Leeds EPC Data Methodology

Author: Douglas Phillips

Date: 2017/12/15

Methodology

This methodology details the process for producing or updating datasets using published EPC data, adding value and making the content more useable. The Department for Communities and Local Government (DCLG) publishes EPC data, licenced under the terms of the Open Government Licence v3.0., providing information on the energy performance of buildings to help contribute towards the Government's climate change adaptation and greenhouse gas emission reduction policies.

Access to data is important, however its usability is paramount. Considering the downloaded EPC data in its initial form, there are 80 individual columns for each certificate which results in the generation of a large cumbersome database. In the case of Leeds-based EPC data, the ~250,000 certificates pulled from the www.opendatacommunities.org website result in a database >200MB in size. This therefore needs to be extensively cleaned and reduced in size before any worthwhile research can be completed.

Reducing the Data

Although the certificates contain some interesting data, a lot of the columns can be removed to reduce the file size.

The guidance for the content can be found at https://epc.opendatacommunities.org/docs/guidance.

The columns that need to be retained for initial use are; ADDRESS1, POSTCODE, BUILDING_REFERENCE_NUMBER, INSPECTION_DATE, CURRENT_ENERGY_RATING, POTENTIAL_ENERGY_RATING, CURRENT_ENERGY_EFFICIENCY, POTENTIAL_ENERGY_EFFICIENCY, PROPERTY_TYPE, CONSTITUENCY, ENERGY_CONSUMPTION_CURRENT, ENERGY_CONSUMPTION_POTENTIAL, CO2_EMISSIONS_CURRENT, CO2_EMISSIONS_POTENTIAL, CO2_EMISS_CURRENT_PER_FLOOR_AREA, LIGHTING_COST_CURRENT, LIGHTING_COST_POTENTIAL, HEATING_COST_CURRENT, HEATING_COST_POTENTIAL, HOT_WATER_COST_CURRENT, HOT_WATER_COST_POTENTIAL, TOTAL_FLOOR_AREA and LOW_ENERGY_LIGHTING.

These remaining columns – reducing from the initial 80, down to a much more manageable 24 – will all have some use for producing the final databases. Although the data in the CO2_EMISS_CURRENT_PER_FLOOR_AREA column is not required for the final output, it is valuable in the process of cleaning the data. By sorting this column from smallest to largest, it highlights certificates with major discrepancies. A number of the EPC certificates contain minus numbers (in particular -99). The certificates with minus numbers are deleted as the focus of the datasets are domestic properties where efficiency improvements can be achieved.

The next step is to remove all duplicate EPC reports. The BUILDING_REFERENCE_NUMBER data attributes a unique number to each individual building – these numbers will not be replicated for any

other building's EPC certificate. Duplicate data can occur from two sources; 1) human/uploading error, and 2) following the update of a property's EPC certificate. In this case, the most up-to-date versions of the EPC certificates are desired. The dataset should be sorted by first using the INSPECTION_DATE column, sorting these from newest to oldest, then by the BUILDING_REFERENCE_NUMBER column – from smallest to largest. This results in the duplicated BUILDING_REFERENCE_NUMBER data being listed by the newest EPC certificate. Next, by using the remove duplicate tool available in Excel – specifying just the data contained in the BUILDING_REFERENCE_NUMBER column – will result in the removal of all the older duplicate certificates.

Capacity for Improvement

Once the data has been sorted it can then be made useful in the context of understanding an areas housing stock. The data in;

BUILDING_REFERENCE_NUMBER, INSPECTION_DATE, PROPERTY_TYPE, TOTAL_FLOOR_AREA, CURRENT_ENERGY_RATING, POTENTIAL_ENERGY_RATING, CURRENT_ENERGY_EFFICIENCY and LOW_ENERGY_LIGHTING

is all valuable and can be utilised without further exploitation. The rest of the columns can also be used, but they first need some simple manipulation. The certificates detail data for each property's current and potential performance for a number of indicators; the difference between the two can therefore be seen as the potential capacity for improvement (CfI). As a result, the CfI of each property has been calculated for the following areas;

Energy Efficiency (£/m²/year)

Energy Consumption (kWh/m²)

CO_2 Emissions (tonnes/year)

Lighting Costs (£/year)

Heating Costs (£/year)

Hot Water Heating Costs (£/year)

Total Costs (£/year)

This is achieved via a simple calculation, taking the difference between the current and potential values for each indicator area. When undertaking this task, it is important to calculate it using the absolute value function – this helps negate any issues regarding human error when the EPC data was initially inputted (there are cases where the current and potential values have been put in the wrong columns, resulting in negative values being produced).

Geographical Locations

In addition to understanding the CfI of the individual houses, it is important to detail the location of properties in a useful format. The location data contained within the published EPC certificates include details for each individual property's local authority, constituency and specific address (once cleaned the three important columns are; ADDRESS1, POSTCODE and CONSTITUENCY). The constituency data isn't detailed enough to show specific areas within a city – the constituencies in Leeds contain populations ranging from ~85,000 to 150,000 – while the address data is too detailed and specific

when considering its use at a city-level. Therefore, establishing a more appropriate resolution between these two levels is important for making the data useful. This section details the method for deriving the LSOA and ward location code for each specific BUILDING_REFERENCE_NUMBER.

Firstly, the address data needs to be turned into a format that can be used in geographic information system (GIS) software. By utilising the postcode data (permitted for use by the Royal Mail Copyright Notice), it is possible to derive easting and northing coordinate points using a batch postcode to grid reference converter tool. This will successfully convert around 95-97%+ of the postcodes, however for the remaining few – consisting predominantly of recent new builds etc – they require more attention. Using google maps and the postcodes in question (and the data from ADDRESS1, if required), the closest streets with convertible postcodes should be identified. The coordinates of a street, or a neighbouring street, are suitably detailed enough when wanting to establish which LSOA and ward they can be attributed to.

Once the coordinates have been derived, these can be imported into GIS software (in this case, QGIS has been used; a free, open-source and cross-platform GIS tool). Overlaying the plotted coordinates with shapefiles containing LSOA and, separately, ward spatial information it is possible to join attributes by location, using QGIS inbuilt tools. To keep inline with other work completed and databases used as part of this project, the LSOA shapefile utilised is based upon the 2001 boundaries (not the 2011 changes). The shapefile used for the ward classification uses the 2011 Census Merged Wards boundaries. The ward and LSOA attribution can be extracted as a .csv file and used as required.

As a result of the above methodology, two EPC datasets have been produced; 1) **Leeds EPC Dataset – Capacity for Improvement**, and 2) **Leeds EPC Dataset – Geographical Locations**. The EPC certificates used in their production were downloaded on the 09th November 2017 therefore if this method is to be used to update the current datasets, it would be advised for this to be done in November 2018.

Licensing Acknowledgements

Author: **Douglas Phillips**

Organisation: **ODI Leeds**

Date: **2017/11/30**

Title: **Leeds EPC Dataset – Capacity for Improvement**

The information contained within this dataset has been attained and derived from
https://epc.opendatacommunities.org/. This information is licensed under the terms of the Open
Government Licence v3.0 (http://www.nationalarchives.gov.uk/doc/open-government-
licence/version/3/).

Author: **Douglas Phillips**

Organisation: **ODI Leeds**

Date: **2017/11/30**

Title: **Leeds EPC Dataset – Geographical Locations**

The information contained within this dataset has been attained and derived from
https://epc.opendatacommunities.org/. This information is licensed under the terms of the Open
Government Licence v3.0 (http://www.nationalarchives.gov.uk/doc/open-government-
licence/version/3/). The derived location data contained within this dataset has used Royal Mail
Group Limited 'address data', permitted for use by the accompanying Royal Mail Copyright Notice
(https://epc.opendatacommunities.org/docs/copyright).

Licensing guidance found at; https://epc.opendatacommunities.org/docs/copyright

The Department for Communities and Local Government has published these data
in order to provide information about the energy performance of buildings. The
publication of these data is intended to contribute to delivering the Government's
policies to adapt to the effects of climate change and reduce greenhouse gas
emissions. It will provide data to facilitate improvements in the energy efficiency of
buildings through research, improved management and innovation.

Royal Mail Copyright Notice
These data include addresses and postcodes in which intellectual property rights are
owned by Royal Mail Group Limited ('address data'). The Department for
Communities and Local Government has published the address data with the
consent of Royal Mail Group Limited, which reserves all its copyright, database
rights, trade marks and other intellectual property rights.

Royal Mail Group Limited permits the use of 'address data' for specific purposes
related to the energy performance of buildings which are as follows:

- • The effective management of properties in the private or public sector with a view
  to promoting energy efficiency (through consideration of factors such as patterns of

use, maintenance, repairs, improvements and upgrades of building structure or energy using systems).

- • Research to provide data, analysis and statistics intended to improve the understanding of energy efficiency of buildings or to facilitate innovation and improvement in use, design, materials and equipment to this end.
- • Research into the effectiveness or impact of energy efficiency improvements including those delivered through Government or other energy improvement programmes.
- • Promoting and marketing energy efficiency improvements that may be made through a Government energy efficiency programme or initiative.
- • Promotion and better understanding of the current energy efficiency of buildings and potential improvement in the building sale or rental markets and/or by building occupiers or users.
- • To enable enforcement authorities to carry out enforcement duties as required by the Energy Performance of Buildings Regulations (England and Wales) 2012 (SI 2012/3118).
- • By local authorities in the exercise of their duties under section 91 of the Building Act 1984, or an approved inspector for the purposes of the inspector's functions under Part 2 of the Building Act 1984.
- • For the prevention or detection of crime, the apprehension or prosecution of alleged offenders, any proceedings in a court or tribunal or complying with an order or a court or a tribunal.

In addition to the purposes set out above there are certain exceptions to copyright which permit you to use copyrighted material for specific and limited purposes. These include non-commercial research and private study, subject to compliance with certain conditions. Further information about these exceptions can be found here: https://www.gov.uk/guidance/exceptions-to-copyright.
Where you publish or otherwise share address data with any other person for any of the purposes permitted under this copyright notice, you must include a copy of this copyright notice with the data.

Any use of address data which is not specified in this notice or covered by one of the copyright exceptions requires an appropriate license and is subject to Royal Mail Group Limited's relevant terms and conditions listed at www.poweredbypaf.com.
By using address data on this website you will be taken to agree to the terms set out in this copyright notice. Failure to comply with this copyright notice may lead to Royal Mail Group Limited taking legal action against you.

You can contact Royal Mail at address.management@royalmail.com.
Non-Address Data
All data fields other than the address and postcode data (address, address 1, address 2, address 3, postcode) available via this website are licensed under the Open Government Licence v3.0.