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.

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).

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

variable	*n unique values*	*class*	*example*
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

variable	n unique values	class	example
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

variable	value	number	percent
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

variable	value	number	percent
	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

variable	value	number	percent
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

variable	value	number	percent
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

variable	value	number	percent
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