EPC data:   
Technical documentation

|  |  |
| --- | --- |
| Creation date | 2020-09-08 |
| Version | 2020-08 |
| Author | Ellen Webborn |
| Project | Smart Energy Research Lab (SERL) |
| Organisation | University College London (UCL) |
|  |  |

# Introduction

This document describes the England and Wales Energy Performance Certificate (EPC) data collected for SERL participants, stored in the file *SERL\_epc\_data\_v2020\_08.csv*. The data contains 75 columns and 902 rows (one row per participant with available EPC data). This document lists the EPC variables available along with basic information about the values for each variable such as number of unique values and statistics for numerical variables. A guide to the variables is available [here](https://epc.opendatacommunities.org/docs/guidance#glossary_domestic).

A few variables have been added to the EPC data since the data were collected (largely in October 2019, a few individual households had data retrieved later), and these will be made available in future SERL data releases. Data were collected with the Domestic Energy Performance Certificates API using the house number and postcode (details [here](https://epc.opendatacommunities.org/docs/api/domestic)).

The data have not been modified from the original source except for the removal of address data (replaced with our PUPRN (a unique identifier) used in the other datasets).

# Data summary

Table 1 lists all variables currently available in the SERL EPC dataset. The number of unique values is given, alongside the R data class and an example value from the dataset.

Table 1: All EPC variables, the number of unique values found for each variable, the variable (R) class, and an example from the dataset.

| *variable* | *n unique values* | *class* | *example* |
| --- | --- | --- | --- |
| PUPRN | 902 | character | GZ9RJ1P1 |
| current\_energy\_rating | 7 | character | C |
| potential\_energy\_rating | 7 | character | A |
| current\_energy\_efficiency | 78 | integer | 83 |
| potential\_energy\_efficiency | 69 | integer | 98 |
| property\_type | 5 | character | Flat |
| built\_form | 7 | character | Detached |
| inspection\_date | 775 | character | 2017-10-23 |
| local\_authority | 257 | character | E07000103 |
| constituency | 361 | character | E14001060 |
| lodgement\_date | 781 | character | 2013-10-03 |
| transaction\_type | 13 | character | marketed sale |
| environment\_impact\_current | 80 | integer | 84 |
| environment\_impact\_potential | 73 | integer | 94 |
| energy\_consumption\_current | 340 | integer | 190 |
| energy\_consumption\_potential | 292 | integer | 79 |
| co2\_emissions\_current | 107 | numeric | 3.8 |
| co2\_emiss\_curr\_per\_floor\_area | 129 | numeric | 31 |
| co2\_emissions\_potential | 92 | numeric | 1 |
| lighting\_cost\_current | 156 | integer | 81 |
| lighting\_cost\_potential | 101 | integer | 71 |
| heating\_cost\_current | 642 | integer | 685 |
| heating\_cost\_potential | 559 | integer | 886 |
| hot\_water\_cost\_current | 224 | integer | 164 |
| hot\_water\_cost\_potential | 158 | integer | 84 |
| total\_floor\_area | 422 | numeric | 78 |
| energy\_tariff | 7 | character | Single |
| mains\_gas\_flag | 3 | character | N |
| floor\_level | 13 | character | Ground |
| flat\_top\_storey | 3 | character | Y |
| flat\_storey\_count | 6 | integer | 2 |
| main\_heating\_controls | 28 | integer | 2107 |
| multi\_glaze\_proportion | 39 | integer | 100 |
| glazed\_type | 9 | character | double glazing installed during or after 2002 |
| glazed\_area | 5 | character | Normal |
| extension\_count | 6 | integer | 1 |
| number\_habitable\_rooms | 13 | integer | 5 |
| number\_heated\_rooms | 13 | integer | 8 |
| low\_energy\_lighting | 91 | integer | 71 |
| number\_open\_fireplaces | 7 | integer | 0 |
| hotwater\_description | 16 | character | Electric immersion, off-peak |
| hot\_water\_energy\_eff | 6 | character | Good |
| hot\_water\_env\_eff | 6 | character | Good |
| floor\_description | 34 | character | Suspended, no insulation (assumed) |
| floor\_energy\_eff | 5 | character | Very Good |
| floor\_env\_eff | 4 | character | Good |
| windows\_description | 19 | character | Fully double glazed |
| windows\_energy\_eff | 6 | character | Average |
| windows\_env\_eff | 6 | character | Average |
| walls\_description | 61 | character | Cavity wall, as built, no insulation (assumed) |
| walls\_energy\_eff | 6 | character | Very Poor |
| walls\_env\_eff | 6 | character | Average |
| secondheat\_description | 13 | character | Room heaters, electric |
| sheating\_energy\_eff | 1 | character | N/A |
| sheating\_env\_eff | 1 | character | N/A |
| roof\_description | 62 | character | Pitched, insulated at rafters |
| roof\_energy\_eff | 6 | character | Good |
| roof\_env\_eff | 6 | character | Average |
| mainheat\_description | 25 | character | Boiler and radiators, mains gas |
| mainheat\_energy\_eff | 6 | character | Good |
| mainheat\_env\_eff | 6 | character | Average |
| mainheatcont\_description | 26 | character | Manual charge control |
| mainheatc\_energy\_eff | 6 | character | Good |
| mainheatc\_env\_eff | 6 | character | Average |
| lighting\_description | 91 | character | Low energy lighting in 57% of fixed outlets |
| lighting\_energy\_eff | 7 | character | Very Poor |
| lighting\_env\_eff | 6 | character | Average |
| main\_fuel | 17 | character | oil (not community) |
| wind\_turbine\_count | 4 | integer | 0 |
| heat\_loss\_corridoor | 4 | character | unheated corridor |
| unheated\_corridor\_length | 60 | numeric | 4.998 |
| floor\_height | 76 | numeric | 2.400 |
| photo\_supply | 6 | integer | 0 |
| solar\_water\_heating\_flag | 3 | character | N |
| mechanical\_ventilation | 4 | character | natural |

For variables with fewer than 10 unique values in the EPC dataset, Table 2 shows the number of records with each value and the percent with this value (or non-value in the case of N/A or ‘NO DATA!’ etc.). We also include PUPRN to show the number of records.

Table 2: The number and percent of each value found in the dataset for each variable with fewer than 10 unique values found.

| *variable* | *value* | *number* | *percent* |
| --- | --- | --- | --- |
| PUPRN | - | 902 | 100.00 |
| current\_energy\_rating | A | 2 | 0.22 |
| B | 73 | 8.09 |
| C | 239 | 26.50 |
| D | 413 | 45.79 |
| E | 130 | 14.41 |
| F | 34 | 3.77 |
| G | 11 | 1.22 |
| potential\_energy\_rating | A | 26 | 2.88 |
| B | 416 | 46.12 |
| C | 339 | 37.58 |
| D | 86 | 9.53 |
| E | 28 | 3.10 |
| F | 6 | 0.67 |
| G | 1 | 0.11 |
| property\_type | Bungalow | 150 | 16.63 |
| Flat | 164 | 18.18 |
| House | 573 | 63.53 |
| Maisonette | 13 | 1.44 |
| Park home | 2 | 0.22 |
| built\_form | Detached | 307 | 34.04 |
| Enclosed End-Terrace | 11 | 1.22 |
| Enclosed Mid-Terrace | 6 | 0.67 |
| End-Terrace | 107 | 11.86 |
| Mid-Terrace | 202 | 22.39 |
| NO DATA! | 19 | 2.11 |
| Semi-Detached | 250 | 27.72 |
| energy\_tariff | NO DATA! | 1 | 0.11 |
| Single | 679 | 75.28 |
| Unknown | 81 | 8.98 |
| dual | 78 | 8.65 |
| dual (24 hour) | 1 | 0.11 |
| off-peak 7 hour | 1 | 0.11 |
| standard tariff | 61 | 6.76 |
| mains\_gas\_flag |  | 63 | 6.98 |
| N | 114 | 12.64 |
| Y | 725 | 80.38 |
| flat\_top\_storey |  | 752 | 83.37 |
| N | 92 | 10.20 |
| Y | 58 | 6.43 |
| flat\_storey\_count |  | 864 | 95.79 |
| 2 | 16 | 1.77 |
| 3 | 12 | 1.33 |
| 4 | 5 | 0.55 |
| 5 | 4 | 0.44 |
| 8 | 1 | 0.11 |
| glazed\_type | INVALID! | 2 | 0.22 |
| NO DATA! | 63 | 6.98 |
| double glazing installed before 2002 | 304 | 33.70 |
| double glazing installed during or after 2002 | 270 | 29.93 |
| double glazing, unknown install date | 201 | 22.28 |
| not defined | 41 | 4.55 |
| secondary glazing | 11 | 1.22 |
| single glazing | 5 | 0.55 |
| triple glazing | 5 | 0.55 |
| glazed\_area | More Than Typical | 17 | 1.88 |
| Much Less Than Typical | 1 | 0.11 |
| Much More Than Typical | 8 | 0.89 |
| NO DATA! | 63 | 6.98 |
| Normal | 813 | 90.13 |
| extension\_count |  | 63 | 6.98 |
| 0 | 500 | 55.43 |
| 1 | 245 | 27.16 |
| 2 | 77 | 8.54 |
| 3 | 11 | 1.22 |
| 4 | 6 | 0.67 |
| number\_open\_fireplaces |  | 27 | 2.99 |
| 0 | 741 | 82.15 |
| 1 | 101 | 11.20 |
| 2 | 28 | 3.10 |
| 3 | 2 | 0.22 |
| 4 | 2 | 0.22 |
| 7 | 1 | 0.11 |
| hot\_water\_energy\_eff | Average | 151 | 16.74 |
| Good | 580 | 64.30 |
| N/A | 5 | 0.55 |
| Poor | 62 | 6.87 |
| Very Good | 64 | 7.10 |
| Very Poor | 40 | 4.43 |
| hot\_water\_env\_eff | Average | 133 | 14.75 |
| Good | 601 | 66.63 |
| N/A | 5 | 0.55 |
| Poor | 75 | 8.31 |
| Very Good | 64 | 7.10 |
| Very Poor | 24 | 2.66 |
|  |  |  |  |
| floor\_energy\_eff | Average | 1 | 0.11 |
| Good | 9 | 1.00 |
| N/A | 519 | 57.54 |
| NO DATA! | 342 | 37.92 |
| Very Good | 31 | 3.44 |
| floor\_env\_eff | Average | 1 | 0.11 |
| Good | 9 | 1.00 |
| N/A | 861 | 95.45 |
| Very Good | 31 | 3.44 |
| windows\_energy\_eff | Average | 475 | 52.66 |
| Good | 276 | 30.60 |
| N/A | 3 | 0.33 |
| Poor | 55 | 6.10 |
| Very Good | 46 | 5.10 |
| Very Poor | 47 | 5.21 |
| windows\_env\_eff | Average | 475 | 52.66 |
| Good | 276 | 30.60 |
| N/A | 3 | 0.33 |
| Poor | 55 | 6.10 |
| Very Good | 46 | 5.10 |
| Very Poor | 47 | 5.21 |
| walls\_energy\_eff | Average | 62 | 6.87 |
| Good | 451 | 50.00 |
| N/A | 3 | 0.33 |
| Poor | 131 | 14.52 |
| Very Good | 52 | 5.76 |
| Very Poor | 203 | 22.51 |
| walls\_env\_eff | Average | 62 | 6.87 |
| Good | 451 | 50.00 |
| N/A | 3 | 0.33 |
| Poor | 131 | 14.52 |
| Very Good | 52 | 5.76 |
| Very Poor | 203 | 22.51 |
| sheating\_energy\_eff | N/A | 902 | 100.00 |
| sheating\_env\_eff | N/A | 902 | 100.00 |
|  |  |  |  |
| roof\_energy\_eff | Average | 169 | 18.74 |
| Good | 376 | 41.69 |
| N/A | 106 | 11.75 |
| Poor | 47 | 5.21 |
| Very Good | 88 | 9.76 |
| Very Poor | 116 | 12.86 |
| roof\_env\_eff | Average | 169 | 18.74 |
| Good | 376 | 41.69 |
| N/A | 106 | 11.75 |
| Poor | 47 | 5.21 |
| Very Good | 88 | 9.76 |
| Very Poor | 116 | 12.86 |
| mainheat\_energy\_eff | Average | 98 | 10.86 |
| Good | 708 | 78.49 |
| N/A | 5 | 0.55 |
| Poor | 25 | 2.77 |
| Very Good | 45 | 4.99 |
| Very Poor | 21 | 2.33 |
| mainheat\_env\_eff | Average | 57 | 6.32 |
| Good | 725 | 80.38 |
| N/A | 5 | 0.55 |
| Poor | 18 | 2.00 |
| Very Good | 59 | 6.54 |
| Very Poor | 38 | 4.21 |
| mainheatc\_energy\_eff | Average | 292 | 32.37 |
| Good | 479 | 53.10 |
| N/A | 5 | 0.55 |
| Poor | 52 | 5.76 |
| Very Good | 29 | 3.22 |
| Very Poor | 45 | 4.99 |
| mainheatc\_env\_eff | Average | 292 | 32.37 |
| Good | 479 | 53.10 |
| N/A | 5 | 0.55 |
| Poor | 52 | 5.76 |
| Very Good | 29 | 3.22 |
| Very Poor | 45 | 4.99 |
|  |  |  |  |
| lighting\_energy\_eff |  | 1 | 0.11 |
| Average | 178 | 19.73 |
| Good | 178 | 19.73 |
| N/A | 3 | 0.33 |
| Poor | 101 | 11.20 |
| Very Good | 301 | 33.37 |
| Very Poor | 140 | 15.52 |
| lighting\_env\_eff | Average | 178 | 19.73 |
| Good | 178 | 19.73 |
| N/A | 3 | 0.33 |
| Poor | 101 | 11.20 |
| Very Good | 301 | 33.37 |
| Very Poor | 141 | 15.63 |
| wind\_turbine\_count |  | 36 | 3.99 |
| -1 | 2 | 0.22 |
| 0 | 862 | 95.57 |
| 1 | 2 | 0.22 |
| heat\_loss\_corridoor | NO DATA! | 752 | 83.37 |
| heated corridor | 30 | 3.33 |
| no corridor | 55 | 6.10 |
| unheated corridor | 65 | 7.21 |
| photo\_supply |  | 420 | 46.56 |
| 0 | 477 | 52.88 |
| 20 | 1 | 0.11 |
| 35 | 1 | 0.11 |
| 40 | 2 | 0.22 |
| 50 | 1 | 0.11 |
| solar\_water\_heating\_flag |  | 401 | 44.46 |
| N | 496 | 54.99 |
| Y | 5 | 0.55 |
| mechanical\_ventilation | NO DATA! | 63 | 6.98 |
| mechanical, extract only | 2 | 0.22 |
| mechanical, supply and extract | 3 | 0.33 |
| natural | 834 | 92.46 |

Table 3 provides basic summary statistics for numeric variables. The column ‘n’ shows the number of values that were possible to include in the statistics (N/A and similar responses are excluded).

Table 3: Basic statistics for integer and numeric variables. ‘n’ is the number of values used in the calculations (i.e. the non-NA values).

| *variable* | *n* | *min* | *max* | *mean* | *standard deviation* |
| --- | --- | --- | --- | --- | --- |
| current\_energy\_efficiency | 902 | 1.0 | 185.00 | 63.15 | 14.01 |
| potential\_energy\_efficiency | 902 | 19.0 | 197.00 | 78.19 | 11.25 |
| environment\_impact\_current | 902 | 4.0 | 220.00 | 60.45 | 15.67 |
| environment\_impact\_potential | 902 | 20.0 | 232.00 | 76.31 | 12.92 |
| energy\_consumption\_current | 902 | -754.0 | 1226.00 | 245.54 | 118.89 |
| energy\_consumption\_potential | 902 | -827.0 | 1030.00 | 139.79 | 97.23 |
| co2\_emissions\_current | 902 | -12.9 | 36.00 | 4.35 | 3.00 |
| co2\_emiss\_curr\_per\_floor\_area | 902 | -147.0 | 188.00 | 44.51 | 21.19 |
| co2\_emissions\_potential | 902 | -14.1 | 28.00 | 2.50 | 2.18 |
| lighting\_cost\_current | 902 | 18.0 | 419.00 | 83.23 | 36.81 |
| lighting\_cost\_potential | 902 | 13.0 | 226.00 | 56.92 | 20.79 |
| heating\_cost\_current | 902 | -2233.0 | 5989.00 | 727.61 | 520.70 |
| heating\_cost\_potential | 902 | -2224.0 | 4503.00 | 544.51 | 357.14 |
| hot\_water\_cost\_current | 902 | 0.0 | 672.00 | 140.68 | 67.30 |
| hot\_water\_cost\_potential | 902 | 0.0 | 351.00 | 95.06 | 38.31 |
| total\_floor\_area | 902 | 0.0 | 371.00 | 97.72 | 46.96 |
| flat\_storey\_count | 38 | 2.0 | 8.00 | 3.05 | 1.29 |
| main\_heating\_controls | 841 | 2101.0 | 2706.00 | 2137.28 | 104.27 |
| multi\_glaze\_proportion | 830 | 0.0 | 100.00 | 91.03 | 24.92 |
| extension\_count | 839 | 0.0 | 4.00 | 0.54 | 0.77 |
| number\_habitable\_rooms | 839 | 1.0 | 13.00 | 4.76 | 1.76 |
| number\_heated\_rooms | 839 | 0.0 | 11.00 | 4.67 | 1.78 |
| low\_energy\_lighting | 872 | 0.0 | 100.00 | 47.67 | 33.91 |
| number\_open\_fireplaces | 875 | 0.0 | 7.00 | 0.20 | 0.56 |
| wind\_turbine\_count | 866 | -1.0 | 1.00 | 0.00 | 0.07 |
| unheated\_corridor\_length | 66 | 0.0 | 19.57 | 6.11 | 3.28 |
| floor\_height | 304 | 2.0 | 3.50 | 2.44 | 0.18 |
| photo\_supply | 482 | 0.0 | 50.00 | 0.38 | 3.88 |