Department of Energy & Climate Change

National Energy Efficiency Data-Framework: Making data available

Consultation response

May 2014

Any enquiries regarding this publication should be sent to us at
EnergyEfficiency.Stats@decc.gsi.gov.uk.

This publication is available from our website at www.gov.uk/decc.

Contents

1. Introduction

Background

Information held by DECC as part of the National Energy Efficiency Data-Framework (NEED) is a potentially valuable resource for researchers looking at energy efficiency and energy consumption in households.

The UK helped secure the G8's Open Data Charter, which establishes the presumption that the data held by governments will be publicly available, unless there is good reason to withhold it. As part of this commitment to Open Data, DECC is proposing to publish an anonymised dataset[1] of data from NEED.

NEED was set up by DECC to provide a better understanding of energy use and energy efficiency in domestic and non-domestic buildings in Great Britain. The data framework matches – at individual property level – gas and electricity consumption data with information on energy efficiency measures installed in homes. It also includes data about property attributes and household characteristics.

The consultation published on the 21 November 2013 proposed publication of two datasets:

1) **Public use (or training) dataset**: Approximately 20,000 records including information on energy consumption, energy efficiency measures installed in properties and property attributes. This dataset would be made available to all.
2) **End user licence dataset**: Approximately four million records including more variables than the public use dataset. It would be published in a slightly more restricted format; all individuals would be required to agree to an end user licence before having access to the data.

The consultation proposed that the two datasets would be samples of domestic properties in England and Wales. Data would be anonymised to prevent any individual household or business being identified. The data would be published in a format that could not be used for targeting specific households. It was envisaged the data would primarily be used by researchers looking at how energy is used in households, including the impact of installing energy efficiency measures.

DECC used feedback from NEED users to inform the proposals set out in the consultation, including feedback received from a seminar with energy suppliers and an event held for NEED users.

The consultation sought views on these proposals, including the content of the dataset and approach to anonymisation and publication. The consultation document can be found here: https://www.gov.uk/government/consultations/national-energy-efficiency-data-framework-making-data-available.

General response

DECC received 15 written responses to the consultation from a range of respondents. The table below summarizes the respondents.

[1] Anonymised data are data relating to a specific individual or property where the identifiers have been removed to prevent identification of that individual or property (directly or indirectly).

Type of respondent	Number of responses
Energy industry organisations	4
Academic institutions or individuals	3
Local authorities	2
Representative bodies	2
Non-Govermental Organisation (NGO)	1
Open Data User Group	1
Information Commissioner's Office	1
Private sector researchers	1

DECC wishes to thank respondents for their input and the time and effort required to provide these responses.

2. Responses

All responses to the consultation were supportive of more data from NEED being published. The majority of respondents were broadly in agreement with the proposals set out subject to small variations. Disagreement was almost always in the form of a request to publish more detailed data or make the data more widely available. The rest of this section provides a summary of the responses to each of the consultation questions and the Government's response.

Question 1: Do you agree DECC should release anonymised NEED data?

Fourteen of the fifteen respondents to the consultation answered this question. All responses were positive with two highlighting the need to comply with the Data Protection Act and engage with the Information Commissioner's Office (ICO).

Two respondents stated the dataset would be "invaluable" and views were also expressed about how important robust data is to help improve the country's housing stock and change attitudes towards energy efficiency.

While supportive of the proposals in the consultation, some respondents expressed disappointment that the proposals did not go further. For example, allowing targeting of measures or wider access to the data.

Government response

Government plans to proceed with the proposal to publish anonymised data from NEED. We will continue to engage with the Information Commissioner's Office and ensure that the approach to anonymisation and publication is in line with the requirements of the Data Protection Act 1998.

Question 2: Do you agree with the proposed approach to publishing two separate datasets for different purposes?

There were eleven responses to this question. Eight of these responses were supportive. There was an understanding that this approach would allow more data to be made available and that the different data sources could serve different purposes. The ICO agreed with the approach, "as it allows the measures taken to protect individuals' privacy to be tailored to each dataset, bearing in mind the purpose for which each dataset is released, who is likely to use them and the different levels of risk to individuals' privacy".

Two respondents disagreed with the proposals. One raised concerns that the small dataset (20,000 records) could lead to potential bias and incorrect inferences as a result of small numbers of some combinations of attributes, the other wanted to see all the data made available to all users.

One respondent had no view.

Responses to this question also outlined a number of other issues for consideration:

- One respondent suggested discussion with the Open Data Institute to gain more insight into what form the public use dataset should take to ensure greatest value from this dataset.

- There were concerns that the larger (four million record) dataset would have limited value if published under a restricted licence and a request for DECC to re-examine the proposed licence to ensure that it is fair to users of all sizes and encourages wide use of the NEED data. For example, whether "a single database at postcode or another suitably anonymous geography in order to simplify the process, reduce cost and deliver a fair level of access to all potential users (thus maximising the potential economic and social benefit). This dataset would best be provided under the Open Government Licence[2] and would, ideally, contain full national data."
- A respondent asked for clarity on what would constitute "commercial organisations which assist in the delivery of Government policies … specifically, what constitutes policy and how broadly the description 'assist in the delivery' would be applied".
- There was a request for an additional even more restricted dataset to be published under a special licence in line with other data available via the UK Data Archive. This would have the potential to allow more detailed data to be included in the dataset.

Government response

Following the strong support for the proposals set out in the document DECC intends to go ahead with the publication of two separate datasets; a public use dataset and an end user licence dataset. In both cases, the trade-off between risk of disclosure and utility is the primary factor determining the format of the final datasets.

DECC will publish a 50,000 record dataset as Open Data. This is more than twice the size of the dataset proposed in the consultation. This is being done to create a more useful dataset while retaining the high level of confidence in anonymisation which is possible for a smaller dataset. The sample will be selected to be representative of the England and Wales housing stock. It will have the same structure and format as the end user licence dataset with the exception of the exclusion of some variables; it will contain over 30 variables.

DECC will publish a dataset of approximately four million records via the UK Data Archive[3] under an end user licence. Publication of the larger dataset under an end user licence is in line with Government best practice for detailed statistical data. The decision to publish this dataset under an end user licence has been taken following input and advice from the Information Commissioner's Office and the UK Anonymisation Network. It provides the best approach to protecting against disclosure while retaining utility of the dataset. If this dataset were made publically available the utility of the dataset would have to be reduced to a point where it would not be possible to carry out the majority of analysis users have expressed a desire to undertake.

The UK Data Archive description of how the data may be used as: *Any individual employed by, or undertaking research for, any organisation, may use data even if this entails monetary reward, where a public good results from the use. Public good can*

[2] http://www.nationalarchives.gov.uk/doc/open-government-licence/version/2/

[3] http://ukdataservice.ac.uk/get-data/how-to-access/conditions.aspx

be defined as an activity which widens access to information sourced from our collection and has social or economic benefit.[4]

The dataset may be used to aide more efficient delivery of Government schemes; subject to individuals and organisations complying with the conditions set out in the end user licence.

DECC will continue to look at ways to make the data more accessible to a range of users. This will include further work with data owners and experts in anonymisation to see what more could be published as Open Data and what potential there is to publish a more detailed dataset in a secure environment[5]. DECC will also continue to review the outputs published on the DECC website, for example, whether postcode level average consumption can be published in future. More details of future plans are available in Section 3, Next Steps, including a commitment to publish an initial report on future plans for making more data available by 30 September 2014.

Question 3: In relation to i) the public use dataset and ii) the end user licence dataset, what are your priorities for variables in the dataset?

Most respondents answered this question in relation to the end user licence dataset, with a small number responding in relation to the public use dataset. Some respondents did not provide preferences for specific variables but expressed a desire to see as much data made available as possible. Where detailed responses were provided priorities varied between different potential users, but in all cases users wanted additional variables considered as important or priority. There was only one case where a response included details of some variables that were considered less important and no respondents wanted a variable excluded from the dataset.

a) Do you agree with the priority variables set out in Table 4.1? If not, which of the variables listed do you consider to be priorities?

There were nine responses to this question. Two respondents agreed with the priorities set out in the consultation document. The majority stated some preference for additional variables to be considered priority, in summary:

- All variables except environmental impact band were highlighted by at least one respondent.
- There were three requests for main heating fuel to be a priority variable; two which specified this as the only additional priority variable and one which included this in a small number of additional priority variables.
- There were four requests for access to mains gas to be included in the dataset; in all four cases this variable was one of a number of additional variables requested.
- One respondent requested an additional 15 variables to be considered as priority (all but environmental impact band, main heating fuel, loft insulation and weighting).
- Three respondents considered detailed geography or Local Authority as an additional priority.

[4] The UK Data Archive definition of non-commercial use: http://ukdataservice.ac.uk/get-data/how-to-access/registration/commercialusers.aspx.

[5] For example: http://ukdataservice.ac.uk/get-data/how-to-access/conditions/controlled-data.aspx.

- There was also one request for measures installed (cavity wall insulation, loft insulation, solid wall insulation and boiler) to be considered as a priority along with one response stating these variables were less important and could be dropped if necessary.
- There was one request for consideration of the inclusion of income in the end user licence dataset.

Government response

As a result of the responses received, two additional variables have been prioritised for both the end user licence and public use datasets; main heating fuel and access to mains gas. The later will be incorporated into the gas consumption variable using information from the Energy Performance Certificate data to supplement the meter point data.

Region variable will be considered a priority for the end user licence dataset. However Local Authority will not be included, despite a strong desire for this information (see response to 3e).

Environmental impact band will not be included in either dataset, as it was the only variable which no respondent highlighted as a priority.

Table 2 shows the prioritisation of variables based on responses to the consultation, including rationale for decisions.

b) Do you agree with the variables assigned as important in Table 4.1? If not, which of the other variables listed do you consider to be important?

There were nine responses to this question. In a number of cases respondents referred to their answer to question 3a. There was general agreement with the proposals, but as seen in the response to a), most respondents wanted a number of additional variables to be classified as important. In summary:

- Four respondents wanted Region to be considered important and three respondents expressed a desire for local authority to be considered important.
- Three respondents wanted to see loft insulation thickness and wall construction as important.
- Two respondents wanted weighting to be categorised as important.
- One organisation felt the data related to the Green Deal had been over prioritised and stated that “in order to get the most value from NEED, we would encourage DECC to incorporate fields focussing on location, fuel poverty, environmental impact and types of fuel available” to increase potential for a wider range of users.

Government response

Loft insulation thickness and wall construction have been considered important (rather than “under consideration”).

Table 2 shows the prioritisation of variables based on responses to the consultation, including rationale for decisions.

c) Do you agree that those variables listed as "under consideration" are less important than the variables listed as priority or important?

There were eight responses to this question. All responses reiterated the information provided in reply to parts a and b, deeming a number of additional variables be upgraded from under consideration to important or priority.

Government response

Table 2 shows the prioritisation of variables based on responses to the consultation, including rationale for decisions.

d) Are there any variables included in the proposals which you think should not be included?

There were nine responses to this question. Five of these responses were "no", with no further elaboration. The other four responses can be summarised as follows:

- Two respondents wanted additional variables included; household characteristics and small-scale renewables.
- One organisation restated its view that the more data fields that could be included, the more valuable the dataset would be to a wider audience.
- One organisation stated that there were no variables that should be excluded in principle, but if the scope of the end user licence dataset were very broad then it would have concerns over the inclusion of some variables (it did not state which).

Government response

No privacy concerns were raised about specific variables under the proposed approach to publication of the datasets. Therefore decisions on which variables to include will be made on the basis of their use for analysis, while continuing to ensure that the datasets are anonymised.

DECC has also engaged in further discussion with the respondent which raised concerns about inclusion of some variables if the scope of the end user licence dataset were very broad. The approach to anonymisation and planned testing along with the restrictions of the end user licence for the larger dataset have addressed those concerns.

Table 2 shows the prioritisation of variables based on responses to the consultation, including rationale for decisions.

e) Do you agree that inclusion of a lower level geography identifier is less important than a wider range of variables?

There were eight responses to this question. Views were split:

- Three respondents agreed that lower level geography is less important than the range of variables.

- Two respondents disagreed and stated a preference for the lower level of geography even if this had an impact on the range of variables.
- Three respondents felt both were equally important. In two cases there was a request for information at lower level e.g. “label each field with the relevant geographic level” or “consider a simple ‘X% of treatable properties have been completed’”. One respondent argued that making data available under a special licence or secure conditions might be the solution.

Government response

Following consideration of the response to this question DECC intends to prioritise the range of variables over the more detailed geographic information. The need to ensure anonymisation means the utility of the dataset would become so limited that it would not provide much benefit beyond data already published in aggregate form by DECC (e.g. typical consumption by number of bedrooms at local authority level) if the data were published with a local authority identifier in the dataset.

DECC is also reviewing the possibility of publishing more detailed consumption data as part of the suite of sub-national consumption outputs. Data are currently available for gas and electricity consumption for domestic properties by lower level super output area, including total consumption, number of meters and average consumption. DECC is considering publication of these data at postcode level in future. There is an opportunity to input into these and other proposals for sub-national consumption data in response to the latest sub-national publication[6].

Table 2 shows the prioritisation of variables based on responses to the consultation, including rationale for decisions.

f) Which lower layer super output area (LSOA) data are most useful? Index of multiple deprivation, output area classification or percentage of households in fuel poverty?

There were nine responses to this question. Respondents had different priorities, with preferences shown in the table below:

Table 1: Lower layer super output area variable preferences

	First choice	Second choice	Third Choice
Index of Multiple Deprivation	3	2	
Fuel Poverty per cent	4		
Output Area Classification	2		1

Government response

DECC is intending to include Index of Multiple Deprivation (IMD) in the public use file and both Index of Multiple Deprivation and percentage of properties in fuel poverty in the end user licence dataset.

[6] https://www.gov.uk/government/publications/msoa-igz-and-lsoa-factsheet

Both of these variables are based on modelled data and will be assigned to each property based on the geographic location (LSOA) of the property. Following initial testing for the anoymisation of the dataset it has been decided that just one of the two variables should be included in the public use file. Index of Multiple Deprivation has been prioritised due to the stability of the variable and the fact that it was rated as useful by more users (first or second choice).

Given the limited support for Output Area Classification and additional risk of disclosure if included, this variable will not be included in either of the published datasets.

Table 2 shows the prioritisation of variables based on responses to the consultation, including rationale for decisions.

g) Would a weighting variable be useful?

Responses to this question ranged from “crucial” to no. Of the nine responses received, six responses were “crucial” or “yes”. Two respondents considered it “may be useful” or “of value” and just one respondent did not consider a weighting variable to be useful.

Government response

A weighting variable will be included with the final published dataset. This will be based on property attribute data held by the Valuation Office Agency (covering property type, floor area band and property age[7]) and Country/Region.

Table 2 shows the prioritisation of variables based on responses to the consultation, including rationale for decisions.

Question 4: Proposed bandings for variables in the dataset are set out in Annex B. Do you agree with these proposals in relation to i) the public use dataset and ii) the end user licence dataset? Please bear in mind that greater granularity of data will reduce the number of variables that can be included in the final dataset.

a) Annex B sets out options for banding variables please let us know which you would prefer for each variable of interest to you.

Responses to this question are summarised in Table 2 along with decisions on the banding that will be applied to the dataset.

b) Are there any variables that can be banded further than proposed without significant loss of utility?

Five respondents answered, “no”. One had no comment and the other respondent reiterated the desire that Energy Performance Certificate (EPC) data should not be grouped. One of the respondents also restated that it would be preferred if consumption were not banded, and that the proposed bandings were adequate to retain the anonymity of the data.

c) Are there any variables which would no long be useable for analysis if the proposed banding – or one of the proposed options - is applied?

[7] The weighting will be based on the Valuation Office Agency council tax property attributes data. For property type and floor area the bandings used for weighting will be the same as those used in the final NEED dataset. For property age only three property age bands can be used in the weighting (pre-1930, 1930-1982, 1983 or later) rather than the six planned for use in the dataset. This reflects the overlap in categories for EPC and VOA data.

Of the seven responses to this question, none suggested any variables which would no longer be useable. However, two respondents took the opportunity to restate their preferences:

- EPC banding is useful but should also be made available without banding.
- It would be preferable for energy consumption to be provided as an actual number rather than a banded/rounded figure. The respondent highlighted the difficulties with attempting to detect small changes in energy use and understand the causes of those changes and the increase in the level of uncertainty in any findings.

d) For variables such as consumption and floor area, is it preferable to have bands of the same size (which may have to be larger) or more detail in the centre of the distribution with larger bands at the extremes?

There were a range of responses to this question. In summary, the six respondents' views were:

- Two preferred banding of the same size as they are easier to work with and provide clarity when presenting results.
- One respondent wanted more detail in the centre of the distribution with larger bands at the extremes.
- Two respondents had no strong opinion.
- One respondent set out the benefits of each approach and requested consistency with other similar variables in frequently used data sources such as the EHS and SAP/RdSAP.

More detailed responses relating to banding for specific variables have been summarised in Table 2.

Government response

The Government's responses to this question are summarised in Table 2, which also shows the prioritisation of variables based on responses to question 3 and the rationale for decisions made. Dummy datasets to illustrate how the datasets will look are provided as an Excel spreadsheet at Annex D.

Table 2: Summary of data to be included in anonymised dataset, including banding

Variable	Banding options proposed in consultation	Banding preferences	Proposed banding	New priority (previous priority in brackets)	Rationale
Gas Consumption	a) Gas consumption deciles. b) Equal bands size at all levels/rounding e.g. to nearest 5,000kWh. c) Variable size bands e.g. 100-2,500 2,500 - 5,000 … 1,000 kWh bands... 18,000 - 20,000 20,000 - 22,500 22,500 - 25,000 25,000 - 30,000 30,000 - 40,000 40,000 - 50,000 kWh d) More detail than above (e.g. rounded to nearest 100kWh) at the possible detriment of other variables in the dataset.	• Option a - one respondent • Option b - none • Option c - two respondents • Option d - three respondents • One respondent preferred no banding, but content with c if necessary.	A combination of c) and d). Rounding data with more detail than set out in option c) of proposals: • off gas: coded 1 • 0 - 99kWh:coded 99 • 100-7,999kWh: rounded to nearest 500kWh • 8,000-15,999kWh: rounded to nearest 100kWh • 16,000-24,999kWh: rounded to nearest 500kWh • 25,000-34,999kWh: rounded to nearest 1,000 kWh • 35,000-50,000kWh: rounded to nearest 5,000 kWh • All values greater than 50,000kWh: included as 50,000 kWh	Top Priority (Priority)	This is a key variable and in order to carry out robust analysis as much detail as possible is required. However, this is also the most sensitive variable in the dataset and therefore must be protected. The level of rounding applied ensures there are no unique consumption values (analysis based on 2011 consumption) when comparing with a single visible variable, e.g. property size, age or floor area band, within a region.
Electricity consumption	a) Electricity consumption deciles b) Equal bands at all levels/rounding (e.g. 5,000 kWh). c) Variable band sizes e.g. 100-1,000 1,000 - 2,000 … 500 kWh bands... 6,000 - 7,000 7,000 - 8,000 8,000 - 10,000 10,000 - 15,000 15,000 - 25,000 kWh d) More detail than above (e.g. rounded to nearest 100kWh) to the possible detriment of other variables in the dataset.	As above: • Option a - one respondent • Option b - none • Option c - two respondents • Option d - three respondents • One respondent preferred no banding, but content with c if necessary.	A combination of c) and d). Rounding data with more detail than set out in option c) of proposals: • Invalid or less than 100kWh:coded 99 • 100-9,999kWh: rounded to nearest 50kWh • 10,000-11,999kWh: rounded to nearest 100kWh • 12,000-14,999kWh: rounded to nearest 500kWh • 15,000 -19,999kWh: rounded to nearest 1,000 kWh • 20,000 -25,000kWh: rounded to nearest 5,000 kWh • All values greater than 25,000kWh: included as 25,000 kWh	Top Priority (Priority)	This is a key variable and in order to carry out robust analysis as much detail as possible is required. However, this is also the most sensitive variable in the dataset and therefore must be protected. The level of rounding applied ensures there are no unique consumption values (analysis based on 2011 consumption) when comparing with a single visible variable, e.g. property size, age or floor area band, within a region.

Variable	Banding options proposed in consultation	Banding preferences	Proposed banding	New priority (previous priority in brackets)	Rationale
Economy 7 Flag	n/a	n/a	Flag indicates all properties with a profile 2 meter.	Under consideration	There was no strong case made for inclusion of this variable. However, as it is not a visible variable and relates to the meter not the tariff it has limited additional risk of disclosure and will therefore be included in the dataset.
Energy Efficiency Band	a) As per EPC but with two groupings: - A and B grouped - F and G grouped b) As per EPC (at detriment of other variables in dataset). c) more groupings allowing more detail/other variables.	• Option a - one respondent • Option b - three respondents (one happy with a as an alternative, and one happy for A and B to be grouped, but not F and G). • Option c - one respondent • One respondent preferred SAP score, but happy with banding if necessary.	As per EPC but with bands A and B grouped.	Top Priority (Priority)	This was highlighted as important by a number of responses and is therefore a top priority. Most respondents wanted as much detail as possible. However, only 0.04 per cent of records on the dataset the sample will be drawn from are band A. If published as a separate band this would quickly become disclosive when consider alongside information available in the public domain. Therefore bands A and B will be combined. Each of the other bands will be kept separate as per EPCs.
Environment Impact Band	a) As per EPC but with two groupings: - A and B grouped - F and G grouped b) As per EPC (at detriment of other variables in dataset). c) more groupings allowing more detail/other variables.	As above, except request for SAP score.	n/a	Drop (Under consideration)	No users highlighted this as a variable that should get increased priority. Therefore the benefits to users do not outweigh the additional risk of including this variable (given its visibility on EPC certificates), so it will not be included in the final dataset.

Variable	Banding options proposed in consultation	Banding preferences	Proposed banding	New priority (previous priority in brackets)	Rationale
Property Age	a) As per EPC: b) Fewer bands, e.g.: pre-1930 1930-1949 1950-1966 1967-1982 1983-1990 1991-1995 1996 onwards	• Option a -six respondents • Option b - one respondents • There was interest in the greater detail the EPC data provided at the extremes i.e. pre-1930 and post 1996.	pre-1930 1930-1949 1950-1966 1967-1982 1983-1995 1996 onwards	Top Priority (Priority)	Though respondents wanted more detail in this variable, especially for the oldest and newest properties, these categories could not be split due to the relatively small number of records and therefore higher risk of disclosure. More detail in these bands would have reduced the detail available for other variables, including the level of rounding required for consumption.
Property Type	Combination of built form and property type variables: 1) Detached house 2) Semi-detached house 3) End terrace house 4) Mid terrace house 5) Bungalow 6) Flat (inc. maisonette)	Four respondents agreed with the proposals, one respondent requested a further breakdown of bungalows (by detached and semi detached) and one respondent request a breakdown in line with the four part classification used in AddressBase.	As proposed.	Top Priority (Priority)	The majority of respondents agreed with the proposals. Bungalows are already the least common property type and splitting this further would restrict the information provided for other variables, therefore this will not be split.
Floor area band	a) 50m^2 bands and category for all over 200m^2 (i.e. same as NEED outputs) b) 25m^2 bands and over 200m^2	• Option a -three respondents • Option b - two respondents • One respondent considered both OK.	50m^2 bands and category for all over 150m^2.	Top Priority (Priority)	As more than half chose or were content with option a) (50m^2 bands), this banding will be used, as it allows for more variables or greater detail in other variables. However, unlike option a) the top category will be over 150m^2. This is being done in order to avoid additional grouping for other variables. The majority of unique records when considering combinations of variables in the data were for records in the over 200m^2 category. Joining these records with the group below allows greater utility of other variables to be retained.

Variable	Banding options proposed in consultation	Banding preferences	Proposed banding	New priority (previous priority in brackets)	Rationale
Main Heating Fuel	a) Gas, electricity, other. b) Gas, other	All seven responses to this question showed a preference for option a.	Option b) Gas, other	Priority (Important)	Responses were unanimous in wanting the more detailed breakdown for this variable. However, the small number of properties in “other” means the disclosure risk of including a split for the non-gas fuels means it is not possible to implement this option alongside the other priority data included in the dataset.
Mains Gas	a) Yes/no	Where responses were provided they agreed with the proposals.	Off gas properties coded 1 in gas consumption variable for each year - this will be based on a combination of the EPC variables and whether or not a property appears to have a gas meter in the meter point data.	Priority (Under consideration)	This variable was highlighted as one that should have increased priority by a number of users. In order to provide 100 per cent coverage, the off gas EPC variable will be combined with information on properties with a gas meter to provide a more comprehensive variable.
Loft insulation thickness	a) 50mm bands up to 250mm of loft insulation. b) loft insulation flag (yes/no based on 150mm or more)	• Option a -two respondents (with one of these stating option b would be better than nothing). • Option b - two respondents. • One respondent suggested more detailed bands (25mm) below 150 and then 150mm+ combined.	b) loft insulation flag (yes/no based on 150mm or more) - this will be based on a combination of the EPC variable and information on lofts insulated through Government schemes.	Important (PUF - no, EUL - under consideration)	There was some interest in the different levels of insulation below 150mm. However, due to disclosure, the relative importance of this variable and the gaps in the EPC dataset a flag for loft insulation will be included. In addition to the information provided on the EPC, it will also be assumed that any property recorded as having had insulation through a Government scheme has more than 150mm of insulation.
Wall construction	a) cavity wall or other b) cavity wall, solid wall, other	All responses provided stated a preference for option b).	Option a) cavity wall or other.	Important (Under consideration)	Due to the small number of "other" properties, option a) will be included in the published dataset.
Cavity wall insulation installed	n/a	n/a	Flag indicating properties which have had cavity wall insulation installed through a Government scheme.	Important (Important)	Variable to be included as planned.

Variable	Banding options proposed in consultation	Banding preferences	Proposed banding	New priority (previous priority in brackets)	Rationale
Cavity wall insulation year	a) Calendar year b) Gas year (1 October - 30 September)	There were only four responses to this question. Three preferences for a) and one for b) or installation quarter.	a) calendar year	Important (Important)	The possibility of including installation quarter was considered. However, this would quickly lead to a high risk of disclosure in cases where individuals are aware of when cavity wall insulation has been carried out on a property. Therefore this variable will contain information on calendar year of installation - as the option which was preferred by most respondents.
Loft insulation installed	n/a	n/a	Flag indicating properties which have had loft insulation installed through a Government scheme.	Important (Important)	Variable to be included as planned.
Loft insulation install year	a) Calendar year b) Gas year (1 October - 30 September)	Same as CWI: Four responses to this question. Three preferences for a) and one for b) or installation quarter.	a) calendar year	Important (Important)	The possibility of including installation quarter was considered. However, this would quickly lead to a high risk of disclosure in cases where individuals are aware of when loft insulation has been installed in a property. Therefore this variable will contain information on calendar year of installation - as the option which was preferred by most respondents.
Solid wall insulation installed	n/a	n/a	n/a	Drop (Important - if sufficient records to avoid disclosure)	Only 0.5 per cent of records on the dataset the sample will be selected from have a record of having had solid wall insulation installed. Given the high visibility of this measure and the small number of properties it applies to this variable will be excluded from the dataset. There was some interest from respondents in using this variable for analysis, but only one respondent requested it be upgraded from Important to Priority.

Variable	Banding options proposed in consultation	Banding preferences	Proposed banding	New priority (previous priority in brackets)	Rationale
Solid wall install year	a) Calendar year b) Gas year (1 October - 30 September)	Same as CWI: Four responses to this question. Three preferences for a) and one for b) or installation quarter.	n/a	Drop (Important - if sufficient records to avoid disclosure)	See above.
New boiler	n/a	n/a	Flag indicating properties which have had a new boiler installed.	Important (Important)	Variable to be included as planned. Note there is a gap in the available data for 2008-09 we are working to include data for this period in the published dataset if possible.
New boiler install year	a) Calendar year b) Gas year (1 October - 30 September)	Same as CWI: Four responses to this question. Three preferences for a) and one for b) or installation quarter.	a) calendar year	Important (Important)	The possibility of including installation quarter was considered. However, this would quickly lead to a high risk of disclosure in cases where individuals are aware of when a boiler has been installed in a property. Therefore this variable will contain information on calendar year of installation - as the option which was preferred by most respondents.
Region	n/a	n/a	Former Government Office Regions and Wales.	EUL - priority PUF - drop (Under consideration)	Geographic information was considered important by many respondents. Some respondents thought Local Authority would be more valuable and some felt Region would be sufficient (for example to understand weather). This variable will not be included for the public use file in order to reduce the risk of disclosure.

Variable	Banding options proposed in consultation	Banding preferences	Proposed banding	New priority (previous priority in brackets)	Rationale
Local Authority	n/a	n/a	n/a	Drop (PUF - no, EUL - under consideration)	Although LA was viewed as important by a number of respondents, it will not be included in the dataset. If LA were included, the damage to the dataset required to reduce the risk of disclosure would be so great that the utility of the dataset would be limited.
Index of multiple deprivation	a) Deciles	n/a	Quintiles	Priority (PUF - Under consideration EUL - Important)	This variable was highlighted as valuable to a range of respondents and therefore will be included in the dataset as quintiles, subject to disclosure checking.
Output Area Classification (OAC)	a) Seven Super Groups: - Blue collar communities - City living - Countryside - Prospering suburbs - Constrained by circumstances - Typical traits - Multicultural	There was a request that 2011 Census OACs should be used if possible.	n/a	Drop (Under consideration)	Of the three variables put forward (IMD, OAC and fuel poverty indicator), this variable received the least support and therefore will not be included in the dataset.
Fuel poverty indicator	a) Quantiles. E.g. lowest quantile would be allocated to all households which are in an LSOA with fewer than 7 per cent of households estimated to be in fuel poverty.	n/a	Quintiles	EUL - Important PUF - drop (Under consideration)	This variable was considered important by a number of respondents and it is intended that it will be included in the end user licence dataset. However, IMD will take priority in the public use file where this variable will not be included.
Weighting variable	n/a	n/a	n/a	Priority (Under consideration)	Only one respondent did not consider a weighting variable useful. As there is very limited additional disclosure risk resulting from inclusion of this variable it will be included in the final dataset.

Variable	Banding options proposed in consultation	Banding preferences	Proposed banding	New priority (previous priority in brackets)	Rationale
Date of EPC inspection	n/a	n/a	a) Two groups: - 2007-2009 - 2010 or later.	EUL - important PUF - no (not considered in consultation)	One respondent requested information on the date of the EPC inspection on the end user licence dataset. This would allow users to understand how recent the data for a specific household is and therefore how reliable it is likely to be. Year of EPC was considered. However, this was seen as a useful variable by intruders when attempting to identify properties during initial testing. In order to include an indication of timing, while ensuring other variables do not need to be damaged, two groups rather than more detailed year will be included.

Question 5: Do you agree with the proposed approach to anonymisation for

i. The public use dataset; and

ii. The end user licence dataset?

There were ten responses to this question, all of which were positive. All respondents except ICO responded "yes", with some making a small number of additional comments, such as:

- A suggestion that, if greater granularity or detail is required for specific studies the case for these should be set out separately and data accessed from the appropriate sources.
- Support for the use of the ICO anonymisation code[8].
- A request to revisit the inclusion of Valuation Office Agency data in light of the recent HMRC consultation on data sharing[9].
- The anonymisation techniques appear sound and to follow best practice.
- A request for transparency regarding the disclosure control processes applied.
- A recommendation that DECC consult with the ICO.
- A desire for greater data sharing for the purpose of improving obligation delivery.

The ICO welcomed DECC's consideration of the "Anonymisation: managing data protection risk"[10] code of practice in taking steps to minimise risk of individual households being identified. However, it also highlighted that this does not cover all circumstances and techniques and recommended seeking additional guidance from sources such as the UK Anonymisation Network.

ICO were also pleased that DECC had recognised that different types and levels of disclosure may require different steps to ensure that individuals' privacy is protected.

While ICO felt it could not provide a complete assurance that the proposed approach taken would be fully data protection compliant in all circumstances, it considered that "on the basis of the information available, it would appear that DECC's approach generally reflects the guidance in the "Anonymisation: managing data protection risk" code of practice".

Government response

The positive responses to this question support DECC's planned approach. DECC will use the anonymisation techniques set out in the proposal as far as required, balancing disclosure risk with utility and potential damage to the dataset.

Full details of the approach to anonymisation and the testing carried out will be published alongside the dataset. As part of the anonymisation and publication process DECC has sought input from a range of parties including the Information Commissioner's Office and the UK Anonymisation Network (including the Office for

[8] http://ico.org.uk/for_organisations/data_protection/topic_guides/anonymisation

[9] https://www.gov.uk/government/consultations/sharing-and-publishing-data-for-public-benefit

[10] http://ico.org.uk/for_organisations/data_protection/topic_guides/anonymisation

National Statistics, University of Southampton, University of Manchester and Open Data Institute). These discussions have reinforced the need to view anonymisation in the context of how the data will be made available, supporting the more cautious approach taken to protecting data in the public use file. DECC will continue to engage with these organisations prior to and following publication of the datasets.

DECC intends to include the same variables in both the public use and end user licence files, with the exception of Region, fuel poverty indicator and date of EPC inspection. The exclusion of these three variables will reduce the risk of disclosure in the public use file. Banding of variables and rounding consumption values has also been used to protect the data in both datasets. Decisions set out in this consultation have drawn on results from initial testing of the dataset; primarily the end user licence dataset. Further anonymisation techniques (e.g. record swapping and further banding or dropping of variables) will be applied if deemed necessary following further testing.

DECC is continuing to work with VOA to get more straight forward access to the property attribute data held by VOA, with the potential to include these data in an anonymised dataset in future. HM Revenue and Customs (HMRC) have consulted on legislation which may allow DECC to have access to VOA data in anonymised form and if this legislation is implemented DECC will work with HMRC and VOA to see if it would be possible to include the data in a future version of the anonymised NEED dataset.

DECC will continue to look at the potential to include new data sources in NEED, as well as reviewing the approach to access to the data (see next steps for more details).

Question 6: Do you agree with the proposed approach to publication and access?

Three responses addressed both parts of this question together. In all three cases the responses were broadly in agreement with the proposals:

- One respondent agreed with the approach and also suggested DECC provide an indication of plans including timing of future updates and requested that the licence required for the first dataset should continue to apply for updates.
- One respondent was in agreement, but had concerns that public use file may be too small to be representative of some less common types of homes. The respondent requested an increase in the size to around 50,000 records or release of sub-sets of data at local authority level.
- The ICO restated its support for the approach of releasing two different datasets. Its response said “We would broadly agree that limiting the size of the publically available dataset, and the restrictions on access to the more extensive dataset, are sensible steps to reduce the risk that individuals can be identified”.

i. Do you agree with the proposal for a smaller publically available dataset?

There were nine responses to this question. Eight of these responses were yes, with one of these including a request to make the sample representative. The other response was from ODUG which supported the release of the public use file, but wanted to see more data released as Open Data. It also requested the release of the underlying datasets as Open Data and highlighted the benefits of doing this.

ii. Do you agree with the proposed restrictions on access to a more extensive dataset?

There were nine responses to this question. Six respondents were supportive of the approach answering “yes”, while three respondents had more extensive answers:

- ODUG encouraged any licensing to permit as wide a use of the data as possible without major financial burdens on potential licencees and said that any licence used should be in line with National Archive best practise.
- One respondent agreed that access to this dataset must be restricted. However it desired greater clarification on the definitions provided in the consultation regarding access under the end user licence before being able to fully comment.
- One respondent was not convinced that the restriction was necessary given the precautions being put in place to avoid disclosure. However, was supportive of the approach on the basis that more fine grain data could be released as a result.

Government response

Responses to this question were all broadly in agreement and therefore DECC intend to go ahead with the approach outlined in the consultation, with some small changes to reflect respondents comments.

The public use file will be published on the NEED pages of the gov.uk website and available to anyone under the Open Government Licence[11]. This will be made up of approximately 50,000 records, rather than the 20,000 originally proposed. This will allow for the sample to be more representative for less common house types while still allowing DECC to produce a dataset which it can be confident has been sufficiently anonymised. The sample will be selected to be representative of the England and Wales housing stock as far as possible.

The end user licence dataset will be made available through the UK Data Archive under its standard end user licence[12]. This will be free to access and available under the UK Data Archive description of allowable use: *Any individual employed by, or undertaking research for, any organisation, may use data even if this entails monetary reward, where a public good results from the use. Public good can be defined as an activity which widens access to information sourced from our collection and has social or economic benefit.*

The dataset may be used by organisations where the data would support more efficient delivery of Government schemes, subject to these organisations complying with the conditions set out in the end user licence.

We intend that individuals will only need to sign up to the licence once in order to access future versions of the dataset. However we cannot rule out the possibility that future versions of the dataset may require users to sign up to a new licence, for example, if the UK Data Archive revises the terms of its end user licence.

DECC will also continue to look at ways to make the data more accessible to a range of users. This will include further work with data owners and experts in anonymisation to see what more could be published as Open Data and what potential there is to

[11] http://www.nationalarchives.gov.uk/doc/open-government-licence/version/2/

[12] http://ukdataservice.ac.uk/get-data/how-to-access/conditions.aspx

publish a more detailed dataset in a secure environment[13]. DECC will also continue to review the outputs published on the DECC website as Open Data, for example, whether postcode level average consumption can be published in future. More details of future plans are available in Section 3, Next Steps.

Question 7: If you are a potential user, please tell us how you think you would use these data.

There were nine responses to this question, with respondents identifying a range of potential uses for the datasets. These included academic, local authority and energy industry research; it covered use as a primary data source and to contextualize analysis from other sources. Some examples of the possible uses include:

- Modelling energy demand and understanding trends in demand for certain household characteristics;
- Understanding the savings from installing energy efficiency measures;
- Informing probabilistic bottom up stock models;
- Understanding the relationship between theoretical consumption/EPC bands and actual energy use;
- Highlighting behavioural trends such as preference in taking efficiency in savings or comfort;
- Potential to use the data to improve the accuracy of industry settlement in future;
- Use in future 'hack' events (similar to the Open Data Challenge[14]);
- Impact of off gas and E7 efficiency;
- Support prioritisation of action against housing energy performance; and
- To improve the knowledge of council staff in dealing with residents enquiries.

There were some concerns raised that the dataset did not go far enough in supporting targeting of the Energy Company Obligation, but areas where it could help included:

- Investigating relationships between variables to understand behaviours and possible indicators for need;
- Strategic analysis; and
- Identifying geographic or demographic groups which may benefit from interventions.

One user highlighted the many additional uses that could be made of the data if access to a more detailed dataset where made available, especially if it enabled linking with other sources of data.

Question 8. Do you have any other comments on the proposals?

Respondents were given an opportunity to provide any further comments. Six respondents made use of this opportunity. A number of responses welcomed the initiative and reiterated their

[13] For example: http://ukdataservice.ac.uk/get-data/how-to-access/conditions/controlled-data.aspx.

[14] http://www.nesta.org.uk/project/open-data-challenge-series

support for release of the dataset and the benefits it could bring. Some also provided comments which were not covered by any of the previous responses including requests for further data to be published:

- more demographic variables within the dataset such as age, income, household composition and tenure;
- FiTs data (e.g. indicator of households with solar PV);
- a similar dataset for non-domestic data classified by sector and size.

Other comments are summarised below:

- A request for banding to be in line with other commonly used dataset (such as EHS and SAP/RdSAP).
- A reiteration of the request for an additional level of access under a special/secure licence.
- A request for data to be consistent with the new fuel poverty definition to help delivery of the strategy.
- Highlighting the risk of NEED data to be used to justify interference with the roll out of smart meters and lead to changes to roll out plans and additional cost of roll out. There were also concerns with customer consent once smart meters have been rolled out.
- Welcoming the inclusion of the EPC data, and requesting information on:
  - how often EPC data will be updated in NEED;
  - how greater proliferation of EPCs will change dataset numbers over time;
  - how DECC will ensure anonymity is maintained over time with updates and changes to the number of EPCs.

Government response

DECC is grateful for the support for its plans to publish an anonymised dataset. It is intended that these two datasets will be the first of a number of future datasets.

DECC will continue to look at developing the data available through NEED, subject to legal requirements and protecting against disclosure of personal data. This includes additional data sources and making more data available. Specifically, if a data source with information on demographic variables which can be used in the dataset can be found then we would seek to include this in future. DECC is also investigating whether it would be possible to model some of these variables itself to an adequate level of accuracy to provide meaningful results.

A number of the other comments have been addressed through responses to earlier questions. Where they have not been covered, some further responses are below:

- Wherever possible banding used is in line with that used in the EHS or published NEED tables.
- The fuel poverty definition used in the dataset is based on the low income high cost definition for the 2011 fuel poverty dataset.

- DECC will continue to ensure that outputs from NEED including the release of anonymised data are compliant with the Data Protection Act.
- We intend to update EPC data in NEED on an annual basis and to incorporate it into any future release of the data. It is not intended that any future dataset would be made larger as a result of this change, but with the increase in the number of properties having had an EPC it should be possible to better reflect the UK housing stock in the four million household sample (so the weighting variable would become less influential). As the number of properties with an EPC increases, the chance of each property with an EPC being in the NEED dataset will decrease, this will therefore also reduce the risk of identification of a property in the NEED dataset.

3. Summary and next steps

Following the positive responses to the consultation DECC intends to publish anonymised NEED data.

The majority of consultation responses were supportive of the approach outlined and therefore the datasets published will broadly be in line with the proposals; subject to a small number of refinements to reflect responses to the consultation. As planned, two dataset will be published - a public use dataset and an end user licence dataset.

Public use dataset

This dataset will contain approximately 50,000 records. Subject to testing, it will contain over 30 variables including gas and electricity consumption for 2005 to 2012, energy efficiency measures installed in properties, property attributes and a weighting variable. It will be selected to be representative of the England and Wales housing stock. The dataset will be made available to all via the Government website and data.gov.uk.

End user licence dataset

The end user licence dataset will contain approximately four million records covering the same variables as the public use dataset and three additional variables; Region, fuel poverty indicator and EPC inspection date. All individuals wishing to use these data will be required to sign up to an end user licence before being granted access to the data. This licence will allow use of the data at no cost, described by the UK Data Archive as: *Any individual employed by, or undertaking research for, any organisation, may use data even if this entails monetary reward, where a public good results from the use. Public good can be defined as an activity which widens access to information sourced from our collection and has social or economic benefit.*

This approach to publication of two datasets with different content and different access requirements is in line with ICO guidance, and supported by the ICO as it "allows the measures taken to protect individuals' privacy to be tailored to each dataset, bearing in mind the purpose for which each dataset is released, who is likely to use them and the different levels of risk to individuals' privacy".

These proposed datasets will be the first publication of data from NEED at property level. The approach and final content of the datasets have been developed with input from members of the UK Anonymisation Network and support from the Information Commissioner's Office. The work on anonymisation has been seen as an example of good practice with the Office for National Statistics and Information Commissioner's Office requesting a reference report on the work which can be used by their organisations as a case study.

DECC is also extremely grateful to all the parties who have worked with DECC to allow this project to progress, including a range of data providers and potential users of the data. There have also been amendments to the Energy Performance of Buildings Regulations 2012[15] to allow Energy Performance Certificate Data to be included in the dataset.

The next steps for the project are set out below.

- Complete testing of the end user licence and public use files.
- Finalise dataset and accompanying documentation.

[15] http://www.legislation.gov.uk/id/uksi/2014/880

- 29 May 2014 – Publication of public use file on Government website and submit end user licence file to the UK Data Archive (publication on UK Data Archive to follow a few weeks later).

Following publication of the datasets outlined above, DECC will work with a range of organisations to review the publication of the datasets. This will include looking at:

- how the datasets have been used;
- the range of data included in the dataset, such as which variables have been most valuable, increasing coverage to include Scotland and more information on household characteristics; and
- the anonymisation of the data released and approach to release - including whether more data can be made available as Open Data and whether a more detailed dataset can be made available through a secure environment.

DECC will get initial feedback through a seminar with energy suppliers and a planned event for NEED users in summer 2014. Alongside this, DECC will continue to work with the Cabinet Office Transparency Team, Public Sector Transparency Board, Open Data User Group and the Open Data Institute to engage with potential users of an Open Data dataset and understand the value of and priorities for a larger Open Data dataset. A significant part of this will be to understand which variables are most valuable to this group of users; as it will not be possible to publish all variables if a larger public use file is produced, due to the additional risk of disclosure. It is anticipated that the public use file to be published in May will give Open Data users an opportunity to understand the data and its potential uses in order to consider priorities.

DECC will also continue to liaise with anonymisation experts and the ICO to ensure future publications include as much useful data as possible while remaining consistent with ICO guidance and Government best practice.

DECC will produce a short report on its plans for future publication of Open Data resulting from this review by the end of September 2014. DECC plans to publish an updated dataset in spring 2015. This dataset or datasets will be informed by the review and include gas and electricity consumption data for 2013.

DECC will also continue to publish outputs from its own analysis using NEED. The next report will be published at 9:30am on 26 June 2014. This will include analysis of 2012 consumption data and estimates of the typical reduction in annual gas consumption following the installation of energy efficiency measures installed in 2011[16].

[16] Available at: https://www.gov.uk/government/collections/national-energy-efficiency-data-need-framework.

Annex A: Consultation questions

Consultation Question	
1.	**Do you agree DECC should release anonymised NEED data?**
Consultation Question	
2.	**Do you agree with the proposed approach to publishing two separate dataset for different purposes?**
Consultation Question	
3.	**In relation to i) the public use dataset and ii) the end user licence dataset, what are your priorities for variables in the dataset?** **a) Do you agree with the priority variables set out in Table 4.1? If not, which of the variables listed do you consider to be priorities?** **b) Do you agree with the variables assigned as important in Table 4.1? If not, which of the other variables listed do you consider to be important?** **c) Do you agree that those variables listed as "under consideration" are less important than the variables listed as priority or important?** **d) Are there any variables included in the proposals which you think should not be included?** **e) Do you agree that inclusion of a lower level geography identifier is less important than a wider range of variables?** **f) Which lower layer super output area data is most useful? Index of multiple deprivation, output area classification or percentage of households in fuel poverty?** **g) Would a weighting variable be useful?**
Consultation Question	
4.	**Proposed bandings for variables in the dataset are set out in Annex B. Do you agree with these proposals in relation to i) the public use dataset and ii) the end user licence dataset? Please bear in mind that greater granularity of data will reduce the number of variables that can be included in the final dataset.** **a) Annex B sets out options for banding variables please let us know which you would prefer for each variable of interest to you.** **b) Are there any variables that can be banded further than proposed without significant loss of utility?** **c) Are there any variables which would no long be useable for analysis if the proposed banding – or one of the proposed options - is applied?** **d) For variables such as consumption and floor area, is it preferable to have bands of the same size (which may have to be larger) or more detail in the centre of the distribution with larger bands at the extremes?**
Consultation Question	

5.	**Do you agree with the proposed approach to anonymisation for** **i. The public use dataset; and** **ii. The end user licence dataset?**
Consultation Question	
6.	**Do you agree with the proposed approach to publication and access?** **i. Do you agree with the proposal for a smaller publically available dataset?** **ii. Do you agree with the proposed restrictions on access to a more extensive dataset?**
Consultation Question	
7.	**If you are a potential user, please tell us how you think you would use these data.**
Consultation Question	
8.	**Do you have any other comments on the proposals?**

Annex B: Glossary

Open data is information that is available for anyone to use, for any purpose, at no cost.

An **Anonymised dataset** is a dataset in which direct identifiers have been removed. Further protection may be required if indirect identifiers are present in the data.

A **Public use dataset** is typically record level data which can be accessed by any individual, with no restrictions on use. It will not contain personal data. It may be of more use as a training tool than for researchers.

An **End user licence dataset** will have more detail than one for Public Use. Users will have to sign an agreement with one of the requests being that no attempt will be made to use the data to identify any individual, household or organisations.

Anonymisation involves removing the direct identifiers from a microdata record. This term is used frequently when microdata are being protected. Direct Identifiers are variables which will enable a property to be identified with a high degree of confidence such as address.

Microdata are individual level data, for example data about individual people, properties or households.

Indirect Identifiers are variables in a dataset that assist with identification of a household or property without directly referring to them. For example combinations relating to a property in a table could allow an individual to be identified with a great degree of confidence.

Direct Identifers are variables in a dataset will help an intruder easily identify an individual. These include Property Reference Number.

Disclosure Control refers to a number of techniques which can be applied to the data to limit disclosure risk. The most common techniques include recoding, suppression and rounding.

Disclosure risk occurs if information about an individual can be ascertained either exactly or to within a defined narrow bound by an intruder with a high level of confidence. This risk can be mitigated by applying disclosure control.

Intruder refers to a group or individual who wishes to identify people in the table or attributes relating to these people. Also known as an attacker they may or may not have malicious intent.

Granularity is the level of detail provided in the data. High granularity refers to record level data or similar. Low granularity would be aggregated or summarised data.

Key variable is a variable which is commonly used in tabulations. If a large number of tables are produced they are likely to be linked via one or more key variable. By combining these tables an intruder may be able to identify an individual or associated attributes.

Visible variable is a variable which enables identification of an individual or other statistical unit by placing them in a certain category for particular key variables.

Lower layer super output area is a geographic area made up of a number of output areas. Super output areas were designed to improve the reporting of small area statistics. Each LSOA contains between 400 and 1,200 households. There are 32,844 lower layer super output areas in England and 1,909 in Wales.

The Energy Efficiency Commitment (EEC) set targets on energy suppliers to achieve improvements in energy efficiency by providing energy efficiency measures to households across Great Britain. The first scheme (EEC1) ran from 2002 to 2005 and the second (EEC2) ran from 2005 to 2008. EEC2 had a

requirement for at least 50 per cent of the target to be met in relation to Priority Group consumers, defined as those in receipt of certain income-related benefits and tax credits.

The Carbon Emissions Reduction Target (CERT) ran between 1 April 2008 and 31 December 2012 and followed EEC. It required all domestic energy suppliers with a customer base in excess of 250,000 customers (increased from 50,000 at the end of December 2011) to make savings in the amount of carbon dioxide emitted by households in England, Scotland and Wales.

The Community Energy Saving Programme (CESP) targeted households across Great Britain, in areas of low income, to improve energy efficiency standards, and reduce fuel bills. There were 4,500 areas eligible for CESP. Like CERT, CESP was funded by an obligation on energy suppliers and electricity generators.

Annex C: Responses

A number of respondents put their responses on their websites. Where this was the case, links are provided below.

Information Commissioner's Office

http://ico.org.uk/about_us/consultations/~/media/documents/consultation_responses/ICO-response-to-DECC-National-Energy-Efficiency-Data-Framework-consultation-on-anonymised-data.pdf

National Energy Foundation (NEF)

http://www.nef.org.uk/themes/site_themes/agile_records/images/uploads/NEF_response_NEED_consultation_Jan_2014.pdf.

Elexon

http://www.elexon.co.uk/wp-content/uploads/2014/01/ELEXONs_Consultation_Response_NEED.pdf

Decisions on the final dataset have also been informed by discussions at a NEED stakeholder event in May 2013 and DECC's NEED project board in October 2013 (see consultation document Annex C for more details[17]). Input from the ONS and guidance provided by the Information Commissioner's Office has also informed the approach to anonymisation of the dataset and ensuring that household data is adequately protected.

[17] https://www.gov.uk/government/consultations/national-energy-efficiency-data-framework-making-data-available.

Department of Energy & Climate Change
3 Whitehall Place
London SW1A 2AW
www.gov.uk/decc
URN 14D/138