

Data Documentation: NIC project

Jonathan Halket, Mateusz Mysliwski, Lars Nesheim, Polly Simpson

Centre for Microdata Methods and Practice
IFS

July 26, 2018

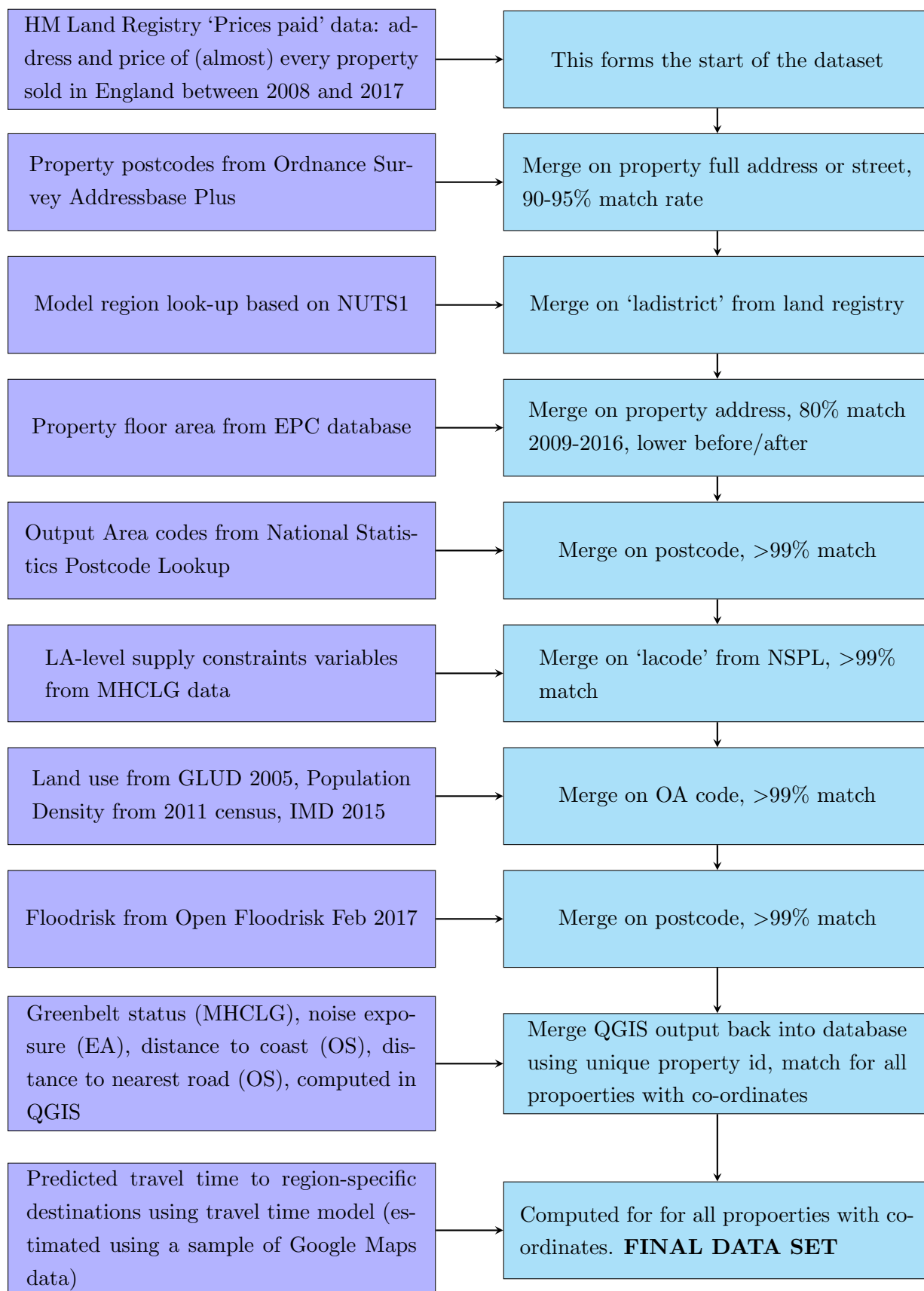
Abstract

This documents provides details of the data used for the Property Value Uplift (PVU) Tool.

Contents

1	Overview of data construction	3
2	Data sources	4
2.1	Land Registry Price Paid	4
2.2	Addressbase Plus	4
2.3	Energy Performance Certificate database	5
2.4	ONS Local authority to Region lookup	6
2.5	National Statistics Postcode Lookup (NSPL)	6
2.6	Local plans	6
2.7	Number of dwellings by tenure and district, England	6
2.8	DCLG LA-level land constraints data land	7
2.9	Generalised Land Use Database 2005	7
2.10	Population Density	8
2.11	Index Multiple Deprivation	8
2.12	Open Flood Risk	8
2.13	Greenbelt	9
2.14	Road Noise	9
2.15	Google Maps API	10
2.16	VOA Rating list	10

1 Overview of data construction



2 Data sources

2.1 Land Registry Price Paid

Brief description: Covers (almost all) residential property sales in England Wales from 1995 to the current day. Exceptions include properties sold or transferred for below market value (e.g. right to buy sales at a discount, gifts, compulsory purchase orders)

Variables: Information on transacted price as well as property characteristics, including:

- **pricepaid:** Sale price stated on the transfer deed
- **propertytype:** D = detached, S = semi-detached, T = terraced, F = flat, O = other
- **newbuild:** Y = new build, N = established residential building
- **tenure:** F = freehold, L = leasehold. (Leasehold is very strongly correlated with being a flat)
- For more information see: <https://www.gov.uk/guidance/about-the-price-paid-data>.

Source and version used: Available from gov.uk <https://www.gov.uk/government/statistical-data-sets/price-paid-data-downloads>. The model uses data from 2008 to November 2017.

Download details: Data is released monthly, but 'year-to-date' available in a single file. Although the data is available as a single file containing all years, it's advisable to download only a year at a time as processing data files of that size can be slow. Yearly files are collated on the date of the transaction/deed date rather than the date that the information was lodged with HM Land Registry.

Licence: Contains HM Land Registry data©Crown copyright and database right 2017. This data is licensed under the Open Government Licence v3.0.

Limitations: Data doesn't contain sales of land or non-domestic property.

2.2 Addressbase Plus

Brief description: Addressbase is the Ordnance Survey's flagship database of Addresses in the UK.

Variables: The data contains co-ordinates for each address, as well as unique identifiers (TOIDs) which can be used to link addresses to OS Mastermap topography and transport layers.

The best way to store address information is not immediately obvious. The Ordnance Survey chose to break down addresses into their most basic building blocks - this creates a large number of variables but can be very precise. Central to their system are primary and secondary addressable objects (PAO/SAO respectively). Take the address Flat 4, 17 Stapleton Road, Bristol. Here the primary addressable object is 17. Stapleton road is the street or thoroughfare. Flat 4 is the secondary addressable object (as it is a subset of a larger building). PAO and SAO

can both either be a number, or text, or both. For example:

- `pao_start_number` - number on the street that the building is e.g. 15 Newbridge Road, comes after `pao_text`
- `pao_start_suffix` - e.g. the 'A' bit of 15A Newbridge Road
- `pao_end_number` - e.g. the 16 bit of 15-16 Newbridge Road
- `pao_end_suffix` - e.g. the 'A' bit of 15-16A Newbridge Road
- `pao_text` - e.g. the 'Radley House' bit of Radley House, 15-16A Newbridge Road

Source and version used: AddressBase Plus June 2017

Download details: Data for England comes in 36 csv files (15.4GB).

Licence: Normally, access to AddressBase Premium is subject to a fee. However it is available to public sector organisations for free under the public sector mapping agreement.

2.3 Energy Performance Certificate database

Brief description:DCLG Energy Performance Certificate (EPC) data is based on assessments of energy efficiency which are mandated for almost all properties in England and Wales when constructed, sold or let. Since 2008 assessors have been required to submit the underlying data for these assessments to DCLG. DCLG publishes this data, except where:

- the holder of the energy certificate has 'opted-out' of disclosure
- energy certificates are excluded on grounds of national security
- energy certificates are marked as 'cancelled' or 'not for issue'
- DEC's that can be identified as 'voluntary' (some organisations choose to have a DEC produced even though they are not required to do so by the regulations) are also excluded

Variables: As well as address and assessment data, the data also contains information on:

- Total floor area
- Window glazing
- Heating systems
- Energy efficiency
- Story of the building that the property is on

Source and version used: Available to download here:

<https://epc.opendatacommunities.org/>. Downloaded for the project 01/12/2017.

Download details: Bulk download comes as one file per local authority.

Licence: The data is at the address level and is therefore considered personal data subject to the Data Protection Act. The data can be used for research, but only after the user registers and

confirmed that they understand and agree to a set of licence conditions. The guidance does not place specific restrictions on linking this data, and DCLG have confirmed with us that permission is not needed to link the data to other sources. See <https://epc.opendatacommunities.org/docs/copyright>

Limitations: The address field is structured in a different way to Land Registry/OS data.

2.4 ONS Local authority to Region lookup

Brief Description: Geographic lookup from local authority code to statistical region name and code.

Source and version: Local Authority District to Region (December 2015) Lookup in England. <https://ons.maps.arcgis.com/home/item.html?id=17a30c262cf44ad0b9780d97c5d4a856>

Download Details: Available as csv file.

Licence: Open Government Licence. Contains National Statistics data © Crown copyright and database right 2018

2.5 National Statistics Postcode Lookup (NSPL)

Brief Description: Geographic lookup from postcode to most other levels of administrative/statistical geography in the UK. Includes population-weighted centroids for each postcode.

Source and version: Available to download from ONS Open Geography Portal. Latest version (as used for project) is November 2017.

<http://geoportal.statistics.gov.uk/datasets/national-statistics-postcode-lookup-latest-centroids>

Licence: Open Government Licence. Contains National Statistics data © Crown copyright and database right 2018

2.6 Local plans

Brief Description: Local authorities are required to produce local plans for future housing development. DCLG have collated details of the published plans by local authority. The data includes the total number of planned houses over the lifetime of each authority's plan, and the annual equivalent number of houses.

Source and version: Correspondance with DCLG (now MHCLG). Current as of June 2017.

Licence: N/A

2.7 Number of dwellings by tenure and district, England

Brief Description: Total number of dwellings by local authority district in England.

Source and version: Table 100 from the DCLG live tables on dwelling stock. Current as of

July 2017.

<https://www.gov.uk/government/statistical-data-sets/live-tables-on-dwelling-stock-including-vacants>

Licence: Open Government Licence

2.8 DCLG LA-level land constraints data land

Brief Description: DCLG statistics on the area of land in a local authority that is a) Green-belt b) Built-up c) AONB/SSSI/National Park etc. Includes total land area for calculation of percentages.

Source and version: Correspondence with DCLG in July 2017. Current as of 2015.

Licence: N/A

2.9 Generalised Land Use Database 2005

Brief Description: In 2001 and 2005, DCLG produced (as experimental statistics) the Generalised Land Use Database which classified all land in England into 10 land categories:

- Domestic buildings
- Domestic gardens
- Non-domestic buildings
- Roads
- Paths
- Rail
- Greenspace
- Water
- Other land uses (largely hard standing)
- Unclassified

The data is available as multiple geographical levels: Census Output Areas (OAs), Lower Layer Super Output Areas (LSOAs), Middle Layer Super Output Areas (MSOAs), Local Authorities (LAs), and Government Office Regions (GORs). The area of coverage is England.

Source and version: Census ward level data for 2005 downloaded from the national archives

<http://webarchive.nationalarchives.gov.uk/20060206095010/http://www.odpm.gov.uk/index.asp?id=1146084>

Download details: Available as spreadsheets that give the area of land (thousands of square metres) in each administrative area. I think it's also available in a way that means you can do more specific mappings but that you have to already have a public sector licence to use OS MasterMap.

Notes: You can find useful guides to the data here:

http://www.andywrightman.com/docs/GLUD_2005.pdf http://www.andywrightman.com/docs/GLUD_2006.pdf

Licence: Open Government Licence

2.10 Population Density

Brief Description: Population density at the output area level collected for the 2011 Census

Source and version: Downloaded from NOMIS (ONS official labour market statistics portal)

<https://www.nomisweb.co.uk/>

Licence: Open Government Licence. Contains National Statistics data © Crown copyright and database right 2018

2.11 Index Multiple Deprivation

Brief Description: The index of multiple deprivation (IMD) is a measure of very local level deprivation across multiple dimensions including health, income and education. Various summary measures of the IMD are available. For this analysis we use IMD decile at the output area level.

Source and version: IMD 2015 data downloaded from NOMIS (ONS official labour market statistics portal)

<https://www.nomisweb.co.uk/>

Licence: Open Government Licence. Contains National Statistics data © Crown copyright and database right 2018

2.12 Open Flood Risk

Brief Description: “Open Flood Risk by Postcode combines the Environment Agency’s Risk of Flooding from Rivers and Sea with English postcodes from Open Postcode Geo. Each postcode is supplemented with the information from the flood risk area in which it is located. Postcodes are treated as a single point in space, allowing for the postcode to be looked up and the flood risk determined for that point. Note the value of SUITABILITY for a given area for how suitable the risk level is for a given purpose.”

Categorizes floodrisk into the following categories, recorded in the prob_4band variable:

- High - Each year, there is a chance of flooding of greater than 1 in 30 (3.3%).
- Medium - Each year, there is a chance of flooding of between 1 in 30 (3.3%) and 1 in 100 (1%).
- Low - Each year, there is a chance of flooding of between 1 in 100 (1%) and 1 in 1000 (0.1%).

- Very Low - Each year, there is a chance of flooding of less than 1 in 1000 (0.1%).
- None - This value is added by GetTheData to indicate a postcode which is not in a flood risk area.

Source and version: OpenFloodRisk available <https://www.getthedata.com/open-flood-risk-by-postcode>, current as of February 2016. Details of original floodrisk information are available https://www.getthedata.com/docs/RoFRS_Product_Description_vs1.5.pdf

Download details: Available as a single CSV file, one observation per postcode

Licence: Derived from Risk of Flooding from Rivers and Sea. Contains Environment Agency data licenced under the Open Government Licence v3.0. Derived from Open Postcode Geo (<https://www.getthedata.com/open-postcode-geo>). Open Postcode Geo is derived from the ONS Postcode Directory which is licenced under the Open Government Licence and the Ordnance Survey OpenData Licence. Northern Irish postcodes have been removed as these are covered by a more restrictive licence.

2.13 Greenbelt

Brief Description DCLG produce a greenbelt dataset, in polygon form, of greenbelt boundaries in England. They do this by merging together information from different local authorities. They caveat that due to differences in local authorities digitisation methods the accuracy of the given boundaries may vary.

Source and version: Latest data available here: <https://data.gov.uk/dataset/a6c492ba-5ff8-44d8-bbd8-025162f4b773/english-local-authority-green-belt-dataset>. We use latest data available as of June 2017.

Download Details: Download is a shapefile (.shp).

Licence: Open Government Licence. Contains National Statistics data © Crown copyright and database right 2018

2.14 Road Noise

Brief Description: The EU Environmental Noise Directive requires the UK government to measure the noise exposure of large settlements and the area around major roads, railways and airports.¹ This was last undertaken in 2012. The Environment agency publishes the resulting assessment as open data. Multiple measures of road noise exposure are available, we use LAeq, 16h - an indicator of the annual average noise levels for the 16-hour period between 0700 2300 (in db).

Source and version: Details about the requirements and data, as well as download links are available here: <https://www.gov.uk/government/publications/open-data-strategic-noise-mapping>.

Download Details: Data is available in TAB, GML or shapefile format (we use shapefile).

¹http://ec.europa.eu/environment/noise/directive_en.htm

Licence: Open Government Licence

2.15 Google Maps API

Brief Description: Google Maps Distance Matrix API allows users to request travel times between a pair of co-ordinates or addresses, with varying assumptions on the travel mode used and traffic conditions. Requests can be made using various programming languages including Python and R. We use the R package ‘ggmap’ to make geo-coding and travel time requests. More details on how we use the google maps data are in the travel time model section of the final PVU tool report. Google’s online documentation can be found here: <https://developers.google.com/maps/documentation/distance-matrix/>

Licence: Travel times were downloaded using a Google Maps APIs Premium Plan.

2.16 VOA Rating list

Brief Description: The Valuation Office Agency is responsible for valuing all non-domestic property in England and Wales in order to determine the tax liability of occupants (specifically, Business Rates, a local tax on business property). They maintain a database containing the estimated market rent of almost all non-domestic property in England and Wales. These estimated rents are called ‘rateable values’ because they are used as the basis of business rates. The data covers 2 million commercial properties in England and Wales as of April 2017. Property coverage is almost, but not quite, complete. For example it doesn’t cover:

- “Some types including hotels, public houses, universities, schools and hospitals are covered in the Rating List but usually have no floor area data.
- “Very large industrial hereditaments are valued separately, and are omitted from both Rating List and SMV.
- “Three classes of activity are exempt from rates altogether: agricultural premises, places of worship and properties of Her Majesty the Queen. This last group besides Royal palaces includes the Ministry of Defence estate, which is extremely large. Floor area data in all these cases must be estimated or obtained from other non-VOA sources. Some area data are available from Display Energy Certificates (DECs) and Energy Performance Certificates (EPCs) (Department for Communities and Local Government, 2015).”²

Version and Source: Data available from the VOA. If you have agreement from VOA data can be downloaded here: <https://voaringlists.blob.core.windows.net/html/rlidata.htm>. The latest available data is based on the 2017 revaluation (release April 2017). Comprehensive property valuation carried out in 2010 and 2017, but incremental updates are carried out continuously as appeals take place.

Licence: VOA data has restricted licensing conditions, the terms of which are available at:

²Detail from Evans et al 2017 guide to 3DStock, <http://journals.sagepub.com/doi/pdf/10.1177/0265813516652898>

<https://www.tax.service.gov.uk/business-rates-find/terms-and-conditions>

It cannot, legally, be downloaded and used for research without making a data sharing agreement with the VOA. For this project a data sharing agreement was made with the VOA.

Additional details of data structure: There are two available datasets of rateable values from the VOA:

The Rating List: Covers 2 million properties, giving addresses and primary activities. Each observation is a different property. There is only one observation per property.

The Summary Valuation database: Covers a smaller number of properties (100,000 fewer), giving sub-activities and floor areas for each of these sub-activities.

The Summary Valuation database has a somewhat unconventional structure. V1 is an indicator telling you the type of record of the observation. These range from 01 to 07. A 01 record is the main record type, and is essentially the same as in the rating list. It tells you what assessment and what property subsequent observations are going to relate to so will be what we use to link the detailed info to the ratings list.

Subsequent records (before the next 01 record) all relate to this property. These are numbered 02 to 07. The vast majority of records are type 2. These tell you about each individual room of a 'property' (e.g. the most common are "Office", "Internal Storage", "Retail Zone A/B/C", "Staff Toilets", "Kitchen", "Workshop", "Store", "Warehouse"), what floor it's on, their floor space and contribution to the rateable value.

Type 3 records are 'additional items'. Includes things like bowling green, clubhouse, car wash, helipad, swimming pool etc.

Type 4 records are plant and machinery - only value is given.

Type 5 records are car parking which gives the amount and value of car parking.

Type 6 records are adjustments. These are explained here:

<http://app.voa.gov.uk/corporate/publications/manuals/ratingmanual/ratingmanualvolume4/sect7/d-rat-man-vol4-s7-app1.html>. Briefly, adjustments to value are often made at the end of valuation that reflect factors which will change the overall valuation of the property. The top factors here are listed below and are not insignificant determinants of property prices (median adj is -5%)

- Access
- Adjustment for Quality
- Divided or split unit
- Merged units
- Non-standard frontage
- Planning restriction
- Poor access

- Services
- Shape
- Shared access
- Size or quantity allowance
- Toilet facilities
- Variations in floor level
- Width to depth ratio

Type 07 gives you the total value of the property before adjustments and the cash value of the adjustment.

Notes: Guide to valuation <https://www.gov.uk/guidance/how-non-domestic-property-including-plant-and-machinery-is-valued>