Department of Energy & Climate Change

DECC Stakeholder Event

17 May 2013

The second stakeholder event for users of the National Energy Efficiency Data-Framework (NEED) was held on 17 May 2013. The event was well attended by users from academia, local government, private and third sector energy researchers and Energy UK. It built on on-going engagement with energy suppliers.

The event aimed to give users an update on progress with NEED, and to discuss options and preferences for the release of anonymised data in the future. As part of the event DECC outlined its plan to release an anonymised record level dataset. The discussion relating to this item is summarised below.

Presentations given at the event, covering domestic and non-domestic NEED, are provided on the NEED webpage at: https://www.gov.uk/government/statistical-data-sets/national-energy-efficiency-data-need-update.

Overall Comments

Feedback on NEED was positive with a particular interest in the progress being made with non-domestic NEED.

Users felt an anonymised dataset of any description would be beneficial, but there are also a lot of cases where the aggregate outputs already published - or planned - can be used without the need for the anonymised dataset, such as validating models, or understanding distributions. Not all users were aware of some of the tools and data tables that could help with these endeavours.

Consumption by multiple attributes (unformatted): https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/65986/6950-gas-and-electricity-consumption-multiple-attribute.xls.

Table Creator: https://www.gov.uk/government/statistical-data-sets/need-table-creator

Other data tables including headline consumption and distributions: https://www.gov.uk/government/publications/national-energy-efficiency-data-framework-need-report-summary-of-analysis-2013-part-1

Main uses

Attendees came from a range of spheres and had varied plans for how an anonymised dataset would be helpful. This included work with an interest in behaviours, physical

properties of the housing stock and the ability to target energy efficiency measures. Specifics included:

- Creating and validating models
- Providing a tool for students to better understand energy use and efficiency
- Imputation of missing values in household surveys
- Modelling housing stock
- Fuel poverty impacts

There were also academic users who had resource for analysis and where open to being guided by Government priorities for policy development.

Variables

In general users wanted all possible variables. Specifically:

- Consumption
- Local authority
- Property type
- Property age
- Number of bedrooms
- Floor Area
- Tenure
- Number of occupants
- Measures installed

In addition to those currently held within NEED, users were interested in:

- Wall Type
- EPC band
- Heating fuel/Heating system
- Store height
- Fuel poverty flag
- Stay at home flag (e.g. indicator of those most likely to be at home during the day, pensioners, people who work from home, mothers with young children)
- ESRI mapping information on building direction
- Additional EHS variables
- a weighting variable - to understand the proportion of the population with particular characteristics

It was also felt that it would be helpful to provide totals so users can develop scaling factors. Information on uncertainty would also be valuable.

Banding/level of geography

Attendees where sympathetic to the need to have some banding in order to reduce the likelihood of users being able to identify a specific household within the data. However, there were differing views on the best approach.

Geography - Some users were happy with region as the most detailed geographic identifier, while other felt the dataset would not be able to meet their requirements unless data were available at lower level super output area (LSOA).

Number of occupants – banded by 1,2,3,4, 5 or more.

Consumption data – some users felt distribution (e.g. percentiles, perhaps deciles) was more useful than consumption bands. Another suggested approach was to have 100kWh bands for consumption in the centre of the distribution with larger bands at the extremes.

As a result of the divergent views of attendees each table reached the same suggestion that it would be best to produce two different datasets. One at a detailed geography with fewer variables and one with region as the most detailed geography and more detail in other variables. For some users it was actually post code level consumption data that would be most useful rather than the range of other data which is available through NEED.

Sample size

Views varied from a sample the same size as the English Housing Survey (about 16,000 for England) being sufficient to a request for all households in Great Britain. This depended on intended use.

Access

It was suggested that two different datasets could be made available with different levels of access; a fully open dataset which could be made available to all users and a more detailed dataset which is only available to registered users. It was felt that some users would be happy to comply with any access requirements in order to get hold of detailed data, while others would rather have less detailed data if it meant there were fewer requirements.

Users felt that there is still value in a dataset which cannot be linked to other sources, but that there would be even more advantages if data could be linked.