number of connections >= 300 Mbit/s (number of lines)	Count of connections that have average measured speeds of 300Mbit/s or greater
Number of connections >= 30 Mbit/s (number of lines)	Count of connections that have average measured speeds of 30Mbit/s or greater
Average data usage (GB)	Average data usage, download and upload, in GB of all connections
Average data usage (GB) for lines < 10Mbit/s	Average data usage, download and upload, in GB of all connections that deliver speed less than 10Mbit/s
Average data usage (GB) for lines 10<30Mbit/s	Average data usage, download and upload, in GB of all connections that deliver speed less than 30Mbit/s
Average data usage (GB) for SFBB lines	Average data usage, download and upload, in GB of all connections that deliver speeds of 30Mbit/s or greater
Average data usage (GB) for UFBB lines	Average data usage, download and upload, in GB of all connections that deliver speeds of 300Mbit/s or greater

Relevant Documents

The following documents are included in the folder attached to this document.

Document Name	Name file
About this data fixed local	About-this-data-fixed-local-and-unitary-authority-
and unitary authority	201809.pdf
About this data fixed output	About-this-data-fixed-output-area-201809.pdf
area	
About this data fixed	About-this-data-fixed-parliamentary-constituency-
parliamentary constituency	201809.pdf
About this data fixed	About-this-data-fixed-postcode-201809.pdf
postcode	
About this data mobile local	About-this-data-mobile-local-and-unitary-authority-
and unitary authority	201809.pdf
About this data mobile	About-this-data-mobile-parliamentary-constituency-
parliamentary constituency	201809.pdf
Connected Nations 2018_	Connected Nations 2018_ Data downloads - Ofcom.pdf
Data downloads - Ofcom	
Connected Nations 2018	connected-nations-2018-methodology.pdf
Methodology	

Table A.56. Relevant Documents: Technology Data

Quality

Overall, Ofset provides enough methodological supplementary documents to allow researchers to understand the methodology, the process, data sets and assumptions used in compiling and presenting data. However, there is no metadata contained in the data to identify quality differences between the records processed.

Data Profile 17: Political Data

Name of the Data: Political Data 34

Description of the Data

The British Local Elections Database is a unique database of local election results in Great Britain. The data is structured in two parts. First, the period of the twentieth century before wholesale local governments reorganization in 1973. The second period, from 1973 to 2003, covering the existing English, Welsh and Scottish local elections.

Data Profile

Field	Description
Dataset ID	17
Dataset Name	Political data
Temporal Extent	1911 – 2003
Supplier/data	Various
owner	
Analytical Units	Polygons (Westminster constituency)
Data Format	CSV
Geographical Extent	Great Britain
Access	Open access
Cost	Free
Observations	Circa 200,000 individual local election results
Web page	https://impact.ref.ac.uk/casestudies/CaseStudy.aspx?Id=4601 http://www.visionofbritain.org.uk/atlas/theme/T_POL

Table A.57. Data Profile: Political Data

Structure of the Data

The British Local Elections Database is structured in two periods: The first period, from 1889 to 1972, cover 50,000 election results of mainly urban areas, including English, Welsh and Scottish county boroughs between 19911 and 1972, and all London area contests in the periods 1989-1971. The second period, form 1973 to 2003, contains in excess of 150,000 individual local election results. The following tables summarize the available data for each period.

Table A.58. Summary	of British I	Local Elections –	1889 - 1972
---------------------	--------------	-------------------	-------------

Council type	N of cases	First year	Lst year	
County Borough Council	356,678	19 <i>′</i>	1	1972
County Council	5,270	195	58	1970
Greater London Council	96	196	64	1970

³⁴ This document is a summary of relevant elements contained in the document 'project_guide.pdf' and Rallings, C.S., Thrasher, M.A.M. and Ware, L., British Local Election Database, 1889-2003. Colchester, Essex: UK Data Archive, June 2006. SN: 5319, http://dx.doi.org/10.5255/UKDASN53191

London Borough Council	7,361	1900	1971
London County Council	1,214	1889	1961

Table A.59. Summary of British Local Elections – 1973 - 2003

Council type	N of cases	First year	Lst year
County Council	27,633	1973	2001
District Council	87,066	1973	2003
Greater London Council	276	1973	1981
London Borough Council	5,827	1974	2002
Metropolitan Borough Council	18,995	1973	2003
Metropolitan County Council	1,641	1973	1981
Unitary Authority	9,879	1974	2003

Relevant Documents

The following documents are included in the folder attached to this document "

Document Name	Name file
Impact case study (Res 2014)	DetailsofData.pdf
British Local Elections Database (web)	British Local Elections Database_web.pdf
Summary of Database	project_guide.pdf
British Local Election	British Local Election Database.pdf
Database, 1889-2003	

Table A.60. Relevant Documents: Political Data

Quality

The British Local Election Database, 1889-2003 is a high quality ESRC Project currently deposited with the UK Data Archive (study number 5319).

Data Profile 18: Health Data

Name of the Data: Health Data 35

Description of the Data

General practice prescribing data is a list of all medicines, dressings and appliances that are prescribed and dispensed each month. A record will only be produced when this has occurred and there is no record for a zero total.

For each practice in England, including GP Practices, the following information is presented at presentation level for each medicine, dressing and appliance, (by presentation name):

- Total number of items prescribed and dispensed
- Total net ingredient cost
- Total actual cost
- Total quantity

The data covers NHS prescriptions written in England and dispensed in the community in the UK. Prescriptions written in England but dispensed outside England are included. The data includes prescriptions written by GPs and other non-medical prescribers (such as nurses and pharmacists) who are attached to GP practices.

Practices are identified only by their national code, so an additional data file - linked to the first by the practice code – provides further detail in relation to the practice.

Presentations are identified only by their BNF code, so an additional data file - linked to the first by the BNF code - provides the chemical name for that presentation.

Data Profile

Field	Description			
Dataset ID	19			
Dataset Name	Health data			
Temporal Extent	2010 - 2019			
Supplier/data NHS				
owner				
Analytical Units	Points			
Data Format	CVS			
Geographical Great Britain Extent Great Britain				
Access Open Government Licence				
Cost	Free			
Observations	Approx. 10,000 GPs			
Web page	https://data.gov.uk/dataset/176ae264-2484-4afe-a297- d51798eb8228/gp-practice-prescribing-data-presentation-level			

Table A.61. Data Profile: Health Data

³⁵ This document is a summary of relevant elements contained in the link

https://www.ofcom.org.uk/research-and-data/multi-sector-research/infrastructure-research/connected-nations-2018/data-downloads

Structure of the Data

Practice prescribing data

Table A.62. Description of Variables: Health Data - Practice prescribing data

Variable	Description			
SHA ¹	SHA or AT code - 3 characters			
PCT ²	PCT or CCG code - 3 characters			
PRACTICE	ANNNNN Practice code - 6 characters			
BNF CODE	British National Formulary (BNF) code - 15 characters			
BNF NAME	BNF presentation name - 40 characters			
ITEMS	Prescription items dispensed - whole numbers			
NIC	Net ingredient cost - pounds and pence			
ACT COST	Actual cost - pounds and pence			
QUANTITY	Quantity - whole numbers			
PERIOD	YYYYMM			

¹ For data covering August 2010 to March 2013, the SHA field refers to Strategic Health Authority. Following changes to the NHS structure, from April 2013 onwards, the SHA field refers to Area Team.

² For data covering August 2010 to March 2013, the PCT field refers to Primary Care Trust. Following changes to the NHS structure, from April 2013 onwards, the PCT field refers to Clinical Commissioning Group (CCG).

Practice level prescribing - practice codes, names and addresses

Variable	Description			
PERIOD	YYYYMM			
Practice Code	ANNNNN practice code - 6 characters			
Practice Name	Name of surgery			
Address 1	ANY VILLA SURGERY			
Address 2	1 ANY ROAD			
Address 3	ANYTOWN			
Address 4	ANYSHIRE			
Postcode	XX2 7XX			

Table A.63. Description of Variables: Health Data - Practice codes, names and address Variable Description

Practice level prescribing chemical names and bnf code file

Table A.64. Description of Variables: Health Data - prescribing chemical and bnf codes

Variable	Description
CHEM SUB	Chemical BNF code - 9 characters
NAME	Chemical Name - 60 characters

CSV file can be used to translate the presentation level 15 character BNF code of the main data file in to chemical level by using the first 9 characters. For dressings and appliances, where the concept of a "chemical" is not appropriate, only the 4 digit code is used. summary="Chemical names and bnf code file"

Relevant Documents

The following documents are included in the folder attached to this document.

Document Name	Name file
GP practice prescribing data	GP practice prescribing data - Presentation level -
- Presentation level	data.gov.uk.pdf
Practice level prescribing -	Practice level prescribing - glossary of terms - NHS
glossary of terms	Digital.pdf
Practice level prescribing	Practice level prescribing data_ more information - NHS
data: more information	Digital.pdf

Table A.65. Relevant Documents: Health Data

Quality

All prescription statistics in this data are based on information systems at NHS Prescription Services, part of the NHS BSA. The figures used are collected as an essential part of the process of reimbursing dispensers (mainly pharmacists and dispensing doctors) for medicines supplied. All prescriptions which are dispensed in England need to be submitted to NHS BSA if the dispenser is to be reimbursed and so coverage should be complete.

NHS Prescription Services quality assures the data they provide. They state that due to the complex and manual processes involved there may be inaccuracies in capturing prescription information which are then reflected in the data.

Currently the prescription processing activity is internally audited to 97.5% accuracy (meaning at least 97.5% of prescriptions are recorded accurately). While the BSA can adjust payments to dispensers if an error is identified they are unable to correct the prescribing data.

Data Profile 19: Educational Data

Name of the Data: Educational Data³⁶

Description of the Data

The National Pupil Database (NPD) is one of the richest education datasets in the world holding a wide range of information about pupils who attend schools and colleges in England. It forms a significant part of the evidence base for the education sector and supports a number of key priorities around accountability and school improvement. The first version of the National Pupil Database including information from the first pupil level school census matched to attainment information was produced in 2002.

Data Profile

Field	Description		
Dataset ID	20		
Dataset Name	Education		
Temporal Extent	2007-2018		
Supplier/data Ofsted/DfE owner			
Analytical Units	Points, Polygons		
Data Format	CVS?		
Geographical Extent	Great Britain		
Access	Restricted (Department for Education)		
Cost	Free		
Observations	-		
Web page	https://www.gov.uk/guidance/how-to-access-department-for- education-dfe-data-extracts		

Table A.66. Data Profile: Educational Data

Structure of the Data

A number of National Pupil Database (NPD) de-identified individual level 'standard extracts' for each academic year are available in the Secure Research Service. It will be possible to link these extracts at a pupil level. An overview of the standard extracts is in the table below.

Table A.67. Overview of the standard extracts - NPD

Data extract	Academic years
Early Years Census data and Alternative Provision Census	2007/08 to 2017/18
School Census data (autumn, spring and summer)	2012/13 to 2017/18
3-Term absence data, exclusions data, and post 16 learning aims – from the school census (PLAMS Amended)	2012/13 to 2016/17

³⁶ This document is a summary of relevant elements contained in the link <u>https://www.gov.uk/guidance/how-to-access-department-for-education-dfe-data-extracts</u> and in the document 'npd user guide and protocol july 2012.pdf'

National Client Caseload Information System (NCCIS)	2011/12 to 2016/17
Early years foundation stage profile (EYFSP) attainment data, with key pupil data from the appropriate spring School Census	2012/13 to 2016/17
Key stage 1 pupil data	2012/13 to 2016/17
Key stage 1 pupil data with key pupil data from the appropriate spring School Census, matched to Early years foundation stage profile (EYFSP) attainment data	2012/13 to 2016/17
Key stage 2 pupil data	2012/13 to 2016/17
Key stage 2 pupil data, with key pupil data from the appropriate spring School Census, matched to key stage 1 prior attainment	2012/13 to 2016/17
Key stage 2 exam level data	2012/13 to 2016/17
Key stage 4 pupil data	2012/13 to 2016/17
'Final' key stage 4 pupil data, with key pupil data from the appropriate spring School Census, matched to key stage 3 (where available), key stage 2, key stage 1 prior attainment	2012/13 to 2016/17
Key stage 4 exam level data	2012/13 to 2016/17
Key stage 5 student data	2012/13 to 2016/17
'Final' key stage 5 student data, with key pupil data from the appropriate spring School Census, matched to key stage 4, key stage 3, key stage 2 and key stage 1 prior attainment	2012/13 to 2016/17
Key stage 5 exam level data	2012/13 to 2016/17

For details of each individual data see 'The National Pupil Database, User Guide and Protocol' document.

Relevant Documents

The following documents are included in the folder attached to this document.

Table A.68. Relevant Documents: Educational Data

Document Name	Name file	
The National Pupil Database, User Guide and Protocol	npd user guide and protocol	july 2012.pdf

Quality

The NPD quality is ensured by the Department for Education.

Data Profile 20: Geo demographic segmentation Data

Name of the Data: Geo demographic segmentation Data³⁷

Description of the Data

Acorn produces geo demographic segmentation of residential neighbourhoods in the UK. It classifies each postcode in the country into one of 62 types that give a distinctive picture of the kinds of people who live in an area, their attitudes and how they behave.

The Acorn segmentation has a hierarchical structure. The 62 types aggregate into 18 Acorn groups, which lie within 6 descriptive Acorn categories at the top level. Acorn types are also further subdivided into 313 detailed micro-segments. These micro-segments may be appropriate for analysis of areas such as inner city council areas within which there may be a relatively limited variety of Acorn types.

Data Profile

	Description		
Field			
Dataset ID	21		
Dataset Name Geodemographic data			
Temporal Extent	2003, 2015, 2017		
Supplier/data owner	CACI		
Analytical Units	Points, Polygons (Postcodes and output areas)		
Data Format	CVS		
Geographical Extent	Great Britain		
Access	Available to Registered or Authorised Users, UK Data Service		
Cost	Free		
Observations	Approx. 2.3 million postcodes		
Web page	https://beta.ukdataservice.ac.uk/datacatalogue/studies/study?id=8196		

Table A.69. Data Profile: Geo demographic segmentation Data

³⁷ This document is a summary of relevant elements contained in the link https://beta.ukdataservice.ac.uk/datacatalogue/studies/study?id=8196

Structure of the Data:

Ca	ategories	Тур	Des
1/	Affluent Achievers		
Α	Lavish Lifestyles	1	Exclusive enclaves
	-	2	Metropolitan money
		3	Large house luxury
В	Executive Wealth	4	Asset rich families
		5	Wealthy countryside commuters
		6	Financially comfortable families
		7	Affluent professionals
		8	Prosperous suburban families
		9	Well-off edge of towners
С	Mature Money	10	Better-off villagers
		11	Settled suburbia, older people
		12	Retired and empty nesters
		13	Upmarket downsizers
2	Rising Prosperity		
D	City Sophisticates	14	Townhouse cosmopolitans
		15	Younger professionals in smaller flats
		16	Metropolitan professionals
		17	Socialising young renters
Е	Career Climbers	18	Career driven voung families
		19	First time buyers in small, modern homes
		20	Mixed metropolitan areas
			•
3	Comfortable Communities		
F	Countryside Communities	21	Farms and cottages
		22	Larger families in rural areas
		23	Owner occupiers in small towns and villages
G	Successful Suburbs	24	Comfortably-off families in modern housing
		25	Larger family homes, multi-ethnic areas
		26	Semi-professional families, owner occupied
		nei	ghbourhoods
Н	Steady Neighbourhoods	27	Suburban semis, conventional attitudes
		28	Owner occupied terraces, average income
		29	Established suburbs, older families
	Comfortable Seniors	30	Older people, neat and tidy peichbourboods
•	Comoltable Semols	31	Elderly singles in purpose-built accommodation
<u> </u>		01	
J	Starting Out	32	Educated families in terraces, young children
L		33	Smaller houses and starter homes
4	Financially Stretched		
Κ	Student Life	34	Student flats and halls of residence
		35	Term-time terraces
		36	Educated young people in flats and tenements

 Table A.70. Description of Variables: Geo demographic categories

L	Modest Means	37	Low cost flats in suburban areas
		38	Semi-skilled workers in traditional
		neig	hbourhoods
		39	Fading owner occupied terraces
		40	High occupancy terraces, many Asian families
Μ	Striving Families	41	Labouring semi-rural estates
		42	Struggling young families in post-war terraces
		43	Families in right-to-buy estates
		44	Post-war estates, limited means
Ν	Poorer Pensioners	45	Pensioners in social housing, semis and terraces
		46	Elderly people in social rented flats
		47	Low income older people in smaller semis
		48	Pensioners and singles in social rented flats
5	Urban Adversity		
0	Young Hardship	49	Young families in low cost private flats
		50	Struggling younger people in mixed tenure
		51	Young people in small, low cost terraces
Ρ	Struggling Estates	52	Poorer families, many children, terraced housing
		53	Low income terraces
		54	Multi-ethnic, purpose-built estates
		55	Deprived and ethnically diverse in flats
		56	Low income large families in social rented semis
Q	Difficult Circumstances	57	Social rented flats, families and single parents
		58	Singles and young families, some receiving
		ben	efits
		59	Deprived areas and high-rise flats
6	Not Private Households		
R	Not Private Households	60	Active communal population
		61	Inactive communal population
		62	Business addresses without resident population

Relevant Documents

The following documents are included in the folder attached to this document.

Document Name	Name file
The consumer classification	8196_acorn_group_personas_2017.pdf
Technical guide	8196_acorn_technical_guide.pdf
Classification Table	8196_acorn_user_guide-5.xls
User guide	8196_acorn_user_guide.pdf

Quality

Acorn processes information from different sources and assess the quality of each data source according to the following criteria:

• Supplier's Own documentation of their Quality control systems

- Independently conducted Quality assessments
- Cross-check against Other suppliers of the Same information
- Cross-check against Independent national or regional Aggregate data sources
- Visual Checks (using internet searches, maps, Streetview or similar)
- Check against Independently collected Panel information
- Consistency checks with Other data sources
- Internal Consistency checks

The quality of each data source is also check with other sources when it is possible. Besides, Acron assess his final classification, and other associated data, using GINI scores,

which measure the effectiveness of Acorn in discriminating across a wide range of variables. More details are described in the 'Technical guide' document. Although key issues for some datasets are outlines in the technical documentation, Acron do not publish detailed results of guality assessments.

Data Profile 21: Census Data 2001 and 2011 with InFuse

Name of the Data: Census Data 2001 and 2011 with InFuse 38

Description of the Data

InFuse is a free service providing easy access to aggregate data from the UK 2011 and 2001 censuses

UK censuses are surveys of the entire population of the UK carried out every ten years. Their primary purpose is to provide the UK government with high quality information about demographic and socio-economic characteristics on which to base policy and spending decisions. The information is also used for a wide variety of other purposes across all sectors of society.

Aggregate data from UK censuses are essentially counts of the numbers of the people and households recorded in the census surveys as possessing particular combinations of characteristics within geographical areas at a range of scales from the whole UK down to Output Areas.

Data Profile

Field	Description
Dataset ID	22
Dataset Name	Census data
Temporal Extent	1971, 1981, 1991, 2001, 2011
Supplier/data owner	ONS
Analytical Units	Polygons (Output areas)
Data Format	Geodatabases (GDB), Shapefile (SHP), CVS
Geographical Extent	Great Britain
Access	Open access
Cost	Free
Observations	-
Web page	http://infuse.ukdataservice.ac.uk/

Table A.72. Data Profile: Census Data 2001 and 2011

Structure of the Data

See the document "2011 Topic descriptions_ InFuse.doc" for a detailed description of all variables.

³⁸ This document is a summary of relevant elements contained in the link <u>http://infuse.ukdataservice.ac.uk/</u>

Relevant Documents

The following documents are included in the folder attached to this document.

Table A.75. Relevant Documents. Census Data 2001 and 2011								
Document Name	Name file							
2011 Topic descriptions	2011 Topic descriptions_ InFuse.pdf							
InFuse								

Table A.73. Relevant Documents: Census Data 2001 and 2011

Quality

The quality of Census data is ensured by the Office for National Statistics (ONS).

Appendix B

This Appendix includes the Linkage Potential Index (LPI) scores.

B.1. Linkage Potential Index

Table B.1. Linkage Potential Index (LPI) scores

ID	Name	LPI	QA	TE	QA 1	QA 2	QA 3	TE 1	TE 2	TE 3	TE 4
1	MasterMap Topography Layer	5.33	3.00	2.33	0.00	0.00	0.00	0.00	-0.37	-0.61	-0.69
2	MasterMap Highways Network	5.47	3.00	2.47	0.00	0.00	0.00	0.00	-0.30	-0.57	-0.66
3	MasterMap Building Heights	4.72	3.00	1.72	0.00	0.00	0.00	-0.26	-0.53	-0.71	-0.77
4	MasterMap Imagery Layer	5.40	3.00	2.40	0.00	0.00	0.00	0.00	-0.33	-0.59	-0.67
5	MasterMap Water Network Layer	3.25	3.00	0.25	0.00	0.00	0.00	-0.89	-0.93	-0.96	-0.97
6	MasterMap Green Space Layer	5.33	3.00	2.33	0.00	0.00	0.00	0.00	-0.37	-0.61	-0.69
7	AddressBase Premium	6.77	3.00	3.77	0.00	0.00	0.00	0.00	0.00	-0.02	-0.21
8	Points of Interest	4.60	3.00	1.60	0.00	0.00	0.00	-0.32	-0.57	-0.73	-0.79
9	Land Registry Price Paid dataset	3.61	1.00	2.61	-1.00	-1.00	0.00	0.00	-0.23	-0.53	-0.62
10	Energy Performance Certificate data	2.23	1.00	1.23	-1.00	-1.00	0.00	-0.47	-0.67	-0.80	-0.84
11	National Register of Social Housing (NROSH)	1.49	1.00	0.49	-1.00	-1.00	0.00	-0.79	-0.87	-0.92	-0.93
12	Historic Ordnance Survey data	7.00	3.00	4.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
13	Met Office data	6.93	3.00	3.93	0.00	0.00	0.00	0.00	0.00	0.00	-0.07

14	Pollution data	5.33	3.00	2.33	0.00	0.00	0.00	0.00	-0.37	-0.61	-0.69
15	Crime data	2.98	2.00	0.98	-1.00	0.00	0.00	-0.58	-0.73	-0.84	-0.87
16	Technology data	2.61	2.00	0.61	-1.00	0.00	0.00	-0.74	-0.83	-0.90	-0.92
17	Political data	7.00	3.00	4.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
18	Health data	2.98	2.00	0.98	-1.00	0.00	0.00	-0.58	-0.73	-0.84	-0.87
19	Education	3.35	2.00	1.35	-1.00	0.00	0.00	-0.42	-0.63	-0.78	-0.82
20	Geodemographic	3.84	2.00	1.84	-1.00	0.00	0.00	-0.21	-0.50	-0.69	-0.75
	data										
21	Census data	6.73	3.00	3.73	0.00	0.00	0.00	0.00	0.00	-0.04	-0.23

Notes: QA items are defined as: QA1) The data include variables in the datasets that helps to identify errors or assumptions in the data process QA2) The ownership of the data flags the overall quality of the datasets within the supplementary documents QA3) Was possible to find enough documentation to identify the variables included in the data; TE items are defined as: TE1) Millennium Cohort Study; TE2) Next Steps, TE3) British Cohort Study, TE4) National Child Development Study.