

Introduction

The Open Data User Group (ODUG)¹ is an independent advisory body appointed by the Chair of the Group and the Cabinet Office Minister. It receives some administrative support from staff in the Cabinet Office, but speaks independently on behalf of the data community. The views of ODUG should not be construed as representing the position of Her Majesty's Government. Individual members of ODUG, or their employing organisations, may make separate responses.

1. Do you agree DECC should release anonymised NEED data?

The release of any reliable government statistics should always be applauded. ODUG notes that the NEED data is an amalgamation of datasets from a number of sources including DCLG, the Ordnance Survey and energy suppliers². The content of this data is a mixture of personal and non-personal information and because of this, ODUG agree that any initial release should be anonymised and non-personal.

In the future, DECC should work with a wider group of stakeholders to ascertain the potential value for a properly licensed and controlled dataset of address-level information. There could be potential benefits to consumers (by taking control of their own energy data using a variety of data aggregation products & services both existing today and in development) and the various participants in the energy market (suppliers, retailers, comparison sites, data & insight providers etc).

However, any non-personal information should be released for free under the Open Government License.

2. Do you agree with the proposed approach to publishing two separate dataset for different purposes?

The Open Data User Group regularly argues that the publication of non-personal data in an open format for unrestricted use should be supported. Currently, the plans for NEED are a good start but the value of the unrestricted use of a 20,000 record "sample" dataset may be limited. ODUG would recommend that DECC and NEED Stakeholders engage with the Open Data community via ODUG and bodies such as the Open Data Institute to gather more insight into what form the sample dataset should take.

The larger, 4 million record dataset is likely to have very limited reach in terms of UK economic and social value if released under the currently proposed license. Whilst ODUG recognises that the NEED data requires careful handling because of sensitive usage statistics on individual households, we call on DECC to seriously examine the release of a more detailed dataset under less restrictive terms. The UK Data Archive license³ appears to be different in scope and terms to many of the standard data licenses used by the likes of the ONS and National Archive. ODUG encourages DECC to re-examine the proposed license to ensure that it is fair to users of all size and encourages wide use of the NEED data.

¹ <http://data.gov.uk/odug>

²

https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/260317/DECC_National_Energy_Efficiency_Data-Framework_consultation_on_anonymised_data.pdf See Table 3.1

³ <http://data-archive.ac.uk/media/381244/ukda137-enduserlicence.pdf>

ODUG would support the publication of a single NEED database at postcode or another level of 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 License⁴ and would, ideally, contain full national data.

3. In relation to i) the public use dataset and ii) the end user license 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?

i) ODUG agrees that property variables and the energy use of these properties is likely to be of most value to users of NEED. However, without location, this value will be diminished.

ii) As this data is not considered for Open release, ODUG will not comment on this section.

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?

i) ODUG represents the views of Open Data users of all shapes and sizes. Therefore, a single answer to this question is difficult to formulate. However, ODUG would always make the point that location is of paramount importance in order to add context to data⁵. We would therefore recommend that the Public Use database contains some reference to Region, Output Area, City or Postcode area to allow users to map the data and combine it with other sources to deliver maximum value. One further observation of the current list of 'important' variables is its current skew in favour of variables that would most assist participants in the Green Deal⁶. 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 in order to make the data useful to sectors of UK PLC which may not have been involved in the NEED Stakeholder group.

ii) As this data is not considered for Open release, ODUG will not comment on this section.

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

i) As in earlier answers, the inclusion of Location variables should be a basic expectation of this kind of data to enable users to add 'place' and further context. By having locations, data linkage is made possible which often leads to enhanced value for end users and data service companies. We would also encourage DECC to include information on environmental impact, fuel poverty, fuel & meter type (gas mains, economy 7 etc) and wall construction to further enhance the value of the NEED data.

ii) As this data is not considered for Open release, ODUG will not comment on this section.

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

⁴ <http://www.nationalarchives.gov.uk/doc/open-government-licence/version/2/>

⁵ <http://data.gov.uk/blog/open-data-user-group-response-to-rm-postcode-address-file-licensing-consultation-%E2%80%93-september-2013>

⁶ <https://www.gov.uk/green-deal-energy-saving-measures>

i) We cannot see any variables listed which would not be of value to potential users. We simply echo earlier statements that the more data fields included, the more valuable the data to a wide audience.

ii) As this data is not considered for Open release, ODUG will not comment on this section.

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

i) ODUG would argue that the two options should not be mutually exclusive. With a wide range of potential users, the NEED database should reflect as much of the content as possible at a geography which is relevant to end users (Postcode area being a common language in similar data-based applications such as property comparison). ODUG would encourage DECC to release all variables at the lowest common geography possible or clearly label each field with the relevant geographic level.

ii) As this data is not considered for Open release, ODUG will not comment on this section.

f) Which lower level super output area data is most useful? Index of multiple deprivation, output area classification or percentage of households in fuel poverty?

i) Due to the large number of potential users of the data, with varying requirements; it is difficult to give a single answer. We would encourage further engagement with Open Data users to ascertain whether one SOA is more useful than another to a large majority of users.

ii) As this data is not considered for Open release, ODUG will not comment on this section.

g) Would a weighting variable be useful?

i) This would certainly be of value to enable users to apply the findings of the sample to a national or local average. Of more value would be a larger sample ensuring even coverage across England & Wales.

ii) As this data is not considered for Open release, ODUG will not comment on this section.

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 license 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.

i) Our response takes into account a 'middle ground' for all fields noting that this may not be the best fit for all potential users represented by ODUG. We would encourage DECC to work with the Open Data community via bodies such as the Open Data Institute to study the options in more detail to present back some case studies to a wider group of stakeholders. However, some observations on specific fields follows:

Field	Option	Why
-------	--------	-----

Gas consumption 2005 - 2012	C	Gives more detail for residential properties, harder to work with though. Option B good alternative.
Electricity consumption 2005 – 2012	C	Gives more detail for residential properties, harder to work with though. Option B good alternative.
Energy Efficiency Band	C	Would prefer B but not at detriment of overall data quality. Please detail these groupings.
Environmental Impact Band	C	Would prefer B but not at detriment of overall data quality. Please detail these groupings.
Property Age	A	Makes best use of raw data source.
Property Type		Would prefer two variable fields. For example, House or Bungalow and then End Terrace or Detached. Much like the 4-part classifications used in AddressBase.
Floor area band	A	Appears to give the most data.
Main heating fuel	A	Appears to give the most data.
Loft insulated thickness	A	Appears to give the most data.
Wall construction	B	Appears to give the most data.
Cavity wall insulation year	A	Easier for users to understand.
Loft insulation install year	A	Easier for users to understand.
Solid wall install year	A	Easier for users to understand.
New boiler install year	A	Easier for users to understand.
Output Area Classification (OAC)		Please do use 2011 census outputs

Of further note on Energy Efficiency Band and the other fields taken from EPC data are a number of requests to the Open Data User Group for an open, national register to be made available of both domestic and non-domestic EPC data⁷. ODUG would like to see priority given to the release of open EPC data over the NEED database.

ii) As this data is not considered for Open release, ODUG will not comment on this section.

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

i) The Open Data User Group is unable to make any recommendations at this stage.

ii) As this data is not considered for Open release, ODUG will not comment on this section.

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

i) The Gas & Energy Consumption fields would need careful consideration if a 5,000kWh banding were used as this may group a large number of residential properties together where their actual consumption varies greatly. However, further detail on the numbers of properties this could affect would help make a more accurate assessment.

ii) As this data is not considered for Open release, ODUG will not comment on this section.

⁷ <http://data.gov.uk/data-requests/environmental-performance-certificates>, <http://data.gov.uk/data-requests/non-domestic-epcs-non-domestic-epc-register>, <http://data.gov.uk/data-requests/averaged-sap-rating-based-on-lsoa-level>

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

i) As in c), the Gas & Energy Consumption fields would need careful consideration if a 5,000kWh banding were used as this may group a large number of residential properties together where their actual consumption varies greatly. Floor area faces a similar issue with the difference between residential and commercial properties. However, the addition of the building type may negate some of this risk.

ii) As this data is not considered for Open release, ODUG will not comment on this section.

5. Do you agree with the proposed approach to anonymisation for

i. The public use dataset; and

Making use of the ICO anonymisation code⁸ is a perfectly reasonable approach. ODUG would suggest that the VOA data currently excluded should be re-visited in light of the recent HMRC consultation on data sharing⁹. We would welcome the opportunity to assist in conversations between the VOA and DECC to encourage the release of VOA data for open use and inclusion within NEED.

ii. The end user licence dataset?

As this data is not considered for Open release, ODUG will not comment on this section.

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?

The Open Data User Group support the release of an anonymised dataset for 'public' use but would encourage DECC to increase the sample size of the data to increase use, re-use and the overall benefit to UK PLC and society. We would also encourage DECC to make use of the Open Government License for this dataset.

ODUG would also encourage the owners of the constituent public sector databases (for example; AddressBase, EPC data, VOA data etc) to focus their efforts on releasing these publicly owned datasets as Open Data over their use in closed-user group initiatives such as NEED. It is a general trend within the Open Data movement that the most utility can be made from raw data rather than amalgamations and statistics. Giving access to the raw constituents of NEED would enable many businesses, charities and other bodies to run their own analyses of energy efficiency within the UK to create various products, services and news-worthy items.

We would also welcome the opportunity to speak to the non-public bodies involved in NEED such as xoserve, Gemserv and the utilities companies to discuss how they could make more data available to their customers and business partners (such as switching websites) to create better customer-focused solutions to energy efficiency and use.

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

The Open Data User Group would never recommend the open release of personal data but would always encourage publishers to carefully consider what constitutes personal data. We would also encourage any licensing to permit as wide a use of the data as possible without major financial burdens on potential licensees.

Finally, any license used should be in line with National Archive best practise.

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>

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

The Open Data User Group represents a wide range of potential users making a single response difficult. We would encourage DECC to work with bodies such as the ODI¹⁰ and NESTA¹¹ to promote NEED through workshops, 'hack' events and so on to better understand the potential uses.

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

The Open Data User Group wishes to thank DECC for the opportunity to comment on these proposals and would welcome the opportunity to represent the Open Data community further as the NEED publication continues. We also call upon the owners of NEED inputs to work to make these available as Open Data to increase the value of NEED and the amount of Open Data available to UK consumers and businesses.

¹⁰ <http://theodi.org/>

¹¹ <http://www.nesta.org.uk/our-projects/centre-challenge-prizes>