RCUK Centre for Energy Epidemiology

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

NatCen
Social Research that works for society

UK Energy Lab:

Feasibility Study Final Report

August 2014

Annex D – Ethics and Security

Authors
Gesche Huebner
Simon Elam
Rob Liddiard
Gareth Morrell
David Shipworth
Ian Hamilton
Anni Oskala
Adam Cooper

1. Overview 3
a. Identifying and reviewing issues around data privacy 3
b. Identifying health and safety threats to all parties involved 3
c. Estimating cost impacts of the different ethical issues identified 4
d. Reviewing common ethical procedures 4
e. Ethical issue when researching companies 5
f. Identification of issues related to data privacy & commercial confidentiality and data access where major energy companies are involved 6
2. Data Management and Security Report 7
3. Identification of ethical research in domestic buildings 12
4. Cost implications of ethical issues 16
5. Review of common ethical procedures 19
6. Information on Ethics Approval Procedure UCL 25
7. Non-domestic considerations 28
Appendix A: UK Energy Lab - Data Management Plan Template 32
Appendix B: Details for the individual steps in the UCL ethics procedure 38
Appendix C: Extract of regulations covering data collection for the US energy survey of commercial settings (CBECS) 40

1. Overview

The aims of the workgroup on "Ethics & Security" were to

a. identify and review issues around data privacy
b. identity health and safety threats to all parties involved
c. estimate cost impact of the different ethical issues identified
d. review common ethical procedures
e. discuss ethical issues when researching companies
f. identify issues related to data privacy & commercial confidentiality and data access where major energy companies are involved (no attachment but summarized below)

All these points are summarized below and detailed further in the following sections.

a. Identifying and reviewing issues around data privacy

The review on Data Management and Security reviewed how to ensure secure data management throughout all stages of the data lifecycle. Two basic principles were identified that need to be followed:

1) Data security should be configured so that there are multiple layers of security through the system and, ideally, more than one layer of security protecting data at every step.
2) The separation principle states that identifiers and data content should be separated whenever possible.

A data management protocol will have to be established which is a quasi-legal document summarising the key responsibilities of each organization handling UK Energy Lab data and describing how these organizations will operate and manage the data between them. This Data Management Protocol will state how the data will be managed in order to comply with relevant legislation such as the Data Protection Act 1988 as well as relevant standards such as ONS Code of Practice for Official Statistics.

b. Identifying health and safety threats to all parties involved

Three major areas of ethical issues will need consideration, focused on participants, on fieldworkers, and on detection of a hazard in a home (in addition to issues around data protection which are considered separately). The exact nature of issues will depend highly on what is asked, who is asked, and what data are collected.

There will be some sensitive questions such as around income or health. It would be advisable to allow refusal to a question but to minimize this from happening by emphasising anonymity and confidentiality of all data collected.

It will need to be decided if vulnerable adults and children should be included in the research. If yes, additional information material would have to be generated, e.g. suitable for children. Also, it will have to be decided whether to provide information and translators for data collection in households where English is only poorly spoken and understood. In order to be inclusive, it is highly recommended to do so, and have a boost sample to ensure enough participants of foreign backgrounds can participate.

Depending on what data is collected from sensors or technical kit, a number of issues might arise. In general, using prototypes as opposed to off-the-shelf-risk would be riskier and need additional testing. It needs to be ensured that no physical hazard results from any equipment. There is also the risk that participants will have concerns, such as the data revealing when they are in the house leading to (for instance) an increased chance of burglary. Hence, data security needs to be ensured and emphasized to participants.

On-site data storage might be preferable under this point. Also, rules for compensation need to be in place, in case any damage to the property happens, and if equipment is employed that uses electricity supplied to the house.

Regarding fieldworkers, the main issue is safety, given that homes of strangers are visited, and regarding the installation of kit in the home. Adequate training on safety issues and on kit installation needs to be provided, and liability insurance for fieldworkers needs to be in place.

A final set of issues concerns what to do when detecting a hazard in a building, such as structural damage, or neglect. When detecting a hazard the home occupier needs to be informed and advised on how to proceed. Regarding detection of other issues such as neglect, it is recommended to report to an Ethics board that then considers each individual case.

c. Estimating cost impacts of the different ethical issues identified

There are several areas where ethics issues may incur additional expense to the project. The one with potentially largest costs associated would be regarding translation of material for those not speaking English well enough to participate in an English survey, for children, and into Welsh, should the survey be carried out in Wales. Other ethical considerations that would entail costs are giving participants access to their own data, in addition to receiving regular communication for panel maintenance.

Also, ensuring encryption of all data would be advisable to ensure no sensitive personal information could leak but this would come at a cost. Also, there is balance between the moral obligation of identifying adverse conditions in households and the costs involved, for example when performing screening on all data to detect for example abnormally high and low temperatures.

Finally, issues around electromagnetic radiation might arise; should participants be concerned or report an interference between survey equipment and home equipment, this would necessitate the visit of a fieldworker.

d. Reviewing common ethical procedures

Ethical procedures across institutions (Universities, Research Councils, NatCen) are based on similar principles, and all follow international standards such as the Declaration of Helsinki or Statistical Institute Declaration of Professional Ethics.

The three main standards are:

1. Informed consent

All participants must be fully informed of the study and what is being asked of them, including the potential risks/benefits and exclusion criteria, in order to make a fully informed decision about whether or not to participate in the research. This must be an active step on behalf of the participant and not due to any inducement, coercion or perceived pressure to participate. This is required of all participants in a research study, except where there is a justification for covert research or deception (such cases will be considered on an individual basis by the ethics committee).

2. Benefit not harm

Research involving human participants must have a benefit to society and the risks involved to participants must be balanced against the potential benefit to the overall community.

3. Confidentiality

All participants have the right for their participation to remain confidential in that only the researcher will be aware who has participated. Generally all data will also be anonymous in the final report so that nothing can be attributed back to an individual participant. There are exceptions, for instance where participants wish to be identified or they cannot realistically have their identities kept confidential, but written informed consent must be obtained from the individual participant in advance.

Examples of how to obtain ethical approval are outlined the accompanying documents. It is worth noting that UCL for example would not demand an Ethics application if all data were collected from a third-party organization, i.e. no staff of UCL would go into the field but would only analyse anonymized data. However, it would need to be ensured that the third party complies with the Data Protection Act.

e. Ethical issue when researching companies

The most important issues around researching commercial companies were:

- commercial confidentiality: It is much more likely that an individual company could be identified based on some accompanying information such as type and size of business, and broad location, which could be a competitive disadvantage to companies.
- Risk of physical harm is much higher when researching companies, in particular industrial ones and workshops. Hence, adequate training and equipment needs to be provided.

A range of other ethical issues apply, comparably to domestic properties, such as hazards caused by equipment, access to data, etc.

Given the issue of potential identification of businesses, it is suggested to move away from trying to sample all types of non-domestic properties. In the separate non-domestic report (project stream 8), it is suggested that the non-domestic element of the UK Energy Lab could take the form of a subset of the wider domestic sample, which would cover smaller businesses either operating from domestic premises, or where domestic and non-domestic share the same building, e.g. a dwelling over a shop, where there is no effective separation of the two.

The 'co-habitation' of domestic and non-domestic activities poses some ethical issues in that it becomes problematic to separate the two in terms of space use, economic activity and energy use and most importantly, where does the subject become an individual, as opposed to a business and vice versa? Any study would have to provide clear definitions of who is the unit of research.

f. Identification of issues related to data privacy & commercial confidentiality and data access where major energy companies are involved

There are currently three avenues being explored for the collection of energy data none of which involve engaging with energy companies. It is anticipated that at a minimum participant permission will be sought to access the Annualised Quantity data held by DECC for each MPAN and MPRN of participants' meters. With explicit participant permissions, this data can then be sourced directly from DECC. Two avenues are being explored for the collection of high-frequency energy data (half-hourly and above). The most likely option is the mandating of duel-fuel smart-meters in participants homes as a prerequisite for participation in the survey, followed by instillation of a Consumer Access Device (CAD) that detects the signal from the Home Area Network and relays this back, independently of the energy supplier, to a data repository controlled by the project. This would allow sampling energy data at the frequency limits of the HAN (half-hourly for gas and around every seven seconds for electricity). This again needs participants' permission, but would be part of the wider sensing capacity of the trial and the CAD would also be the base-station for any additional environmental monitoring done within the home. The third option would be to gain participants' permission to access their data through the Data Communications Company (DCC). This would incur the costs of establishing the Energy-Lab as a DCC User, and would again require individual consent from each participant, but would then provide a ready-made secure data transfer route for high frequency energy data via the DCC.

None of these routes would require consent of the energy supplier, however, collaboration with energy suppliers would be required for installation of smart meters in participants' homes.

2. Data Management and Security Report

a. Introduction

The key to managing the data collected and generated by the UK Energy Lab is to ensure that all the steps in the data lifecycle have been identified and there are clear protocols and responsibilities for the secure management of data at every stage.

NB. The detail regarding operational data security cannot be documented until the overall system design and sub-components have been defined in detail.

The following documents will help ensure secure management of data through all stages of the data lifecycle:

- Data Management Protocol
- Data Flow Diagram
- Data Responsibilities Matrix
- Data Management Plan

b. Multi-layered security

Data security should be configured so that there are multiple layers of security through the system and, ideally, more than one layer of security protecting data at every step.

Details of the security to be applied at each step cannot be provided at this stage as it will be dependent on the technical specification of, for example, devices and communications equipment but all components of the data system should follow best practice and standards within the relevant fields. For example, data should be stored on dedicated, secure servers with access restricted to authorized individuals and data should be transferred using secure communication channels. Data should be encrypted, where relevant using industry standards like AES-256 and/or industry-leading software such as PGP (Pretty Good Privacy).

To avoid potential issues with unencrypted data being intercepted by nearby buildings (or in multi-householder dwellings), all monitoring devices should transmit data in encrypted form (even to an in-home communications hub). If this is not technically feasible then any related issues with data security must be re-examined.

c. Separation principle

The separation principle states that identifiers and data content should be separated whenever possible. In effect this forms the first layer of security as there will be very few instances where data and identifiers cannot be kept separate.

In practice, a data table/file will be stored on a secure server containing identifiers (e.g. names and addresses) and corresponding surrogate keys (e.g. survey number, monitoring device serial number etc.) for each record. Data can then be collected, transferred and processed using the surrogate keys which cannot be used to identify a person or address.

d. Data Management Protocol

A Data Management Protocol is a quasi-legal document summarising the key responsibilities of each organization handling UK Energy Lab data and describes how these organizations will operate and manage the data between them.

The Data Management Protocol will state how the data will be managed in order to comply with relevant legislation such as the Data Protection Act 1988 as well as relevant standards such as ONS Code of Practice for Official Statistics.

The Data Management Protocol will have some similarities to DCLG (2013) 'English housing survey: data security strategy and arrangements'[1]

e. Data Management Plan

A Data Management Plan should be created in the early stages of the project and revised, as appropriate, as the project evolves. Data security is one element of a data management plan.
See Appendix A – 'UK Energy Lab - Data Management Plan Template'

f. Data Flow diagram

An actual Data Flow Diagram should be created at an early stage of the project to provide a visual representation of the data flows. It will help identify security risks and the detail of the Data Activity and Responsibility Matrix. The Data Flow diagram can be derived from a Technical/Solution Architecture Diagram. An example of the sort of diagram that might be produced is given at Figure 1.

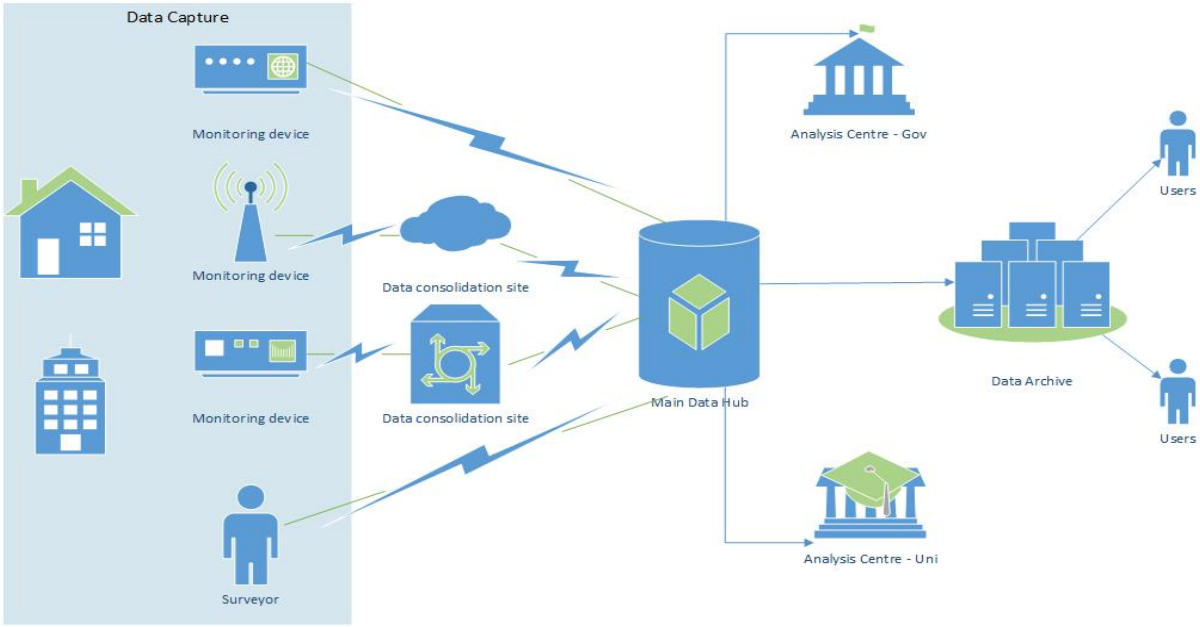

Figure 1: An example data flow diagram

[1] DCLG (2013) 'English housing survey: data security strategy and arrangements' – 10 July 2013. ISBN 9781409839286, URL https://www.gov.uk/government/publications/english-

g. Data Activity and Responsibility matrix

An actual Data Activity and Responsibility matrix should be created at an early stage of the project to document all data activities, the organization and individuals responsible for the activity and the processes/protocols that will to manage the data securely during each activity.

Data Activity	Organization Responsible	Details of Data Security
Householder survey	NatCen	Separation principle; laptops/CAPI devices have encrypted hard drives; Data uploaded to main data hub via secure comms link.
Physical survey	NatCen	Separation principle; laptops have encrypted hard drives; data uploaded to main data hub via secure comms link.
Monitoring Device X		Separation principle; data uploaded daily to data integration hub (cloud) via secure comms link. Data transferred to main data hub via secure comms link.
Monitoring Device Y		Separation principle; data stored on device and manually collected to laptops with encrypted hard drives; data uploaded to main data hub via secure comms link.
Data Analysis	UCL	Separation principle (except where geo-spatial identifiers are required for analysis purposes); Data will be stored on dedicated, secure UCL servers with access control to named individuals. Processing and analysis of data will take place on secure UCL application servers with access control to named individuals.
Data Archival	UK Data Archive	Anonymized data is provided to authorised researchers. Data with some identifiers may be available to researchers who comply with the strict conditions of a Secure Access License.

Table 1: Example Data Activity and Responsibility Matrix. Note: This is an example matrix only. The actual matrix is likely to be considerably more detailed as it should contain details of all relevant sub-activities

h. Data Controllers and Data Processors

The responsibilities of the various organizations as data controllers and data processors as defined by the Data Protection Act will need to be determined.

- "**data controller**" means a person who (either alone or jointly or in common with other persons) determines the purposes for which and the manner in which any personal data are, or are to be processed
- "**data processor**", in relation to personal data, means any person (other than an employee of the data controller) who processes the data on behalf of the data controller.

See the ICO's publication "Data controllers and data processors: what the difference is and what the governance implications are" (available here[2]) for further information on this issue.

In the case of the EHS, for example, DECC and DCLG were data controllers while NatCen were data processors. The UK Energy Lab may define the responsibilities differently depending on the governance framework and leadership structure for the long-term delivery of the survey.

i. Informed Consent

It is notably difficult to strike the right balance between providing enough detail and overwhelming a respondent when it comes to informed consent and data privacy notifications.

For this study, respondents should be clearly informed that:

- Their data will be handled confidentially under the terms of the DPA.
- Uses of data (including potentially linking to other data sources) that will be allowed.
- Uses of data that will not be allowed e.g. commercial direct marketing.

j. Data Privacy Notices – Code of Practice

The Data Privacy notice should conform to the Code of Practice[3] published by the Information Commissioner's Office. The use of a layered approach with additional information available on a website, alongside other relevant additional material for respondents is strongly recommended.

[2]http://ico.org.uk/for_organisations/guidance_index/~/media/documents/library/Data_Protection/Detailed_specialist_guides/data-controllers-and-data-processors-dp-guidance.pdf
[3]http://ico.org.uk/for_organisations/guidance_index/~/media/documents/library/Data_Protection/Detailed_specialist_guides/PRIVACY_NOTICES_COP_FINAL.ashx

k. Sample Data Privacy Notification

Below we present an initial notification, that would need to be expanded, containing the type of statements a data privacy notification should include. It is expected that significant attention will be required in the design and pilot stages before the wording is finalised.

- We will treat any information you provide in the strictest confidence under the Terms of the Data Protection Act 1998.
- The results collected are used for research purposes only and no one looking at the findings will be able to identify you in any way.
- Your information will be used by the Department of Energy and Climate Change (DECC) or other government departments for the production of statistics; and by authorized recipients such as universities and Non-Governmental Organisations for research and analysis purposes. All outputs will be aggregated or anonymized to ensure your identity cannot be revealed.
- Data from this survey may be linked to other datasets to aid research and the production of statistics.
- The data you provide will NOT be used for commercial activities such as direct marketing.
- Once initial processing and analysis have been completed, your data will be transferred for long-term stewardship to a secure national repository such as the UK Data Archive.

3. Identification of ethical research in domestic buildings

a. Overview

Three major areas of ethical issues will need consideration, focused on participants, on fieldworkers, and on detection of a hazard in a home (in addition to issues around data protection which are considered separately). The exact nature of issues will depend highly on what is asked, who is asked, and what data are collected.

There will be some sensitive questions such as around income or health. It would be advisable to allow refusal to a question but to minimize this from happening by emphasising anonymity and confidentiality of all data collected.

It will need to be decided if to include vulnerable adults and children in the research. If yes, additional information material would have to be generated, e.g. suitable for children. Also, it will have to be decided whether to provide information and translators for data collection in households where English is only poorly spoken and understood. In order to be inclusive, it is highly recommended to do so, and have a boost sample to ensure enough participants of foreign backgrounds can participate. Depending on what data is collected from sensors or technical kit, a number of issues might arise. In general, using prototypes as opposed to off-the-shelf-risk would be riskier and need additional testing. It needs to be ensured that no physical hazard can result from any equipment. There is also the risk that participants will have concerns, such as the data revealing when they are in the house with an increased chance of burglary. Hence, data security needs to be ensured and emphasized to participants. On-site data storage might be preferable under this point. Also, rules for compensation need to be in place, in case any damage to the property happens, and if equipment is employed that uses electricity supplied to the house.

Regarding fieldworkers main issues are safety, given that homes of strangers are visited, and regarding the installation of kit in the home. Adequate training on safety issues and on kit installation needs to be provided.

A final set of issues concerns what to do when detecting a hazard in a building, such as structural damage, or neglect. When detecting a hazard that home occupier needs to be informed and advised on how to proceed. Regarding detection of other issues such as neglect, it is recommended to report to an Ethics board that then considers each individual case.

b. Ethical issues focused on participants

Issues around question content

Subjects might find some questions too personal, sensitive, or embarrassing to answer, for example regarding income, or health. Giving participants the right to refuse an answer to a specific question without giving reasons for refusal would be an obvious solution even though it would create missing data. By emphasising to participants confidentiality and anonymity of data, the risk of refusing to answer might be mitigated.

Issues around participant type

When carrying out research with children / underage participants and / or vulnerable results, a DBS check is needed for all fieldworkers. Vulnerable adults are defined as "any person who lacks the absolute most basic (as distinct from mid-level or typical level) human life skills" and more precisely as someone "living in sheltered accommodation, … receives domiciliary care, …, is detained in lawful custody,…" (excerpt). The process of obtaining consent is potentially more difficult if the person in question does not understand the forms; the legal guardians would have to be involved, if the vulnerable adult has one. Given these complications, and also the fact that vulnerable adults will often not pay their own bills and manage their own households, it might be an option to exclude them from the study.

Depending on age of the child, informed consent is necessary from the child or from the parents / legal guardian. Research with children always needs to be in the best interest of the children. Also, all procedures (anonymity, confidentiality, consent, data collection etc.) need to be communicated to the child in a way appropriate for the age. Hence, if aiming at carrying out research with young children, this would involve significant investment in developing additional material, e.g. of pictorial nature.

The research committee will need to make a decision on whether to provide documents in multiple languages, and fieldworkers who speak multiple languages, and / or that translators or whether take the risk of some participants not understanding the questions and hence answering incorrectly, or refusing of grounds of language. It is recommended to ensure that non-English speakers can participate in the research and hence to provide translators and translated documents.

Recruitment & consent issues

It needs to be ensured that participation is voluntary and that participants are aware of their right to withdraw at any time from the study. It needs to be ensured that participants fully understand the study and what is involved, and give informed, written consent to the study.

Effects on the participant's environment

Depending on what sensors / kit will be installed in homes, additional issues might arise around participants' convenience. There might be a physical change to their environment – if using some obtrusive kit, and there is also the possibility that equipment falls off, hitting someone, or creating a risk to stumble. It will need to be ensured that stringent safety procedures are followed, such as installing equipment in such a height it cannot be a trip hazard and cannot be pulled down by toddlers, and ensuring equipment does not bring any fire hazard.

More likely than experiencing physical constrains or harm, participants might experience emotional reaction, such as feeling they are being spied on, or worrying that the sensors might indicate when they are not at home, which in the wrong hands might increase the chances of burglary. There is potentially a non-deniable actual risk, and it will need to ensured that all equipment is tried and tested, securely

implemented in the house, and that mechanisms are in place to ensure data cannot get into wrong hands. How this will be done exactly, will depend on the equipment being used. Regarding data protection, risks will be lower if sensors are used that store data locally instead of for example, transmitting information wirelessly, and all data need to be encrypted.

Giving participants access to their own data
To what extent participants will get access to their own data, needs to be decided once it is decided what data will be collected. It is envisioned that all participants will receive generic updates on the project to keep up their interest. Preferably, no individual access to data would be granted, similarly to procedures in English Housing Survey, as it would be a financial burden to the project and would potentially impact on results. If participants asked for access to data, ideally it would be granted on a yearly basis (to minimize impact on data) and in a summarized form.

Compensation for energy consumed through equipment
A procedure needs to be in place that ensures that householders are compensated for any damages that might occur in their properties, for example through the installation of sensors. In addition, if equipment is employed that uses mains electricity, participants need to be reimbursed for the amount used.

c. Ethical issues focused on fieldworkers

There will presumably be two types of fieldworkers, those carrying out interviews and surveys, and those installing any technological kit. Fieldworkers from NatCen all have a criminal background check; it needs to be ensured that also those installing technological kit (who might not come from NatCen) do have background checks given that there might be children present in the households or vulnerable adults. It needs to be ensured that there is a liability insurance in case fieldworkers cause damage in the property.

Personal safety
There is always a potential risk when fieldworkers go into a stranger's house. However, NatCen provides an extensive training in safety and has mechanisms in place to ensure fieldworker safety.

Accidents during installation of kit
It needs to be ensured that all fieldworkers are adequately trained in installing equipment, and have access to a first aid kit on-site.

d. Ethical issues when detecting a hazard in a building

Different scenarios can be envisioned that might happen, such as:

(1) Detection of a structurally unsafe building or other physical hazards.
(2) Detection of unhealthy living conditions such as extremely high humidity levels or very low temperatures
(3) Detection of welfare issues, such as domestic violence, extremely low hygienic standards, etc.

Example (1) and (3) will be most likely spotted by the fieldworker, in particular those installing kit. A procedure needs to be established on how to react in such a case. In general, the householders need to be informed, and it has to be ensured that the information is actually received, i.e. a letter would not do. It needs to be debated if fieldworkers (installing kit) would ever address an issue personally. It would be preferable not to do so because this would then raise more liability issues, e.g. when fixing something but it only holding temporarily.

Regarding Example (3), it is recommended to report such incidents to an Ethics board and consider them on a case-by-case basis. It is unlikely that (many) of those incidents are encountered, hence it is difficult to come up with a generic solution. It is noteworthy that NatCen employs a standard protocol for disclosure which states NatCen fieldworkers are required to disclose any information in relation to known criminal activity or potential harm to vulnerable individuals that are not able to defend themselves.

Assuming that temperatures and humidity are measured, one could develop a standard routine procedure to check the data for extreme cases, and use an automatic computer program to run those routines and flag up problems. Upon detection of an abnormality, those people identified for ethical issues would have to be notified.

4. Cost implications of ethical issues

There are several areas where ethics issues may incur additional expense to the project. These are outlined below. This does not include the costs incurred in the provision of monitoring equipment that is CEE certified and meets other product safety standards.

a. Translation of materials

Translation of materials into different languages and provision of short print-runs will incur costs to the project.

a. **Welsh language provision**. If the Energy-Lab is to include Welsh households in the sample and funded by a government department, it is a requirement of the Welsh government that data collection instruments are offered in both Welsh and English.
b. **Research with children**. This will need information material about the study, consent forms, and other participant contact information written and presented in a form suitable for children.
c. **Research with ethnic minorities**. This may require providing information in multiple languages and providing translators. Currently other surveys (such as the EHS) do not provide formal translation services for ethnic minorities, instead relying on English speaking friends and relatives. The provision of such services has research benefits but incur considerable costs.

b. Panel maintenance

This is the provision of aggregated data to participants on the trial as a whole for the purposes of maintaining interest and reducing attrition. This would usually be the same information provided to all participants and is thus a relatively low cost and necessary expense as part of the trial. It is anticipated that any such panel maintenance be provided by post to both provide the maximum accessibility to all participants and as this carries more weight with participants than email.

c. Participants accessing their own data

This is providing access to participants to their own data in a form they can make sense of (summarised into some kind of report). This could be done, say, annually in arrears and available on request. This has the potential to impact on the study through becoming a self-selective feedback intervention in its own right. A research decision would have to be taken with respect to: a) the level of detail of provision to participants' own data; b) the frequency of provision to participants' own data; c) how such feedback was managed (should it be provided to all participants if any request to avoid biasing the sample).

The costs of providing participants access to their data will depend on each of these points. From a research design position, the best option is likely to be control the manner and timing of feedback and to either restrict to as few as possible, or to make it part of panel maintenance and provide it to all so as to minimise the likelihood of biasing the sample.

The likelihood of requests for data will also vary depending on which tier of the sample (level 1, 2 or 3) the participant lies in. It is likely that those with the most intensive monitoring, in level 3, would be both most interested in the data and most aware of the monitoring process. In levels 1 and 2 the monitoring is likely to be light (if there at all) and broadly comparable to that in previous studies like the EHS and EFUS. In level 3 there is a particular need to avoid exacerbating any potential Hawthorn effects or other forms of panel conditioning though making any personal data feedback, say, annually in arrears.

At level 3, the costs of any such feedback are likely to be minimal in the context of the study as whole.

For the reasons of avoiding inadvertently creating an intervention study, it is recommended to limit feedback to an aggregated level on an annual basis done as part of panel maintenance.

It is thus not expected that significant additional expenses will be incurred from provision of personalised feedback.

d. Differential costs of handling personal vs. sensitive personal data

The issue of as what point in increasing the temporal or spatial density of sensing energy or environmental monitoring may move from being personal data to sensitive personal data was discussed. It was felt that there may be a point at which it is possible to infer a participant's religious belief through analysis of energy and environmental data. It was felt however that all data gathered in the project would need to be processed to industry standard security protocols including appropriate encryption at the hub level in single household dwellings and thus all energy and environmental data would effectively be being handled to the standard expected for sensitive personal data.

An issue was raised for data transmission from sensors to home hubs in multi-household dwellings. In this instance it was felt that encryption should be handled at the sensor level to ensure data security between households in multi-household dwellings. This may incur an additional cost for this small minority of dwellings.

e. Ethical issues associated with continuous environmental monitoring

The issue of the duty of care due to participants arising from continuous environmental monitoring was discussed and any financial implications that may arise from this. It was felt that while environmental variables would be monitored that could in principle detect adverse environmental conditions for occupants (e.g. fuel-poor people electively living at such cold temperature as to endanger their health) that this did not incur on the project a duty of care to continuously evaluate data-streams and intervene in such situations should they arise. While data would be collected continuously, and some syntactic analysis (algorithms looking for abnormal patterns in the data) would be run, these would be aimed at detecting faults in monitoring equipment. The project would need to be careful and explicit not to

create an expectation of providing continuous monitoring oversight of participants, to both avoid creating disincentives to participate, and to avoid any liabilities that may arise from any expectations of responsibility for detecting and intervening in adverse conditions.

f. Electromagnetic issues

Two issues were reviewed with respect to electromagnetism. Firstly, incidences of participants or participants' relatives claiming problems arising from electromagnetic sensitivity. It may be prudent to prepare an information leaflet on this issue in anticipation of this question arising and this would incur a minor cost. The second issue relates to claims of electromagnetic interference arising from monitoring equipment. It is anticipated that all monitoring equipment would need to conform to CEE standards and comply with wireless transmission protocols so interference would be kept to the minimum. In the event of problems arising, or being perceived to arise, a visit by a field officer would be required and this would incur a cost to the project. Data encryption for all data handled would be advisable but come at a cost.

5. Review of common ethical procedures

a. NatCen ethical procedures

Ethical codes

To maintain our position as an organisation that delivers high quality research it is essential that the work NatCen produces is designed and conducted in a way that meets certain ethical principles, and undergoes appropriate professional and organisational oversight in terms of research governance.

NatCen subscribes to the International Statistical Institute Declaration of Professional Ethics and the Social Research Association's Ethical Guidelines 2002. A range of other governance frameworks and guidelines are also identified as relevant for the work we do and inform the process. These typically include ensuring that studies also meet the ethical requirements of the Economic and Social Research Council, and Government Social Research. NatCen's internal procedures conform to both the above so no extra processes are required to do so. However, staff working on projects that have specific ethical guidelines are required to ensure that they are familiar with and are adhering to the funding body's ethical guidelines at all times.

All staff are required to read the guidelines periodically and abide by their principles at all times. Research staff sit on NatCen's ethical committees and are also often asked to defend the ethical approach of the studies they are designing in front of these committees.

Assessment of ethical issues

Every study must assess, consider and develop procedures to manage the ethical issues arising from a given research design. The Research Ethics Checklist (which acts as the internal REC application form) can be used to help identify issues that may raise ethical considerations or concerns. At a minimum, all studies should consider three core ethical principles:

- Informed Consent
- Confidentiality and anonymity
- Balancing burden (or harm) and benefit

In some studies, other ethical issues may be of consideration:

- Accessibility and vulnerability
- Disclosure of harm
- Data sharing and data linkage

Applications for ethical review are made after funding has been secured and fieldwork procedures (sampling and recruitment of participants, mode of data collection, general coverage of questions etc.) have been agreed (as far as possible). Survey documents (final questionnaires for example) do not need to be finalised before the application is made.

The application form (appended, as a worked example from NatCen's consumer research for ETI's SSH programme) is completed by the Research Project Manager and submitted to the REC administrator. The form dictates whether projects go through a stage 1 or stage 2 ethical review:

> **Stage 1 applications**: A stage 1 project has low levels of ethical risk. If none of the questions on the application form are answered "Yes" then the project only requires a Stage 1 review.
>
> Applications for stage 1 reviews can be submitted at any time and are considered by a REC Chair outside of the fixed REC meetings. Decisions will be given within two weeks of the form being submitted. If the decision is that a stage 2 review is required the Research Project Manager is responsible for amending and resubmitting the application form.
>
> **Stage 2 applications**: Applications for stage 2 reviews are submitted by given monthly deadlines) and are reviewed the following month. In the first instance, all stage 2 reviews are considered by the Chair and Deputy of that month's REC.
>
> - applications that have few major ethical issues are considered by the Chair and Deputy and a decision will be given within a month of the deadline for the submission.
> - applications that have more substantial ethical issues will be referred to the REC for full consideration and a decision will be given within a month of the deadline for the submission.

The REC is comprised of NatCen staff members and external experts and meets on a monthly basis.

b. Specific ethical issues for UK Energy-Lab

On the basis of this procedure, we would expect UK Energy-Lab to require a full Stage 2 application in front of an assembled REC committee. Below we have set out the ethical issues that we anticipate arising as part of this study, indicating any implications for cost or research design.

Informed consent

Informed consent is a fundamental element of any study and we have lots of experience in developing materials that enable participants to full engage with the nature of their participation and the aims and outputs of the research. Typically, informed consent is more challenging to secure if a) participants have limited capacity to provide it or b) aspects of the research are difficult to explain or are best withheld for methodological reasons.

In respect to a), we understand that the Energy Lab will be a general population study. We therefore anticipate that groups who may find it more difficult to give informed consent (for example, people with a disability or learning difficulties, older people, people with English as a second language) will inevitably be involved in the

research. This is standard practice for all of our general population surveys so we do not see it generating insurmountable ethical challenges. The exception would be if an ethnic boost is required, increasing the likelihood of translated materials being needed to ensure informed consent and data quality. On national, general population studies where we do not have an ethnic boost (such as EHS), we do not translate materials or the interview. In relation to the interview itself, for attitudinal data our approach is to aim to control the translation rather than allowing another in the household to translate (due to potential bias); on more factual or behavioural studies, it may be permissible for a family member to act as translator (this happens on the National Travel Survey). In cases where interviewers do not think this is appropriate, an appointment is made with a NatCen interviewer that speaks the relevant language.

If the Energy Lab is to include Welsh households in the sample and funded by a government department, it is a requirement of the Welsh government that data collection instruments are offered in both Welsh and English.

In respect to b), some of the elements of the study, particularly any technical monitoring, may be difficult to explain. Extra efforts will need to be made and extensive piloting take place in relation to clear articulation of the aims of this part of the study, what's involved for the participant and what happens to their data. A simple visual representation of data flows used in previous studies is often the most appropriate approach in this situation.

Further complications would arise if the Energy Lab required the direct participation of children under the age of 16. Additional steps and materials to ensure informed consent would be required.

- **Financial implications**: costs of translated materials is ethnic boost required

- **Methodological considerations**: need to be convinced of value of ethnic boost and involvement of children to off-set additional costs and risks that this generates

Confidentiality and anonymity

All NatCen studies have to ensure confidentiality and anonymity of participants: i.e. keeping secure the knowledge *that* people have taken part and *what* data they provided. NatCen is accredited to ISO27001 for information security and complies with the data protection in act in its collection, handling, storage and transfer of data. This aims to ensure individuals' participation is confidential. As standard for government and research council funded survey we also anonymise data sets by removing personal and identifiable data and replacing this with pre-set serial numbers.

In this study we do not anticipate that there are any circumstances where anonymisation of social survey data will not be possible. Standard survey practices

should ensure that the principles of confidentiality and anonymity are adhered to in relation to social survey data.

The physical energy monitoring proposed as part of Energy Lab may require additional safeguards to ensure anonymity and confidentiality. A potential concern of participants might be that the possibility of accessing real-time data or even retrospective data on energy use and possible occupancy could compromise the privacy of households. It is possible to get around these concerns by ensuring that:

- Any real time data is collected by a secure, encrypted process
- A time delay is put on data sharing from household to hub or hub to cloud
- Monitoring data is held separately to personal, address level details and only linked by pre-set serial number
- Access to data is limited to 'Data Controller' organisations that are ISO27001 compliant.

All of these safeguards are in place on the consumer research NatCen was involved in as part of ETI's SSH programme and do not present insurmountable ethical challenges.

- **Financial implications**: none additional to those associated with standard social surveys; potential implications in relation to secure servers and software for secure storage and transfer of monitoring data.

- **Methodological considerations**: need to be convinced of value of ethnic boost and involvement of children to off-set additional costs and risks that this generates

Balancing burden and benefit

Ultimately, deciding on the ethical status of any research study has to balance the burden or risk of harm to participants against the potential benefits to participants and society in general. The case for the benefits of the Energy Lab is being made elsewhere, but below we identify potential risks of harm and burden to both participants and fieldworkers:

- **Time and inconvenience**: the social survey will likely take up to an hour of people's time; and if this is a longitudinal study, this will be a repeat annual or bi-annual event. Agency fieldworkers should be well trained in minimising this burden and specific guidance can be provided at interviewer briefings on how to do so.
- **Inconvenience and risk**: additional inconvenience is likely to be experienced by the installation of energy monitoring equipment. Installers will need to think carefully about the possible risks associated with this and ensure that participants are fully informed about what participation in this element of the study involves. We would expect some attrition between the social and technical parts of the survey (around 80% of EHS participants in the social survey give consent to take part in the physical survey)

- **Providing energy data**: participants may view their energy data as a resource and want to receive a benefit out of the study from providing it. This may be in the form of a re-packaged summary of their energy use that helps that use energy more efficiently in the future. There is a legal requirement to provide data that participants have provided back to them; there is an ethical obligation to provide this in a way that is intelligible and potentially useful to the participant. However, the cost and resource involved in translating raw data into this format has to be balanced against this ethical obligation.

It is clear that the more sophisticated and detailed the technical monitoring, the more ethical issues are likely to come into play.

- **Financial implications**: re-packaging data for participants in a useful way may be labour intensive

- **Methodological considerations**: need to be convinced of value of each feature of the proposed monitoring study to minimise possible ethical challenges and minimise attrition.

c. Other ethical issues

Disclosure of harm

The standard protocol for disclosure is that NatCen fieldworkers are required to disclose any information in relation to known criminal activity or potential harm to vulnerable individuals that are not able to defend themselves. We do not anticipate any disclosure issues along these lines arising as part of the interview. As non-experts in energy efficiency or the physical properties of buildings, it is not necessary or advisable for fieldworkers to be assessing and responding to possible situations of harm that relate to structural safety of a building or high humidity levels for example.

Risk to fieldworkers

Fieldworkers always face possible risks when conducting household surveys. They are trained in relation to NatCen's working safely procedure (appended) to deal with these eventualities. Additional risk will arise if social survey interviewers, rather than technical specialists, are required to install monitoring equipment and a full risk assessment, pilot and health and safety protocol will need to be developed following full training of interviewers. A similar set of documents was developed as by NatCen's partners in the consumer research for ETI's SSH programme.

Data sharing and data linkage

It is likely that there will be methodological value from being able to link survey response and energy monitoring data to other data sources, such as the EPC register, NEED etc.. Informed consent will need to be obtained in order to do this, bearing in mind that experience suggest that the more sources surveys request to link to, the lower the consent rate.

- **Financial implications**: costs of Health and Safety consultants, protocols and assessments; costs of either training social survey interviewers or employing technical specialists to install monitoring equipment.

- **Methodological considerations**: need to be convinced of value of each feature of the proposed monitoring study to costs of installing safely; need to be convinced of value of each request to link data due to impact on response.

6. Information on Ethics Approval Procedure UCL

a. Summary

According to information from the Ethics Research Coordinator, there would be no need to obtain ethical approval from UCL if no UCL researchers are actually carrying out fieldwork but are only designing the study, analysing and publishing data. However, it needs to be ensured that suppliers of third party data (as NatCen would be considered to be) comply with the Data Protection Act 1998.

However, it is recommended to confirm this with the Ethics Research Coordinator before actually commencing the study because requirements might change and / or study design might change to include fieldwork carried out by UCL.

Also, given that many academic journals require ethical approval of studies, it needs to be ensured that NatCen has its own procedure for ethical approval and / or follows standard conventions, such as those of the Declaration of Helsinki, as provided by research councils UK , etc. The three most important ethical standards are:

(1) informed consent,
(2) benefit not harm,
(3) confidentiality.

b. General information about Ethics approval

All research should conform to commonly agreed standards of good practice, such as: those defined in the Declaration of Helsinki; and publications produced by (i) Research Councils UK (RCUK) and its constituent research councils, (ii) other funding bodies, for example, members of the Association of Medical Research Charities, (iii) and other national and international regulatory and governance bodies.
Three ethical standards need to be complied with:

Informed consent
All participants must be fully informed of the study and what is being asked of them, including the potential risks/benefits and exclusion criteria, in order to make a fully informed decision about whether or not to participate in the research. This must be an active step on behalf of the participant and not due to any inducement, coercion or perceived pressure to participate. This is required of all participants in a research study, except where there is a justification for covert research or deception (such cases will be considered on an individual basis by the ethics committee).

Benefit not harm
Research involving human participants must have a benefit to society and the risks involved to participants must be balanced against the potential benefit to the overall community.

Confidentiality
All participants have the right for their participation to remain confidential in that only the researcher will be aware who has participated. Generally all data will also be anonymous in the final report so that nothing can be attributed back to an individual participant. There are exceptions, for instance where participants wish to be identified or they cannot realistically have their identities kept confidential, but written informed consent must be obtained from the individual participant in advance.

Necessity of obtaining Ethics Approval for the Longitudinal Panel Study
Following discussion with the Ethics Research Coordinator, it might not be necessary to obtain ethical approval for the panel study if UCL researchers would not come into contact with participants but would merely design the study and analyse data. This was based on the assumption that names would not be made available but for example the postcode (first four entries) yes. UCL Ethics would look at this study as being led by NatCen, and as UCL obtaining third party data. Note that this would change if UCL employees would be actually involved in data collection. Also, UCL would need to ensure that the data supplier would comply with the Data Protection Act from 1988. The same would presumably be true for other Universities.

As suggested by the Ethics Research Coordinator, many journals insist on only publishing studies for which ethical approval had been obtained. Under this aspect, it would be worth ensuring that NatCen's ethical requirements are as strict as the ones commonly required by Universities, to enquire with journals if they would accept publications without ethical approval by the University, and / or to submit an application to Ethics after all (the latter option not suggested by the Ethics research coordinator).

For all data coming from third parties, it needs to be ensured that the provider complies with the Data Protection Act 1988; this would also be the case when for example obtaining data from energy suppliers.

General Procedure for obtaining Ethical Approval from UCL
The procedure for obtaining approval for a research project is given below. Further detail on this process is provided at Appendix A of this report.

(1) Register with the Ethics Data base & check application deadlines
(2) Register the project under Data Protection (duration for obtaining approval up to a week).
(3) Obtain Disclosure and Barring Service (DBS) Checks for the fieldworkers (up to one month).
(4) Fill in the Risk Assessment Form.

(5) Complete insurance form (if sample size > 500 subjects).
(6) Fill in and submit the actual ethics application; which includes different types of supporting documents such as recruitment forms, consent forms, information sheets.

7. Non-domestic considerations

a. Identification of ethical issues when researching commercial companies

The ethics surrounding energy surveys and suchlike, in the non-domestic sector, are concentrated upon the ethical behaviour of surveyors towards clients. It appears that the ownership of data lies with the client and that there is an assumption that information will not be shared with third parties. The possibility of a client wishing/allowing data to be used beyond the confines of the specific survey seems not to be considered. For this reason, no mention is made, for example in the RICS code of conduct, of data privacy issues. The same can be said of the BRE guidance for Energy Performance Certificate surveyors.

The lack of literature on the ethics surrounding non-domestic energy surveys etc. may be due to the historical lack of centralised, structured, standardised and regular collection of data on non-domestic buildings, or the energy use of occupiers. Most of the detailed data in the possession of the research community have come from ad hoc surveys of individual premises/buildings/organisations, in which it is presumed that specific permission to use the data for research has been gained from those being surveyed. With limited samples, the anonymisation of data is a particular problem, as with reduced numbers of data it becomes easier to identify participants. But, without sufficiently large samples, the data become far less useful if they are anonymised, due to the heterogeneity of the non-domestic sector.

With regard to the collection of participant energy billing data, the most sensitive piece of information is the price per unit paid by the consumer, particularly where it is a large consumer with significant bargaining power. Although price is a driver in the choice of supplier, it may be the case that this piece of information is not accessible and where it is accessible, full anonymisation of the dataset would be a condition of a participant taking part in the research. On the other hand, there may be other participants who will be quite willing for their data to be open to 'public' scrutiny, but these may be very few.

Possibly the most significant challenge with regard to anonymisation, for surveys of the non-domestic sector, is the very heterogeneity of non-domestic buildings and the activities they contain. This diversity means that it is likely to be easier to identify individual organisations from a fairly limited number of data. For example, the combination of business activity and four-character post code could identify individual organisations, as there may be very few instances of the activity in that post code. This may even happen at a regional or national scale, if the activity is extremely specialised. With some activities, such as office work or retail, this scenario is more likely to be rare, but in manufacturing it could be quite common. Non-domestic (non-residential) buildings/premises are likely to have a far more diverse range of physical hazards for field workers than domestic. There should be training of field workers, prior to the field work, of a nature suitable to ensure their safety under 'generic' conditions. An assumption of compliance with Health & Safety legislation by occupiers may be possible and reasonable. If this is the case, so long as the field workers are given adequate training/instruction/supervision by occupiers,

there should be no additional risks to the field workers under these controlled conditions. But, where field workers operate in ways that fall outside the normal operations of occupiers, hazards and subsequent risk will need to be assessed and managed, before field workers, or occupiers are exposed to the specific hazards. With this in mind, choice of methods and equipment will be important.
Training may take a modular form, with core training enhanced for specific environments according to the risks identified there. The nature of the modules is likely to change over time, even for ostensibly the same type of premises, due to field workers encountering many different types of equipment; this will especially be the case in industrial premises.

b. Giving participants access to their own data / obligation versus impact on study

Potentially, this could be a lot of data for each survey respondent, compared to domestic, dependent upon the complexity of the business/premises in question. For example, the data might include multiple energy use meters and perhaps even data from building management systems. However, these types of data would originally have been provided by the respondent, their energy supplier or building management contractor, so would not be 'new' data and could be made available in exactly the format in which they were supplied. In practice however, it would be more beneficial to provide them in a way that provides additional benefits, such as in a time series graph, comparisons with other similar settings and so on.

The main ethical issues surround data that have been collected by and specifically for the UK Energy Lab. Data might include the following types:

1. Photographs and other visual data
2. Inventories of equipment
3. Hours of occupation and employee numbers
4. Working practices, level of training and suchlike
5. Content of energy supply contracts, perhaps
6. Environmental conditions (temperatures etc.)

Some of these data types will prove more problematic to store and feed back to UKEL participants, as some data are unlikely to be supplied in a standardised format, even though they would be used in a structured manner during the research.
The feedback of data to non-domestic survey respondents is subject to a consideration that might not apply to the domestic sector. The respondent to the survey might not be a member of management and the revelation of survey responses that are truthful, but do not adhere to management practice, could place in jeopardy the employment of the individual respondent, should management become aware of the survey responses (i.e. through a request for stored data). Clearly, the individual respondent may simply state management policy in their survey responses, but where this is not what is actually occurring within the organisation/premises, the truth will be hidden from the survey. It seems unlikely that the survey respondent could be anonymised in such a way that they could be protected from any adverse effects of answering the survey truthfully, as management would be aware of the identity of the individual respondent.

Identification of issues related to data privacy & commercial confidentiality and data access where major energy companies are involved

Energy suppliers will generally quote commercial confidentiality as a reason for not releasing energy use data. The confidentiality is more of an issue for the supplier than it is for the consumer, as the release of data into the public domain allows competitor suppliers to discover the most lucrative accounts to service. However, this suggests that energy companies do not collect or store such data. This is counterintuitive, as it seems likely – especially with consumer ability/necessity to regularly change suppliers, or investigate such – that suppliers have a fair idea of the energy use of commercial organisations/premises.

A possible additional/alternative reason for suppliers to be unwilling to provide data is the work involved in the supply of such data. There may also be issues of not wishing to reveal the state of their own data acquisition, use and maintenance operations.

Where a survey respondent has given permission for their energy supplier to release data on their behalf to the survey, it is difficult to see where energy suppliers can raise objections, unless there is a contractual obligation for the survey respondent (consumer) not to disclose such information. This latter point could also apply where a respondent has supplied billing data to the survey, but seems unlikely.

Studies elsewhere

In the United States, the Commercial Buildings Energy Consumption Survey (CBECS) is carried out on a four year cycle, by the Energy Information Agency (EIA): it is one of the very few large-scale surveys of non-domestic building energy use. Although the survey is not longitudinal in nature, it carries out detailed surveys on businesses and their energy operations and buildings. This survey currently adheres to the requirements of the "Confidential Information Protection and Statistical Efficiency Act of 2002''. The most relevant sections are provided at **Appendix B** of this report. In summary, they set out conditions of confidentiality, meaning data may not be passed to other agencies.

Non-domestic sample as a subset of the domestic sample: small businesses

In the separate non-domestic report (project stream 8), it is suggested that the non-domestic element of the UK Energy Lab could take the form of a subset of the wider domestic sample, which would cover smaller businesses either operating from domestic premises, or where domestic and non-domestic share the same building, e.g. a dwelling over a shop, where there is no effective separation of the two.

The 'co-habitation' of domestic and non-domestic activities poses some ethical issues in that it becomes problematic to separate the two in terms of space use, economic activity and energy use and most importantly, where does the subject become an individual, as opposed to a business and vice versa?

In terms of anonymisation, it is much less likely that businesses operating from a domestic setting will appear as business hereditaments in the Valuation Office Agency (VOA) Rating List and thus one of the data linkages that allow the

identification of individual businesses would probably be removed. However, where the domestic element is essentially attached to the business non-domestic element, such as in VOA 'Composite' hereditaments, it becomes easier to identify the subject, but these 'composites' tend to occur in retail and workshop type hereditaments.

When visiting survey sites, the following need to be considered:

1) Always be accompanied by an authorised member of staff (unless other arrangements have been made)
2) Follow safety procedures as explained by authorised member of staff
3) Ask permission before taking photos in working areas with people and ask permission again if the photos are to be published (mostly used as a reminder of the use of spaces etc.)
4) Don't photograph any personal or commercially sensitive material
5) Try not to disturb people at work, unless both they and their employer are happy for this happen.
6) Be willing and able to explain what you are doing and avoid giving the impression that you are observing staff performance.
7) Gain permission before using measurement equipment (temperature sensors, daylight meters, etc.)

Guidelines / develop a procedure when detecting a health/ safety threat in a building

This only becomes an issue where field workers identify new hazards or hazards that are not being managed adequately by occupiers. At all times, the field worker should take responsibility for their own safety, but also be aware of the occupier's duty of care towards staff and visitors. Drawing the occupier's, or designated Health & Safety Representative person's, attention to the identified risk may result in immediate action to rectify the situation. Where this does not happen, the field worker must not place themselves, or others in a position of increased risk.

There is no statutory obligation, for someone outside of an organisation, to report health and safety issues to the Health and Safety Executive (or similar). Field workers have a duty to report potential serious risks affecting themselves (or others) that result from their work, to their employer or health and safety representative.

Estimate the cost impact of identified issues

Surveyor familiarisation with the health & safety features/constraints of each survey participant will consume a varying amount of time, depending upon the complexity and dangers of the survey environment. This will cause variability in the costs of visiting each type (and perhaps size) of survey participant premises.

Appendix A: UK Energy Lab - Data Management Plan Template

Project Details

Project ID	
Project Name	UK Energy Lab
Funder	*State research funder if relevant*
PI	*Name of Principal Investigator(s) or main researcher(s) on the project*
Project Data Manager(s)	*Name of person with primary responsibility for managing the project's data (if different to above), telephone and email contact details*

Project Description
Questions to consider: *- What is the nature of your research project?* *- What research questions are you addressing?* *- For what purpose are the data being collected or created?* *Guidance:* *Briefly summarise the type of study (or studies) to help others understand the purposes for which the data are being collected or created*
Enter details here:

Related Policies
Questions to consider: *- Are there any existing procedures that you will base your approach on?* *- Does your department/group have data management guidelines?* *- Does your institution have a data protection or security policy that you will follow?* *- Does your institution have a Research Data Management (RDM) policy?* *- Does your funder have a Research Data Management policy?* *- Are there any formal standards that you will adopt?* *- Are there any Data Sharing Agreements that apply to the data utilized by this project?* *Guidance:* *List any other relevant funder, institutional, departmental or group policies on data management, data sharing and data security. Some of the information you give in the remainder of the DMP will be determined by the content of other policies. If so, point/link to them here.*
Relevant UCL & EI Policies: • UCL Research Data Policy, available here • UCL Information Security Policy, available here • UCL Data Protection Policy, available here
Enter details here:

Data Management Plan – Version History

Version number	Date	Name	Summary of Amendments

Data Collection

What data will you collect or create?
Questions to consider: *- What type, format and volume of data?* *- Do your chosen formats and software enable sharing and long-term access to the data?* *- Are there any existing data that you can reuse?* *Guidance:* *Give a brief description of the data, including any existing data or third-party sources that will be used, in each case noting its content, type and coverage. Outline and justify your choice of format and consider the implications of data format and data volumes in terms of storage, backup and access*
Enter details here:

How will the data be collected or created?
Questions to Consider: *- What standards or methodologies will you use?* *- How will you structure and name your folders and files?* *- How will you handle versioning?* *- What quality assurance processes will you adopt?* *Guidance:* *Outline how the data will be collected/created and which community data standards (if any) will be used. Consider how the data will be organised during the project, mentioning for example naming conventions, version control and folder structures. Explain how the consistency and quality of data collection will be controlled and documented. This may include processes such as calibration, repeat samples or measurements, standardised data capture or recording, data entry validation, peer review of data or representation with controlled vocabularies*.
Enter details here:

Documentation and Metadata

What documentation and metadata will accompany the data?
Questions to consider: *- What information is needed for the data to be to be read and interpreted in the future?* *- How will you capture / create this documentation and metadata?* *- What metadata standards will you use and why?* *Guidance:* *Describe the types of documentation that will accompany the data to help secondary users to understand and reuse it. This should at least include basic details that will help people to find the data, including who created or contributed to the data, its title, date of creation and under what conditions it can be accessed.* *Documentation may also include details on the methodology used, analytical and procedural information, definitions of variables, vocabularies, units of measurement, any assumptions made, and the format and file type of the data. Consider how you will capture this information and where it will be recorded. Wherever possible you should identify and use existing community standards.*
Enter details here:

Ethics and Legal Compliance

How will you manage any ethical issues?
Questions to consider: *- Have you gained consent for data preservation and sharing?* *- How will you protect the identity of participants if required? e.g. via anonymisation* *- How will sensitive data be handled to ensure it is stored and transferred securely?* *Guidance:* *Ethical issues affect how you store data, who can see/use it and how long it is kept.* *Managing ethical concerns may include: anonymisation of data; referral to departmental or institutional ethics committees; and formal consent agreements. You should show that you are aware of any issues and have planned accordingly. If you are carrying out research involving human participants, you must also ensure that consent is requested to allow data to be shared and reused.*
Enter details here:

How will you manage copyright and Intellectual Property Rights (IPR) issues?
Questions to consider: *- Who owns the data?* *- How will the data be licensed for reuse?* *- Are there any restrictions on the reuse of third-party data?* *- Will data sharing be postponed / restricted e.g. to publish or seek patents?* *Guidance:* *State who will own the copyright and IPR of any data that you will collect or create, along with the licence(s) for its use and reuse. For multi-partner projects, IPR ownership may be worth covering in a consortium agreement. Consider any relevant funder, institutional, departmental or group policies on copyright or IPR. Also consider permissions to reuse third-party data and any restrictions needed on data sharing.*
Enter details here

Storage and Backup

How will the data be stored and backed up during the research?
Questions to consider: *- Do you have sufficient storage or will you need to include charges for additional services?* *- How will the data be backed up?* *- Who will be responsible for backup and recovery?* *- How will the data be recovered in the event of an incident?* *Guidance:* *State how often the data will be backed up and to which locations. How many copies are being made? Storing data on laptops, computer hard drives or external storage devices alone is very risky. The use of robust, managed storage provided by university IT teams is preferable. Similarly, it is normally better to use automatic backup services provided by IT Services than rely on manual processes. If you choose to use a third-party service, you should ensure that this does not conflict with any funder, institutional, departmental or group policies, for example in terms of the legal jurisdiction in which data are held or the protection of sensitive data.*
Enter details here:

How will you manage access and security?
Questions to consider: *- What are the risks to data security and how will these be managed?* *- How will you control access to keep the data secure?* *- How will you ensure that collaborators can access your data securely?* *- If creating or collecting data in the field how will you ensure its safe transfer into your main secured systems?* *Guidance:* *If your data is confidential (e.g. personal data not already in the public domain, confidential information or trade secrets), you should outline any appropriate security measures and note any formal standards that you will comply with e.g. ISO 27001.*
Enter details here:

Selection and Preservation

Which data should be retained, shared, and/or preserved?
Questions to consider: *- What data must be retained/destroyed for contractual, legal, or regulatory purposes?* *- How will you decide what other data to keep?* *- What are the foreseeable research uses for the data?* *- How long will the data be retained and preserved?* *- Guidance:* *Consider how the data may be reused e.g. to validate your research findings, conduct new studies, or for teaching. Decide which data to keep and for how long. This could be based on any obligations to retain certain data, the potential reuse value, what is economically viable to keep, and any additional effort required to prepare the data for data sharing and preservation. Remember to consider any additional effort required to*

prepare the data for sharing and preservation, such as changing file formats.
Enter details here:

What is the long-term preservation plan for the dataset?
Questions to consider: *- Where e.g. in which repository or archive will the data be held?* *- What costs if any will your selected data repository or archive charge?* *- Have you costed in time and effort to prepare the data for sharing / preservation?* *Guidance:* *Consider how datasets that have long-term value will be preserved and curated beyond the lifetime of the grant. Also outline the plans for preparing and documenting data for sharing and archiving. If you do not propose to use an established repository, the data management plan should demonstrate that resources and systems will be in place to enable the data to be curated effectively beyond the lifetime of the grant.*
Enter details here:

Data Sharing

How will you share the data?
Questions to consider: *- How will potential users find out about your data?* *- With whom will you share the data, and under what conditions?* *- Will you share data via a repository, handle requests directly or use another mechanism?* *- When will you make the data available?* *- Will you pursue getting a persistent identifier for your data?* *Guidance:* *Consider where, how, and to whom data with acknowledged long-term value should be made available. The methods used to share data will be dependent on a number of factors such as the type, size, complexity and sensitivity of data. If possible, mention earlier examples to show a track record of effective data sharing. Consider how people might acknowledge the reuse of your data.*
Enter details here:

Are any restrictions on data sharing required
Questions to consider: *- What action will you take to overcome or minimise restrictions?* *- For how long do you need exclusive use of the data and why?* *- Will a data sharing agreement (or equivalent) be required?* *Guidance:* *Outline any expected difficulties in sharing data with acknowledged long-term value, along with causes and possible measures to overcome these. Restrictions may be due to confidentiality, lack of consent agreements or IPR, for example. Consider whether a nondisclosure agreement would give sufficient protection for confidential data.*
Enter details here:

Responsibilities and Resources

Who will be responsible for data management?
Questions to consider: *- Who is responsible for implementing the DMP, and ensuring it is reviewed and revised?* *- Who will be responsible for each data management activity?* *- How will responsibilities be split across partner sites in collaborative research projects?* *- Will data ownership and responsibilities for RDM be part of any consortium agreement or contract agreed between partners?* *Guidance:* *Outline the roles and responsibilities for all activities e.g. data capture, metadata production, data quality, storage and backup, data archiving & data sharing. Consider who will be responsible for ensuring relevant policies will be respected. Individuals should be named where possible.*
Enter details here:

What resources will you require to deliver your plan?
Questions to consider: *- Is additional specialist expertise (or training for existing staff) required?* *- Do you require hardware or software which is additional or exceptional to existing institutional provision?* *- Will charges be applied by data repositories?* *Guidance:* *Carefully consider any resources needed to deliver the plan, e.g. software, hardware, technical expertise, etc. Where dedicated resources are needed, these should be outlined and justified.*
Enter details here:

Acknowledgements

This template draws heavily on:

DCC. (2013). *Checklist for a Data Management Plan.* v.4.0. Edinburgh: Digital Curation Centre. Available online: http://www.dcc.ac.uk/resources/data-management-plans

Appendix B: Details for the individual steps in the UCL ethics procedure

1. Registration
    - on the website http://ethics.grad.ucl.ac.uk/new_user.php if the PI is not registered yet
2. Data Protection
    - The Data Protection Act protects information (personal data) about identifiable individuals ("data subjects") and enforces a set of standards for the processing of such information. Any member of staff or student of UCL using personal data for whatever purpose must comply with the provisions of the Act.
    - There are eight principles of the Act which in outline these state that personal data shall:
        i. *be collected and processed fairly and lawfully*
        ii. *only be held for specified lawful purposes*
        iii. *only be adequate and relevant and not excessive in relation to its purposes*
        iv. *be accurate and, where necessary, be kept up to date*
        v. *be held no longer than is necessary for the stated purpose*
        vi. *be held and used in accordance with the rights of the data subjects under the Act*
        vii. *have appropriate security surrounding it*
        viii. *shall not be transferred outside the European Economic Area unless the country to which it is being transferred ensures adequate protection for the rights of data subjects.*
    - "One point stressed is that it is rarely necessary to keep electronic personal data on portable devices such as laptops, USB flash drives, portable hard drives, CDs, DVDs, or computers *not owned by UCL*." However, this will not be possible for this study; and hence needs to be addressed.
    - "To minimise the risk of loss or unauthorised disclosure, a secure remote connection to UCL or virtual private network should be used wherever possible. [….] No data should be held on PC hard drives. […]If storing sensitive data on portable devices or media all data must be strongly encrypted." This needs to be considered and addressed.
    - "Paper records containing personal data should be locked away when not in use. They should not be regularly removed from UCL premises." Again, this is likely not possible and hence needs to be addressed in detail.
    - detailed information is available here: http://www.ucl.ac.uk/finance/legal_services/data_protection/dp_research.html
    - related file attached:
3. DBS checks
    - will be necessary if vulnerable adults and children are involved
    - UCL, in accordance with DBS guidelines, does not accept portability of DBS checks which UCL staff or students may have from previous organisations as proof of satisfactory clearance.

- Checks can be arranged through the Department.
- Cost £44 per person.
- More details under http://www.ucl.ac.uk/current-students/services_2/dbs_checks

4. Risk Assessment Form.
    - for all fieldwork, a risk assessment form needs to be filled in.
5. Insurance.
    - when an intervention study (even non-clinical) has more than 500 subjects; an insurance form has to be filled in.
        i. Intervention can be "a device". Depending on what is tested, we might need to fill in the insurance form.
    - Policy contains cover for negligence of UCL or UCL employees.
    - Note that this would not cover fieldworkers given that they are not employed by UCL.
6. Actual Ethics application
    - structure of the application
        i. Section A: Administrative part
        ii. Section B: Details about the project
            1. Summary of project
            2. Methodology used
            3. Location
            4. Dissemination of results
            5. Identification of ethical issues
                a. This is an important section in which it needs to be discussed what happens when adverse conditions are found in homes such as instable structures or unhealthy living conditions
        iii. Section C: Details about participants
            1. eg. age, number of subjects, way of recruitment, ways of obtaining consent etc.
            2. a number of supporting documents are needed such as consent forms and information sheets

Special issues that need to be considered:

- UCL assumes that normally data is stored on UCL machines / servers
- DBS checks are only accepted when obtained by UCL
- If using data or information held by a third part, it needs to be explained in the Ethics application how data will be obtained and confirmed that the information has been obtained in accordance with the UK Data Protection Act 1998. This would be relevant when using information from suppliers or other surveys.

Appendix C: Extract of regulations covering data collection for the US energy survey of commercial settings (CBECS)

“SEC. 512. LIMITATIONS ON USE AND DISCLOSURE OF DATA AND INFORMATION.

(a) USE OF STATISTICAL DATA OR INFORMATION.—Data or information acquired by an agency under a pledge of confidentiality and for exclusively statistical purposes shall be used by officers, employees, or agents of the agency exclusively for statistical purposes.

(b) DISCLOSURE OF STATISTICAL DATA OR INFORMATION.—

(1) Data or information acquired by an agency under a pledge of confidentiality for exclusively statistical purposes shall not be disclosed by an agency in identifiable form, for any use other than an exclusively statistical purpose, except with the informed consent of the respondent.

(2) A disclosure pursuant to paragraph (1) is authorized only when the head of the agency approves such disclosure and the disclosure is not prohibited by any other law.

(3) This section does not restrict or diminish any confidentiality protections in law that otherwise apply to data or information acquired by an agency under a pledge of confidentiality for exclusively statistical purposes.

(c) RULE FOR USE OF DATA OR INFORMATION FOR NONSTATISTICAL

PURPOSES.—A statistical agency or unit shall clearly distinguish any data or information it collects for nonstatistical purposes (as authorized by law) and provide notice to the public, before the data or information is collected, that the data or information could be used for nonstatistical purposes.

(d) DESIGNATION OF AGENTS.—A statistical agency or unit may designate agents, by contract or by entering into a special agreement containing the provisions required under section 502(2) for treatment as an agent under that section, who may perform exclusively statistical activities, subject to the limitations and penalties described in this title.”

Note that:

“SEC. 502. DEFINITIONS. …

(5) The term ‘‘nonstatistical purpose’’—

(A) means the use of data in identifiable form for any purpose that is not a statistical purpose, including any administrative, regulatory, law enforcement, adjudicatory, or other purpose that affects the rights, privileges, or benefits of a particular identifiable respondent; and

(B) includes the disclosure under section 552 of title 5, United States Code (popularly known as the Freedom of Information Act) of data that are acquired for exclusively statistical purposes under a pledge of confidentiality.”