RCUK Centre for Energy Epidemiology

Department of Science, Technology, Engineering and Public Policy
UCL STEaPP

Social Research that works for society

UK Energy Lab:
A feasibility study for a longitudinal, nationally representative sociotechnical survey of energy use

Final Report

October 2014

EXECUTIVE SUMMARY

Authors
Adam Cooper
David Shipworth
Alun Humphrey

With input from
Simon Elam
Ian Hamilton
Gesche Huebner
Rob Liddiard
Gareth Morrell
Anni Oskala
Alison Park
Michelle Shipworth
Phil Steadman

Disclaimer

The views presented in this report are the authors' own. Nothing in the report represents a commitment to funding or a statement of policy by those public bodies involved or mentioned, including the UK government Department of Energy and Climate Change.

For comments, queries, critiques of the content of this report and associated annexes, please contact Dr Adam Cooper at UCL STEaPP via adam.cooper@ucl.ac.uk or Dr David Shipworth at UCL Energy Institute d.shipworth@ucl.ac.uk

What this project has done and the questions it answers

The RCUK Centre for Energy Epidemiology at the UCL Energy Institute, together with UCL Department of Science, Technology, Engineering and Public Policy and NatCen Social Research undertook a 6-month study (January to June 2014) to review the feasibility of setting up a nationally representative longitudinal panel of energy use in the UK. The study was supported by the UK government Department of Energy and Climate Change (DECC) and this report is to provide DECC with the information needed to decide whether to support further stages of development.

The project was split into 8 strands each of which considered different elements necessary for the successful development, deployment and stewardship of such a project. For the duration of the project we have referred to the potential panel as UK Energy Lab[1] (UKEL) to try and emphasise that this would have both technical and social data collection – making it likely more similar to a medical cohort than a standard social survey. Below we synthesise and summarise the findings from the feasibility study to directly address a series of important questions:

- Do we need UKEL now?
- Who wants it?
- What's the minimum survey size that is useful?
- Can it be done?
- What might it cost overall and to government?
- What are the benefits?
- What are the risks?
- When can it be rolled out?

Do we need UKEL now?

Currently the UK plans to decarbonize energy in homes completely by 2040, just 25 years away. This is a key part of the UK's plans to address climate change. Yet there is very limited data about energy demand and what drives it in homes. What studies there are either have limited data collection types or frequencies or are small scale and one-off. **Annex B** to the main report identifies the available data in this area and clarifies the gap a UK Energy Lab-style survey would fill. The established energy industry and emerging energy SME sector companies are seeking to innovate into uncharted territory but are hampered by the lack of available data, knowledge about what drives energy use in UK homes and the analytic capacity to develop it, meaning the UK is currently poorly placed to identify where the opportunities are to innovate effectively to help reach the 2040 target. There are also key areas such as deployment of new renewable technologies in the home and fuel poverty that lack a strong evidence base for making effective policy. If we do not start to collect this data in the next couple of years it will be too late to take full advantage of the outputs in time to meet the 2040 target.

[1] The 'UK Energy Lab' name has been kept for the duration of this project for the sake of continuity. Following this, the project will be called the UK Longitudinal Energy Survey (LUKES).

Who wants it?
In reviewing the major energy research and innovation institutions positions with respect to the kind of data UKEL would collect, we have not encountered any individual or institution that thought it was not needed. Indeed, entirely independently, several major reviews of the research needs in the energy sector have identified the need to significantly improve the data collection in and around demand side factors. This includes Building DECC's Evidence Base, the RCUK Energy Strategy Fellow report 'Investing in a Brighter Energy Future' and the Low Carbon Innovation and Co-ordination Group' (LCICG, comprising 8 core members and 9 associate members including 6 Whitehall departments) publications have made the same point (see **Annex A** to the main report). An open session with academics across the UK showed both very strong interest and demand for such a facility. It is also clear that the data drawn out of a UK Energy Lab-style survey would augment current investments that form the backbone to UK data collection in this space, specifically the National Energy Efficiency Data-framework (NEED) and the English Housing Survey (EHS).

What's the minimum that is useful?
Ostensibly the UKEL could follow the example of the RECS in the US and keep the survey down to a straightforward home visit and interview with the householder (i.e. a standard social survey). Such a social survey, conducted with a relatively standard cross-sectional design would cost in the region of around £0.5-2M per annum depending on the sample size.

While potentially useful, this has two main limitations that suggest such an approach would not be as good value for money as alternatives:

1. Self-report of energy use, and properties of the home have been shown to be highly variable and lack validity. This means the data collected would not be as valid as direct measurement of these elements. There is no doubt such data would be useful, but without also collected direct technical measurements a range of important scientific and policy benefits would be lost.
2. Internal conditions of homes are highly dynamic across a number of temporal and spatial scales. What is critical is the inter-relationship between these conditions and energy consumption. A cross-sectional annual survey would not be able to resolve the dynamic nature of energy consumption within homes.
3. Finally, data that is most effective is data that can address causality. By unpicking causality, this gives policy-makers much better information for understanding what kinds of interventions are most likely to lead to outcomes of interest. Longitudinal data can resolve causality better than cross-sectional data on account of being able to identify conditions existing prior to a change within-cases, and via statistical modelling, determine what factors were most critical in provoking the change.

The trade-off then is to collect only social/self report data for around £1M/yr (roughly a 10,000 sample) and have data that is of marginal importance for a few areas, or invest in a study that generate a wide range of fundamental benefits at a cost of about £5-6M a year over an equivalent 5 year period (with a similar sample size). Since collecting both social/human and technical (energy, materials and environmental) data in an integrated way are critical for a step-change in understanding, this raises two issues regarding scale of the potential study.

1. For the data to be useful and informative for policy development, a sufficient sample size is needed to obtain a reasonable level of precision (that is, 95% confidence intervals on estimates of around ±4% at most, but preferably ±3% or less) for estimates that cut across aspects such as the household and building type (see **Annex A**). To obtain that level of precision requires achieving sub-sample sizes of around 1000, implying an overall sample size in the region of 10,000 where important subgroups occur at a rate of 10% in the population.
2. Combined with this is challenge of collecting the 'right' technical data across this sample. Since there are around 100 different factors that may be important to measure at a range of different spatial locations and temporal resolutions in the home, the potential demands on data collection across 10,000 homes are enormous. On account of this, there are practical, financial and ethical challenges to installing large numbers of sensors in large numbers of homes. For a start, installing 30+ sensors in one home is estimated to cost around £7-8000 per household. To roll this out across 10,000 households would cost over £100M over five years. The high cost, coupled with the practical challenge of fitting 350,000 sensors and managing the data they transmit as well as the ethical challenge means that such an approach is not feasible.

Making the unfeasible, feasible

In the main report, and explained in detail in **Annex C,** we set out a design that addresses this difficulty. We start by setting out 4 different data types (social/human, building, energy and environment) and 3 levels of data collection. The different levels are key to the design, as we address the challenge of breadth and depth via the deployment of different sample sizes at each level. We work on the basic premise, as set out in point 1 above, that the minimum total sample size each wave would need to be 10,000 homes, with around 80% longitudinal.

Level 1: This is data collection carried out by a normal interviewer, who conducts a face-to-face interview, and may take some simple measurements of the home. All 10,000 homes each wave would receive this level of data collection.

Level 2: A trained surveyor would carry out this level of data collection, undertaking a reduced Standard Assessment Procedure and installation of around 5 different devices, including thermometers, humidity and volatile organic compound (VOC) detectors. This would be carried out in around 2800 homes a year for the first 3 years, reaching the entire longitudinal panel after 3 years. This level enables the acquisition of technical data at a scale that means meaningful estimates can be generated at national and subnational levels and for certain subgroups after a couple of years.

Level 3: A trained electrician/sensor network specialist would carry out the installation of 20- 30 different devices and network them in. This would require 2 full day visits to install and decommission, plus a likely further half-day de-bugging. The intensity and cost of this installation means it would be limited to around 100 homes per year. The purpose of this data collection is to quantify the uncertainties in data collection at level 2, giving greater confidence that the estimates are meaningful.

The overall design we suggest is represented diagrammatically in Figure 1. We propose to initiate the study with an initial Level 1 wave across the entire sample. After wave 1, we anticipate some sample attrition from the panel that we propose to address with an ongoing cross-sectional element of about 2000 homes. Around half of these will be drawn into the subsequent wave of panel data collection. This also serves to measure panel conditioning effects. In waves 2-4, level 2 and 3 data are collected as described above, until wave 4 where the entire longitudinal panel has received level 2 data collection, and visits/revisits/refreshes are undertaken for movers and those brought in to refresh the panel from the cross-sectional element. A more detailed diagram of the overall design can be seen in Appendix B to **Annex C** of the main report.

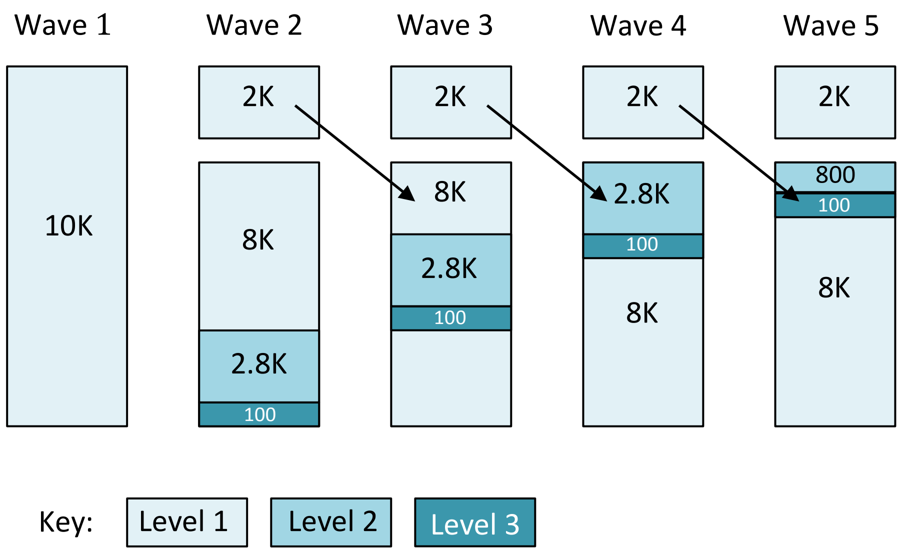

Figure 1: Schematic showing the overall design of the proposed UK energy lab sampling approach over different waves of data collection. Each year a total sample of 10,000 is achieved. For waves 2-4, this comprises 8,000 longitudinally and 2,000 fresh each wave. Within the 8,000, 2,800 homes will receive level 2 monitoring on top of level 1, and 100 homes will receive level 3 in addition to level 1 and 2. Crucially, because the 8,000 homes are longitudinal, level 2 data collected from them for one year in say Wave 2 will likely remain largely relevant to data collection at Waves 3 and 4. This means that the achieved total sample size for level 2 data

collection *adds up* over waves, giving a large overall achieved sample of 8,000 level 2 homes by Wave 5.

Can it be done?
Yes, we have the technology to monitor at Level 2 effectively at scale, and we know roughly what to measure to start with. It will take a medium-term programme of research alongside (via the Level 3 monitoring of home) to improve it. After 3 to 4 waves system for understanding energy use in UK homes is likely to be significantly more effective. What is missing is the integrated research expertise across social science and engineering/physical science but that has been held back in part due to the development of parallel research programmes. A programme like a national energy panel would address that directly by providing a UK 'lab space' to develop better-integrated interdisciplinarity.

What might it cost overall and to DECC in particular?
The estimated minimum cost of the proposed design is in the region **£5-7M per year** or **£28M** over 5 years in today's prices. This includes all the set up piloting and development work required to minimize risks around the roll out of widespread energy and environment monitoring. The costs to DECC will depend on the funding model that can be put in place. The assumption we have worked with in the main report is that research councils would be the best source of lead funding for this, given their role as longer-term guardians of the science infrastructure in the UK, and their role in defending and promoting independent, high-quality research. These are essential qualities for the long-term sustainability of a programme like UKEL as well as ensuring impact in public policy settings. Nevertheless, in order for the UKEL to have the highest impact we propose that government bodies should contribute a significant but minority contribution to the execution and analysis of the survey and subsequent data. On this basis we suggest the maximum contribution to direct funding from government to be 40% of the total budget. Over 5 years, this would mean annual investments of between around £3M for waves 2-4 years, followed by around £2.5M per year. If DECC was a lead funder of a government-led group, contributing half of that 40%, then the wave 2-4 contribution would be in the order of £1.5M per year, and £1.3M on-going after that. Notably, if the total sample size is increased to 14K, and government contribution reduced to 30%, with DECC contributing half of that, the DECC costs would drop by about 10%, while benefiting from great statistical power. Variations on this can be explored through a cost calculator[2] supplied as appendix A to **Annex C** of the main report. An exploration in the governance models upon which a funding split is likely to rest is explored in **Annex E**.

Should the BIS consultation on Science Capital Funding identify a UKEL facility as important to fund, this would alleviate most if not all the set up costs for those supporting it over the longer term. An answer on that is anticipated by early 2015.

[2] Only available to DECC staff internally due to the data supplied being commercial in confidence

What are the benefits?
In the main report, we identify a wide range of benefits that accrue if the minimum level of sampling as set out in the recommended design is followed. We list these as headlines only below, and give some more detail here on the kinds statement that data from the UKEL could support as an illustration of the power (and limits) on the design.

A UKEL can provide benefits to policy evaluation, innovation of technologies and a strategic data 'backcloth' for adding value to a huge range of studies as they are able to understand how the data they create fits into the wider picture in the UK. A deeper analysis of benefits is set out in the main report and **Annex A**.

The kind of data that a UKEL can provide will support conclusions that are currently unavailable via data that is currently produced. An example statement that the UK EL would enable once up and running:

Following [*some change: e.g. head of household becomes unemployed; household buys new white goods*] **in** [*type of home: e.g. 2 person household in a terrace*] **the** [*pattern/level/type*] **of** [*a dependent variable: e.g. energy consumption/comfort*] [*remained stable/changed*] **by** [*X amount/way*] **compared to equivalent UK homes** [*without the change*] **over** [*period of time*] **likely because** [*main reasons for change*].

There are limits on the nature of changes that can be detected given the size of the sample, what is monitored for, frequency of monitoring and so on. As a guide, for a sample size of 8000, we'd expect there to be about 2200 terraces. If half of them (1100) experience the change then we would be able to detect changes in energy use (including load profile over the day) of around 3% or more. Where energy use changes are bigger, sub-groups within terraces could potentially be distinguished. For changes that occur with less prevalence in any one year, it is possible to brigade the multiple instances across the years to generate sufficient power to detect smaller changes. This enables better detection of impacts by controlling for wider background factors.

Other types of statement are also supported including those that identify the longer term and wider types of impact of changes in energy consumption (e.g. on health, employment and so on).

There are currently no sources of data in the UK that can support the same kind of statements with the same power. The baseline design outlined in the main report supports this kind reporting.

What are the main risks?

- **Recruiting enough suitable homes for level 2 monitoring**. Since nothing on this scale has ever been attempting with this level of monitoring it is possible that unforeseen difficulties may arise. The main way of handling this risk is to roll out level 2 monitoring over a number of years. In the baseline scenario this is done over 3 years with a target achieved sample per year around 2800.

Mitigating this risk further can be done with a slower roll out over more years (e.g. 1400 per year over 5 years).

- **Whether we have the capability to remotely collect data effectively**. This is important in getting the data back rapidly and with a high sample response rate. It also reduces costs of data collection. All of these are benefits of the UKEL that would be reduced if monitoring data had to be collected manually.
- **Sufficient numbers of surveyors/interviewer capacity**. We don't know if there are enough surveyors to do the number of properties we are talking about. This is why the slower level 2 roll out is safer than a rapid roll out. Likewise we are not sure whether the interviewers have or could rapidly acquire the right skills to roll out this. This is a low risk medium impact risk.
- The other risks are the **complexity of the data** overwhelming our capacity to utilize it. This will need to be addressed head on with research council funded programmes to make use of this data and ensure that the data collected fits in to current modes of analysis sufficiently.

When could this be rolled out?

If the recommended design is one that remains, then the timing for roll out may depend on the processes used by whichever organization is leading the commissioning. We have set out a 'minimum reasonable timetable' that enables the right consortium and Wave 1 development to take place which indicates a Wave 1 roll out from late 2016 at the earliest. The pace at which more direct buildings, energy and environmental monitoring is rolled out depends in part on budgets and in part on the availability of technologies at scale and surveyors. Our recommended design rolls out direct monitoring of 8000 homes over 3 years following Wave 1, which might commence on Wave 2 (2017-18) or 3 (2018-19). A detailed consideration of the roll out process and the necessary management infrastructure is explored in **Annex F** to the main report.