Sustainable Development Foundation

Report for the Green Construction Board

GCB Project 430 Knowledge Capture and Dissemination

Prepared by: Laura Morgan Forster, Sustainable Development Foundation
Tom Randall, Sustainable Development Foundation
David Churcher, Hitherwood Consulting

Date: 10th June 2014

Table of Contents

Figures

Tables

1 Acknowledgements

This report has been prepared under contract to the Construction Industry Council on behalf of the Green Construction Board Knowledge and Skills Working Group.

The project was overseen by a Steering Group, drawn from the GCB Working Group, comprising:

- Robert Lambe Willmott Dixon
- Malcolm Bell Leeds Metropolitan University
- Liz Reason Green Gauge Trust
- Lynne Sullivan Sustainable by Design.

A Peer Review Group was also established alongside the research team to provide specific expertise and their contribution is gratefully acknowledged. The Peer Review Group comprised:

- Jon Bootland Sustainable Development Foundation
- Roderic Bunn BSRIA
- Simon Elam UCL Centre for Energy Epidemiology
- Paul Ruyssevelt UCL Centre for Energy Epidemiology.

2 Introduction

2.1 Background

This report sets out work commissioned by the Green Construction Board to identify:

- **Currently underutilised data** that could contribute towards a move towards more sustainable construction practices.
- **Actions the Green Construction Board could take** to improve the effective use of appropriate data to encourage more sustainable construction practices.

The tender document stated:

> *'A significant issue for the UK construction & property sectors is that there is a vast array of information and data available that is difficult to access and challenging to differentiate what is of good quality and what is not. Therefore, it cannot be used to full effect to share intelligence and learning to improve knowledge and best practice across the sector.'*

2.2 The Brief

Specific requirements were to:

1. Establish the **current position** of sustainability data capture within the UK construction & property sectors in terms of, availability and accessibility.
2. Produce an overview of what data is **accessible**, publicly and commercially.
3. Review how **quality** of data is evaluated and indicated
4. Identify what/where substantive data exists that is not being used to its' **potential**
5. Identify the key **challenges** and **barriers**
6. Highlight where **others** have, or are also addressing the issue
7. Highlight **best practice** data capture and dissemination
8. Quantify the **opportunity** and **benefits**

2.3 Information Gathering

This report sets out the work undertaken to address the Green Construction Boards brief. The following approach was taken:

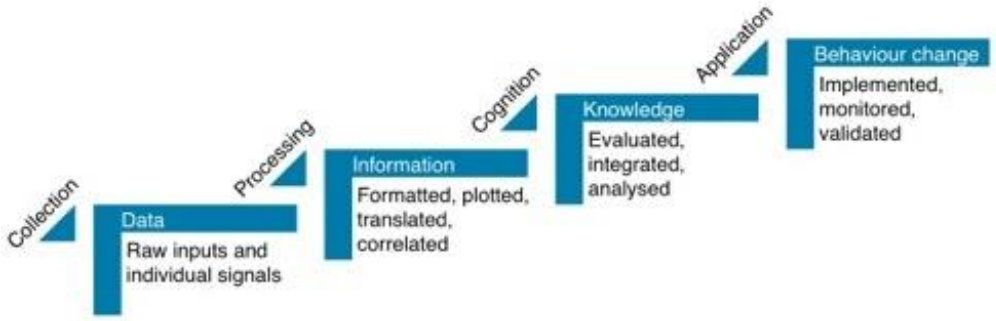
1. **Sector wide survey. To establish the current position of data capture in the UK and obtain some initial insights into accessibility, availability, quality and potential** a survey was issued through the contacts of the Green Construction Board and Sustainable Development Foundation.
   This survey asked respondents to identify potentially relevant data sets they know about and if possible provide information on:
   a. **The type of data** including the issues covered and whether the data is quantitiative or qualitative.
   b. **Building types** covered by the data. Including Domestic, non-domestic and non-domestic sub categories.
   c. **Life cycle stage** covered: planning, design, construction, handover, operation, end of life)
   d. **The quality, accessibility and coverage of the data** including ownership, accessibility, geographic coverage, quality, transparency, use and usefulness.
2. **Targeted Survey**. **To obtain more detailed insights into the quality, usefulness and application** of existing data a 2nd more detailed survey was sent to respondents of the first survey. The survey included asking respondents to score data sets in terms of their quality and usefulness.

2.4 Analysis

The results of the survey were then analysed.

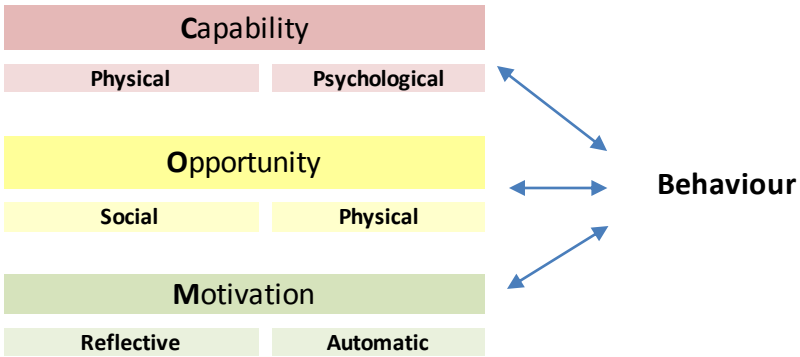
- Analysis of **gaps and general characteristics** of existing data sets was undertaken by plotting various combinations of data set characteristics.
- Detailed analysis of **Quality and Usefulness** was undertaken using the scoring from the second survey.
- The **potential** of data sets was analysed, framed using the following representation

'The Knowledge Ladder' Progression from data collection to behaviour change:



- **Interventions** to affect behaviour change were investigated using an established academic framework of behaviour change and comprehensive list of generic intervention types. These were applied to the sustainable construction areas of interest, identified barriers, and example data sets.

COM-B Behaviour Change Framework:



Generic Intervention Types:

1. Education
2. Persuasion
3. Incentivisation
4. Coercion
5. Training
6. Restriction
7. Environmental restructuring
8. Modelling
9. Enablement

3 Current position of sources of data and information

3.1 Overview

The existence of data and information regarding the selected sustainability indicators was revealed through three parallel activities:

1. **An online survey** (termed 'Survey 1') circulated and forwarded through a list of 100 contacts
2. **Direct discussions** with the Centre for Energy Epidemiology (CEE) at UCL
3. **Supplementary desk research**.

This discovery phase of the research took place between mid-December 2013 and mid-January 2014, although some of the supplementary desk research extended through to early March 2014.

The research resulted in a list of 178 data and information sources but of these 17 were not directly related to sustainability in buildings so were discounted from the remaining analysis. This left a group of 161 data sets – see Appendix C – and these were categorized in a number of different ways:

- By **sustainability** theme
- By **accessibility**
- By the **geographical coverage** of the data or information
- By the **type(s) of building** covered by the data or information
- By the **asset life cycle stage**(s) covered by the data or information.

These categorizations were supplied by the survey respondents, from the information recorded by CEE against each of the data sets they had described, or from a brief examination of the data set. Information supplied by respondents or CEE was taken on good faith and not necessarily checked by the research team due to time constraints.

In addition to these basic characteristics of the data and information sources, some of the data sets were also assessed regarding their quality, usage, current usefulness and potential usefulness. The results of these more detailed assessments are presented in Section 5.

3.2 Results of Survey 1

The online survey obtained responses from 16 individuals, and between them they contributed 42 data and information sources. There was very little duplication in the responses provided. The data sets identified through Survey 1 are those with a Survey Monkey UID (MS1 to MS42) in Table C.1. There were a couple of data sets that were suggested by more than one respondent, but these duplications have been removed from the list.

3.3 Results from Centre for Energy Epidemiology (CCE) at UCL

CEE passed across their list of data sources that had been researched by the team at UCL and co-ordinated by Simon Elam. This contained a total of 115 data sources, but 17 of these were excluded from our study as they were not immediately concerned with buildings. The remaining 98 are listed in Table C.1 with a UCL UID. There was no duplication between the list of data sources obtained from UCL and those obtained from the respondents to Survey 1.

The 17 data sets identified by CEE that were not directly relevant are listed in Table C.2.

3.4 Results from supplementary desk research

An initial review of the lists of data and information sources obtained from the Survey 1 and from CEE indicated some known gaps. These were followed up by the research team and added a further 21 data sources, some of which were added by members of the GCB Working Group.

3.5 Observations on the lists of data and information sources collected

The collective list of data and information sources cannot be considered as comprehensive, although it is hoped that the research has identified the large majority of publicly-accessible data sets. As will be seen in Section 4, some of the data sources identified during this work are privately held. As the research has not been able to ask every potential data holder whether or not they have their own data sets, it is not possible to state an accurate upper limit for the number of privately-held data sets.

4 Gap analysis of data and information sources

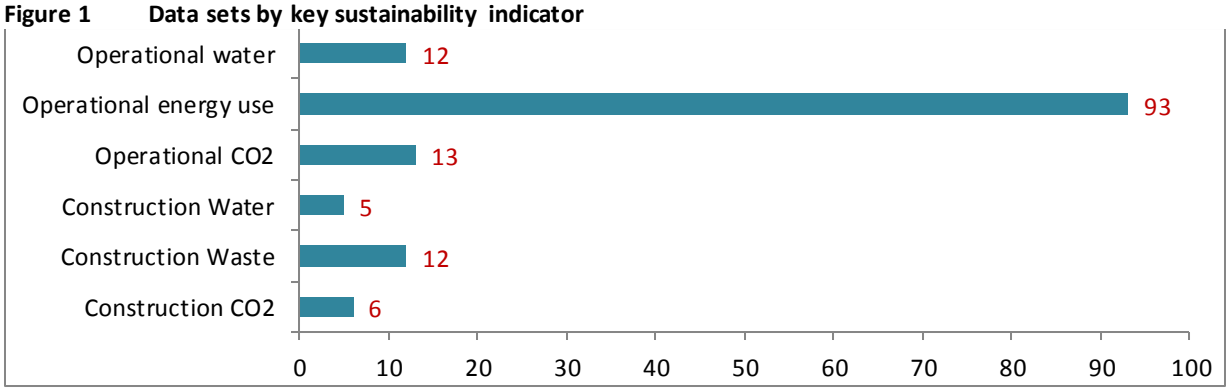
The information concerning data sources was analyzed to highlight gaps in the overall data landscape. Each of the five data set characteristics listed in Section 3.1 was analyzed on its own to describe different aspects of the data landscape. The results are shown in Figures 1 to 7.

4.1 Data across the selected sustainability indicators

GCB had identified the specific sustainability indicators of operational energy, and carbon dioxide emissions and water use during both construction and operation as of particular interest in this project. To this was also added construction waste to make six key sustainability indicators

Figure 1 shows the number of data sets covering each of the key indicators. It is clear that the large majority of data sets are concerned with energy consumption of during the use of buildings.

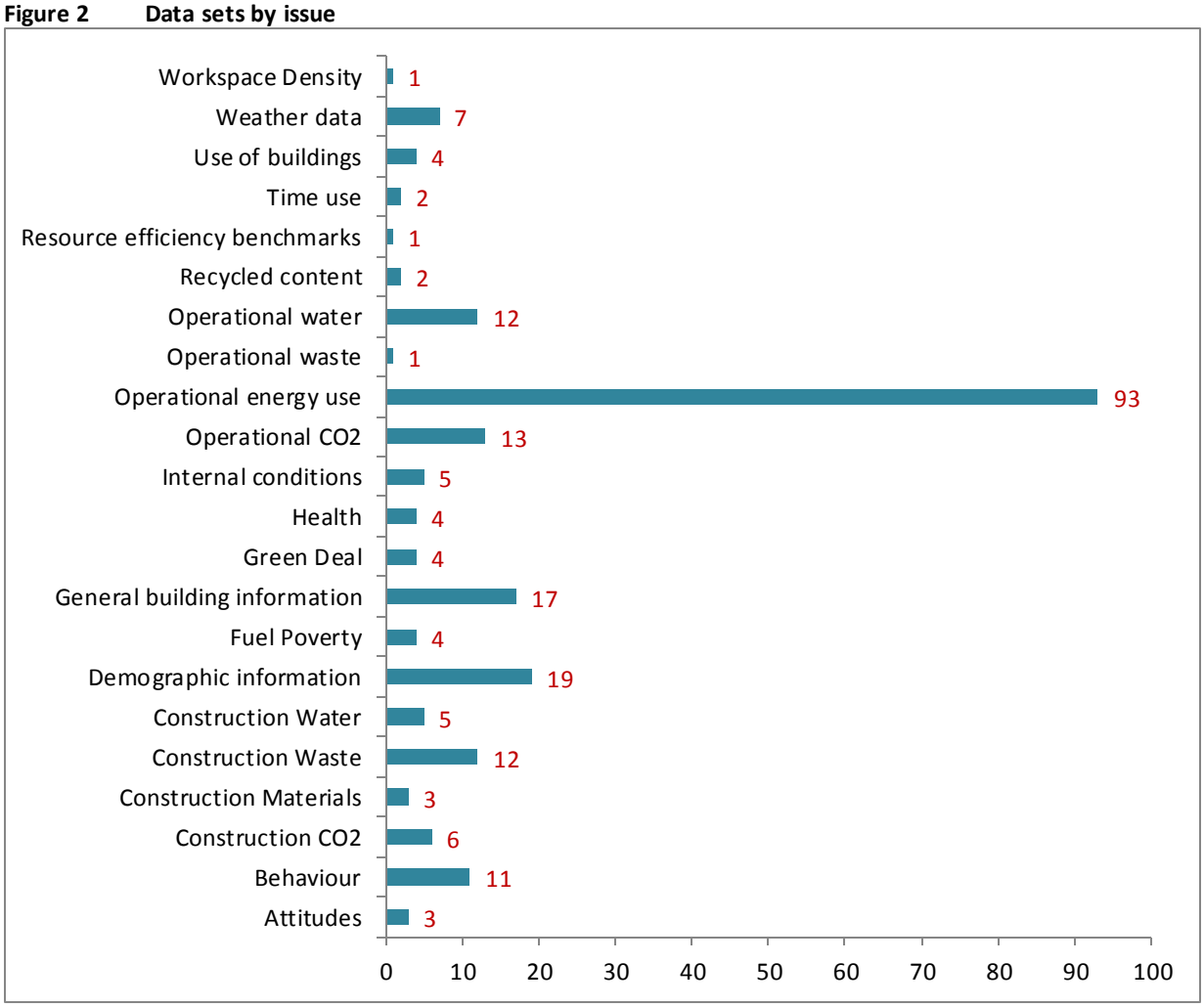
Figure 1 Data sets by key sustainability indicator



Operational water, carbon dioxide emissions and construction waste all have a modest number of data sets. Construction water use and carbon dioxide emissions both have a small number of data sets.

Other issues judged of potential relevance to the research project were also identified. These included issues that may provide context or be used in combination with the initial six issues identified to provide useful knowledge. Some of these other issues had just as many data sets associated with them as the less well represented key indicators. This further emphasised the dominance of operational energy use.

Figure 2 Data sets by issue



4.2 Accessibility of data

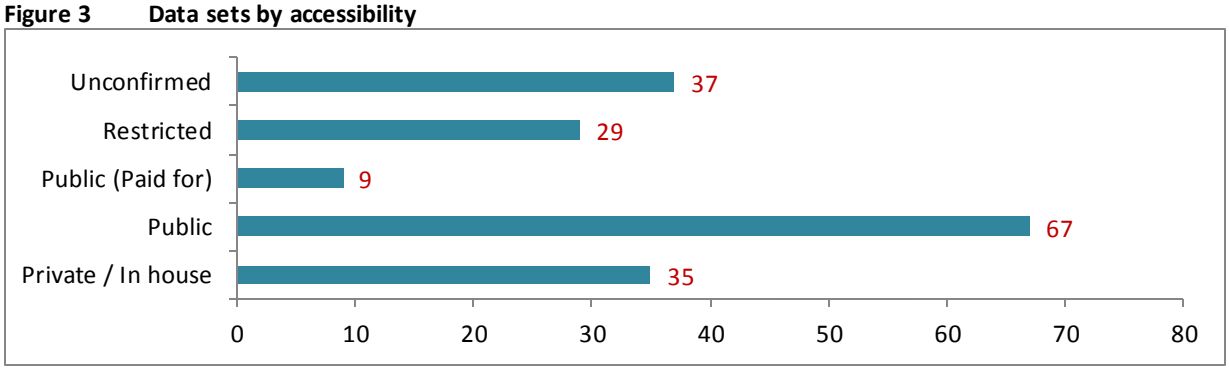
Access to each data or information source was sub-categorized in one of four ways and referred to in our analysis on a scale of 1-4 with an additional 'Unconfirmed' category:

1. **Private** access
2. **Restricted** access
3. **Public** access, **paid-for**
4. **Public** access, **free**

- **Unconfirmed**

Figure 3 shows the number of data sets falling into each of these sub-categories:

Figure 3 Data sets by accessibility



Note: Some data sets classified as 'Public' may also be paid for but this has not been confirmed

There is a fairly even split between data sets that are fully or partially restricted and data sets that have public access. The majority of the public access data sets were submitted by UCL whilst the majority of data sets submitted via the survey were private. It may be that this reflects the characteristics of the organisations submitting the data, with many of the respondents to the survey being from private companies including several contractors.

The most common class of restriction for users is to distinguish between research/policy users and commercial users. Some data sources are accessible to commercial users but only on payment of relatively high charges. Other sources are not accessible to commercial users at any price.

The sub-set of privately held data sources is a sizeable minority of the data landscape, and it is very likely that there are many more privately-held data sources that have not been discovered by the research team. However, extensive searches and enquiries were not possible within the scope of this project.

4.3 Geographical coverage of data sets

The geographical coverage of the data and information sources discovered during the research was sub-categorized as:

- **Continent**
- **Country**
- **UK Nation** (England, Scotland, Wales, Northern Ireland)
- **City or town**

Figure 4 shows the breakdown of data sets by continent and country.

Figure 4 Country / Continent

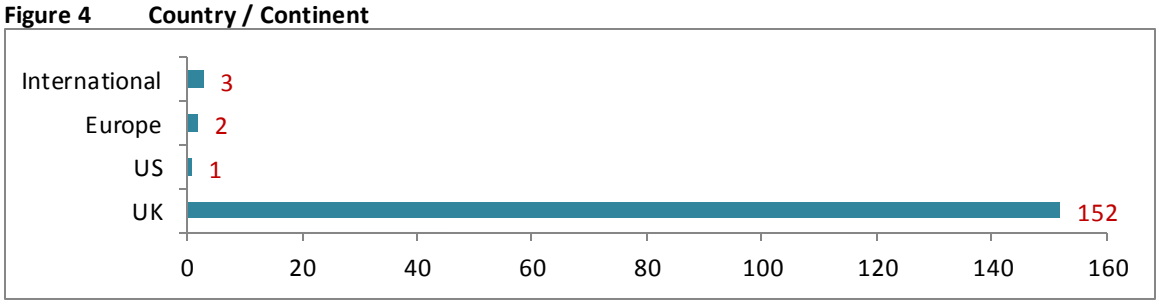


Figure 5 shows the breakdown by UK nation where explicitly stated. (Other data sets may also cover specific nations but this was not stated in the details provided).

Figure 5 UK Nation

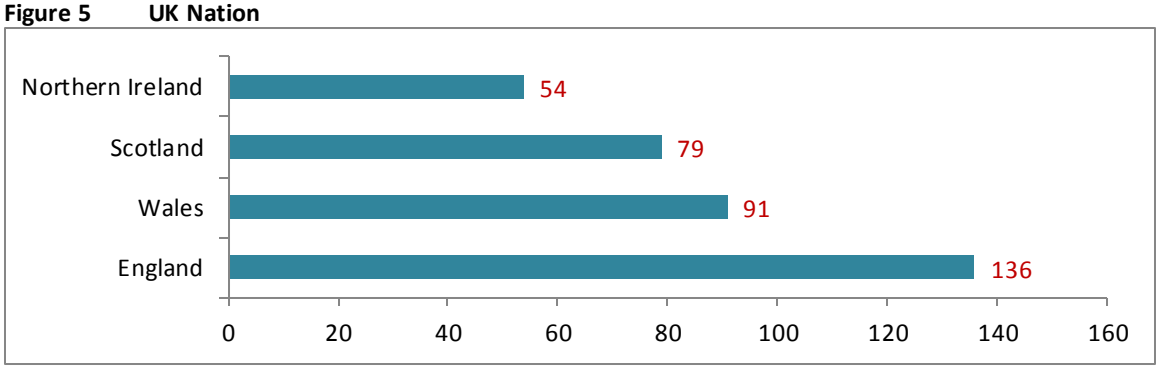
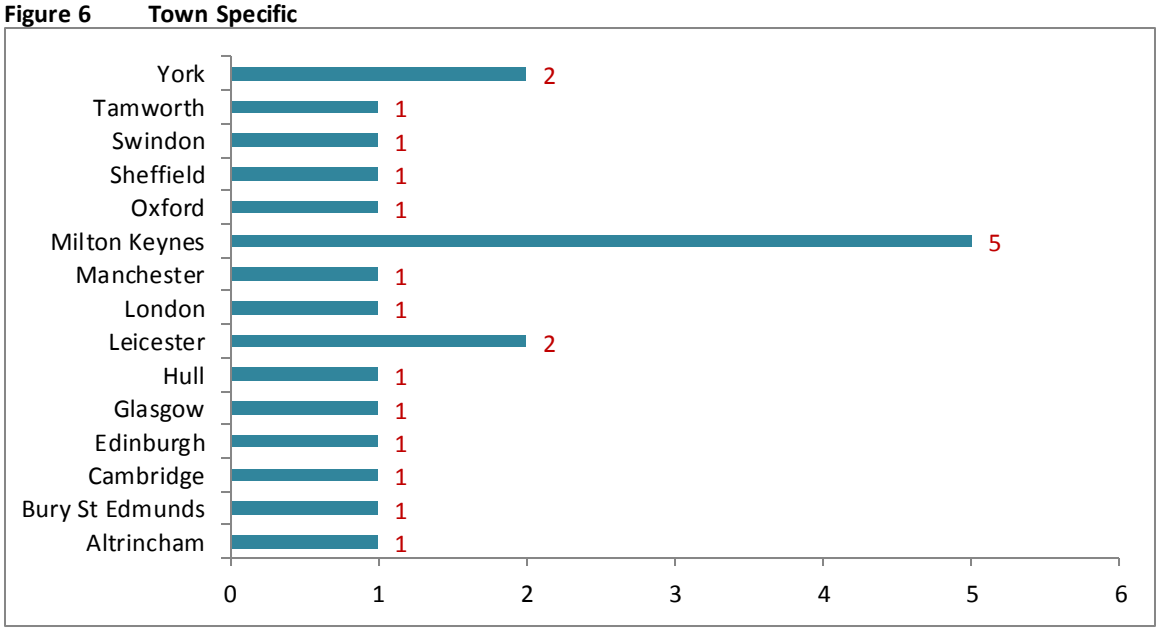


Figure 6 shows data sets it was stated covered specific towns. (Other data sets may also cover specific towns but this was not stated in the details provided).

Figure 6 Town Specific



The large majority of data and information sources apply across the whole of the UK or Great Britain if the data does not include Northern Ireland. This does not mean that each data set has equal levels of coverage of all parts of the UK within it, more that it does not exclude certain countries or regions. Generally speaking, these UK-wide data sets are what might be termed "broad" data sets – they collect a relatively shallow set of data from a very wide base, for example electricity meter readings from all occupied residential properties.

A small number of data sets are focused on very specific parts of the UK, whether a particular county, city or town. Many of these are the result of in-depth case study research projects. They can be termed "deep" data sets – they collect large quantities of detailed data and information from a very narrow base, for example attitudes to energy efficiency retrofit measures in occupiers of early 20th century terraced housing in a particular city. Given the limited time available to collect information during this study, it is entirely possible that there are many other geographically specific data sets that are not included here.

4.4 Building types covered by data sets

The types of buildings covered by each data set were sub-categorized as domestic and various sub-types of non-domestic building. The purpose of this was to see if there were certain building types that were over or under represented in the existing data sets.

Figure 7 shows the number of data sets containing data for the overarching categories of domestic, non-domestic and other. Data sets classified as 'other' typically included data that was not explicitly related to buildings but could be used in combination with building data to obtain potentially useful insights. Examples of 'other' data include weather and demographic. Some data sets were relevant to more than one category.

Figure 7 Data sets by overarching building type

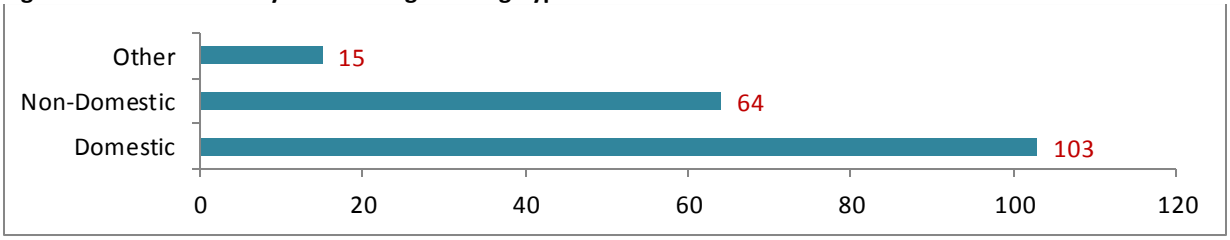
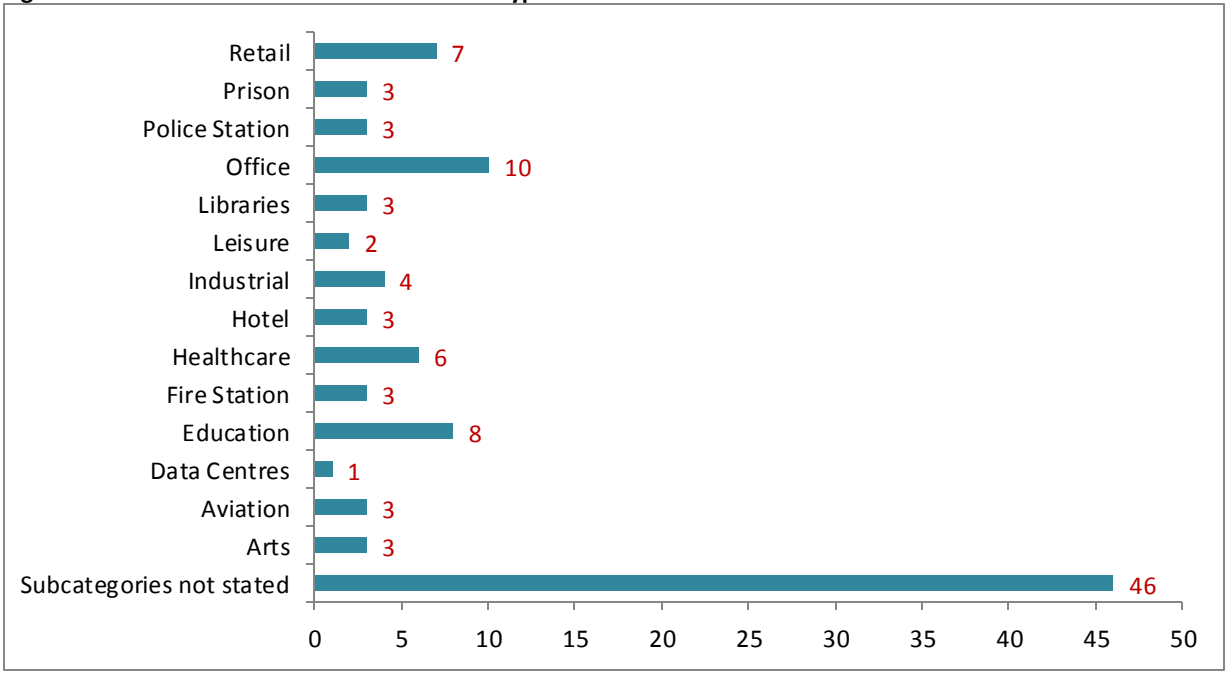


Figure 8 shows the number of data sets for each sub-type of non-domestic building.

Figure 8 Data sets for non-domestic sub-types



Note: Data sets are only labelled with subcategories when explicitly mentioned in the information provided

The results illustrate that data sets exist that have the potential to provide knowledge on a broad range of specific non-domestic sub-categories. The subcategories identified are split fairly evenly between those in the private and public sector. This may reflect that incentives and requirements exist for both sectors to record relevant data.

4.5 Life cycle stages covered by the data sets

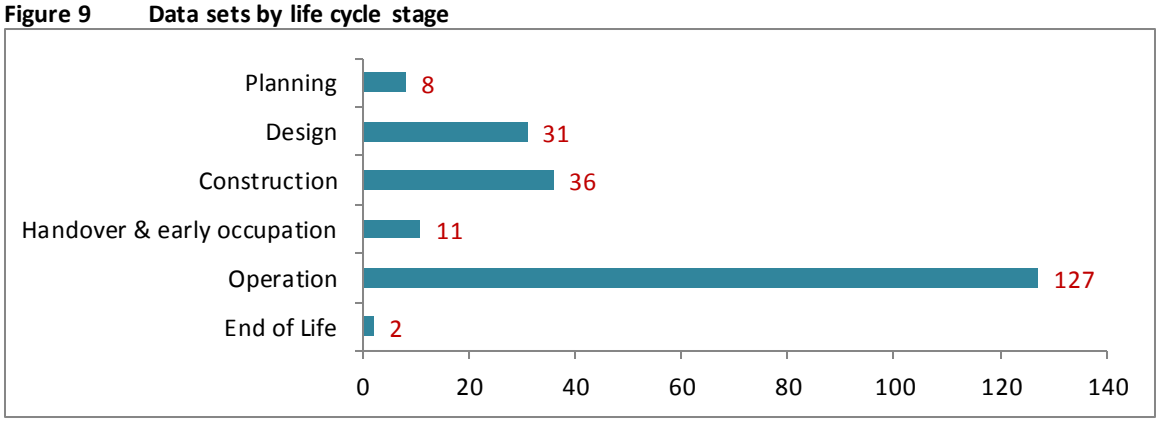
Each data set was categorized according to which asset life cycle stage(s) it related to. Although there are many different built asset life cycle breakdowns in existence, such as the RIBA Plan of Work, or the APM Project Stages, these focused too heavily on detailed breakd owns of the design

and construction phases. This level of detail was not thought to be helpful to the study, so life cycle was sub-categorized into the following stages:

1. **Planning**
2. **Design**
3. **Construction**
4. **Handover**
5. **Operation**
6. **End of life**

Figure 9 shows the number of data sets covering each of the above sub-categories. As with building type, as it was possible for a data set to cover more than one life cycle stage then the totals exceed the number of individual data sets.

Figure 9 Data sets by life cycle stage

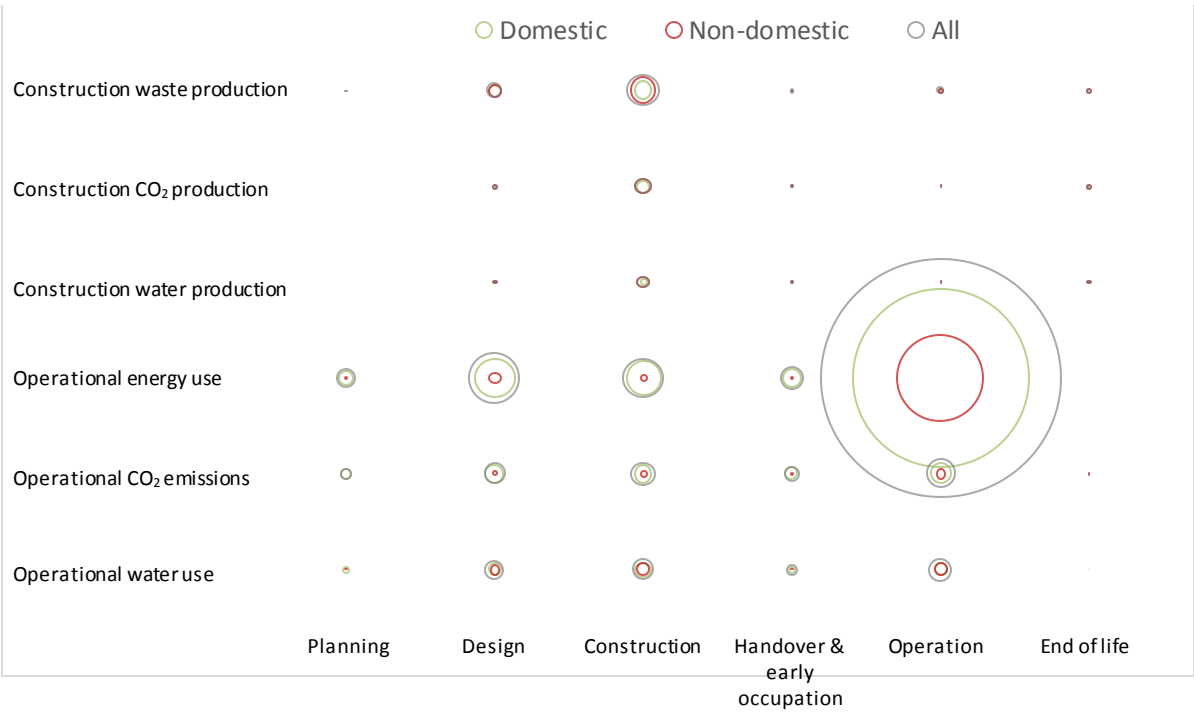


The majority of data sets relate to the operation/in-use phase of a building. This is perhaps to be expected considering that, amongst other things, data sets related to both new build and existing buildings. For the latter, design, construction and handover information may not be obtainable.

4.6 Visual representation of all data

Taking all the data sets found, we have mapped them against the sustainability indicator and life cycle stage they cover (e.g. operational energy use, and what affects it, could be considered at the planning, design and construction stages). This was repeated for the broad building types for some comparison. Further representations are set out in the following section.

Figure 10 Map of all data sets by issue, lifecycle and broad building type
Scale: Smallest Circle = 1 dataset Largest circle = 88 datasets



4.7 Mapping the data and information sources to pairs of characteristics

The above sections 4.1 to 4.5 show the number of data sources for each of the individual characteristics. The following sections show how the landscape of data sets maps to selected pairs of characteristics, as this enables a more nuanced message to emerge.

With five characteristics, it would be possible to form 24 different pair combinations for mapping. However, for practical reasons it made sense to limit the number of pairs to those which were thought capable of illustrating a particular message. In the end, four pairs were selected:

1. **Building type against life cycle stage**. This would show where data sets were concentrated in terms of the project type and life cycle stage they related to
2. **Sustainability indicator against life cycle stage**. This would show where data sets were concentrated in terms of sustainability indicator and may illustrate where additional data would be helpful
3. **Accessibility against life cycle stage**. This would test whether information related to the project phase (design and construction) was more or less accessible than information related to the operational phase (in use and demolition)
4. **Sustainability indicator against building type**. This would test whether certain types of building were particularly well or poorly served in terms of different sustainability information.

4.7.1 Building type against life cycle stage

Table 1 shows the summary of how many data sets were found linking particular building types to particular life cycle stages. As a data set could cover more than one building type and/or more than one life cycle stage, the total of the numbers shown in Table 1 is much higher than the total number of data sets identified.

Each cell in Table 1 also indicates how many of the identified data sets were deemed to be accessible (with accessibility ratings of 'Public, paid-for' (3) or 'Public, free' (4)) or of higher than average quality (with average quality ratings of 2.4 or greater). These frequencies are designated A and Q respectively but are independent of each other, so the data sets with good access ratings are not necessarily the same as those with good quality ratings. More detail is given in Appendix D.

Blank cells indicate that no data sets were found for that combination of building type and life cycle stage.

Table 1 **Data sets per building type & life cycle stage with access (A) & quality (Q) sub-totals**

	Planning	Design	Construction	Handover	Operation	End of life
Domestic	7 (A=4 Q=3)	24 (A=15 Q=10)	27 (A=13 Q=11)	7 (A=5 Q=4)	81 (A=38 Q=12)	1 (A=0 Q=0)
Non-domestic (sub-cat not stated)	1 (A=0 Q=0)	13 (A=8 Q=6)	19 (A=9 Q=10)	2 (A=0 Q=1)	28 (A=9 Q=4)	1 (A=0 Q=0)
Arts					3 (A=1 Q=1)	
Aviation			1 (A=0 Q=0)		2 (A=1 Q=1)	
Data Centre			1 (A=0 Q=0)			
Education		1 (A=0 Q=0)	3 (A=0 Q=0)	2 (A=0 Q=1)	6 (A=2 Q=2)	1 (A=0 Q=0)
Fire Station					3 (A=1 Q=1)	
Healthcare			2 (A=0 Q=0)		4 (A=2 Q=1)	
Hotel					3 (A=1 Q=1)	
Industrial			1 (A=0 Q=0)		3 (A=0 Q=0)	
Leisure					2 (A=1 Q=1)	
Library					3 (A=1 Q=1)	
Office			3 (A=0 Q=0)	1 (A=0 Q=1)	7 (A=1 Q=2)	
Police Station					3 (A=1 Q=1)	
Prison					3 (A=1 Q=1)	
Restaurant						
Retail			2 (A=0 Q=0)		5 (A=1 Q=1)	

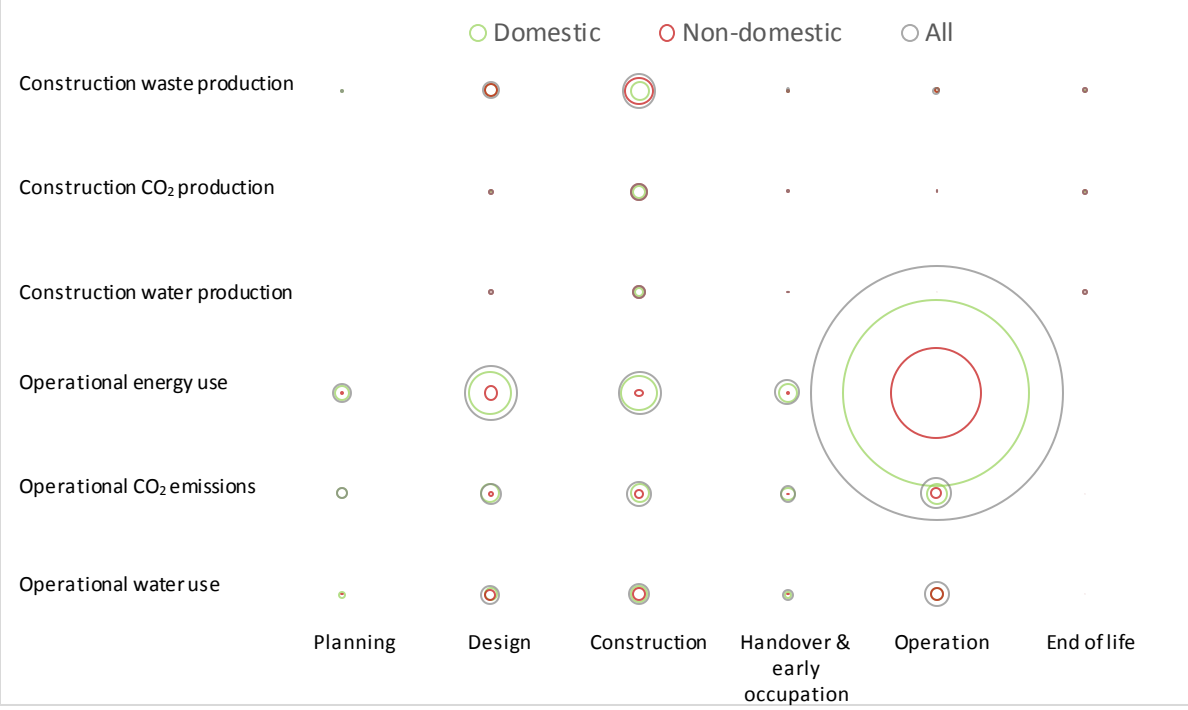
Not only does this show the dominance of the domestic sector in the data sets identified, but it also shows the dominance of operational data amongst domestic buildings. While operational data is still the best represented life cycle stage for non-domestic buildings, the pattern is not so marked.

One particular absence of data sets is that relating to operation of Data Centres. This is surprising given the energy intensity of these types of buildings. The absence of any data sets about restaurants is also notable. Although their energy consumption is on a different scale to data centres, they would also be expected to have relatively high levels of water use.

4.7.2 Sustainability indicator against life cycle stage.

Here we map all the data sets for the sustainability indicator and life cycle stage that they cover. (This is identical to **Error! Reference source not found.**)

Figure 11 Map of all data sets by issue, lifecycle and broad building type
Scale: Smallest Circle = 1 dataset Largest circle = 65 datasets

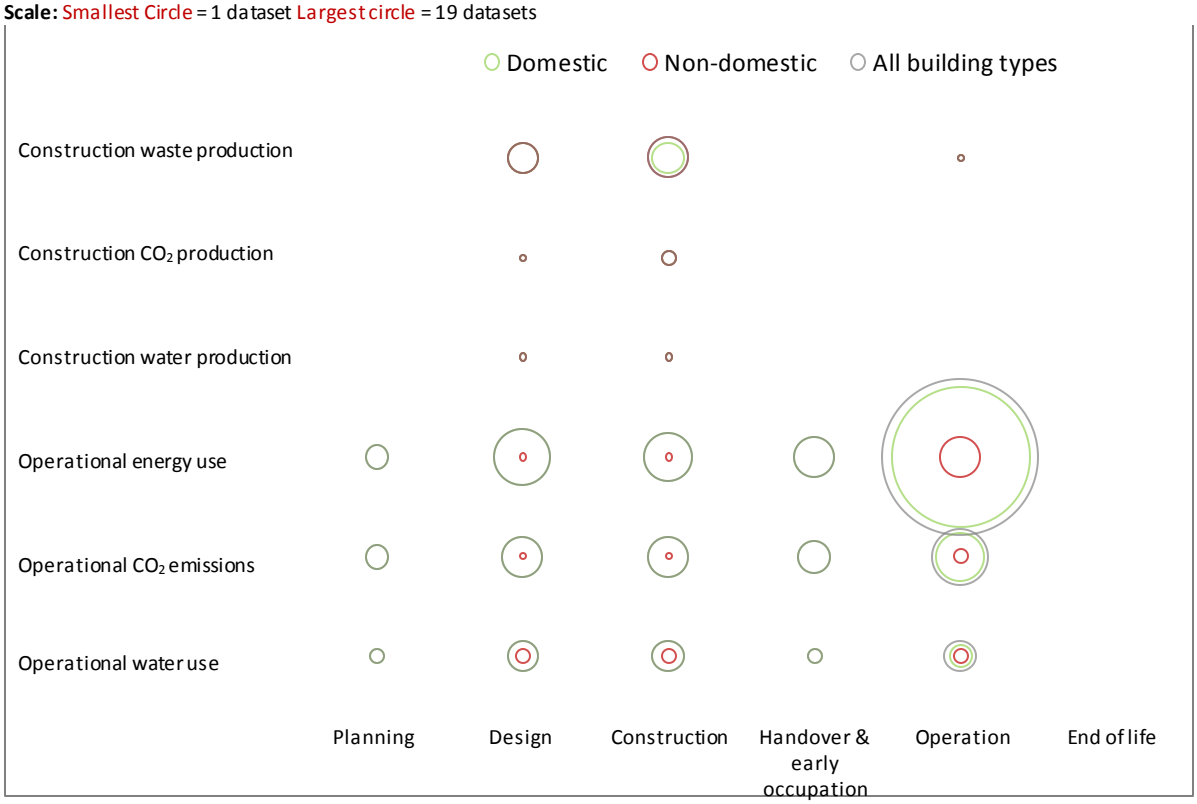


4.7.3 Accessible data and evidence

Here we map the data sets that were indicated as being either 'Public, free' or 'Public, paid-for'. This is thus a subset of all the data sets we had information for.

Figure 12 Map of accessible data sets

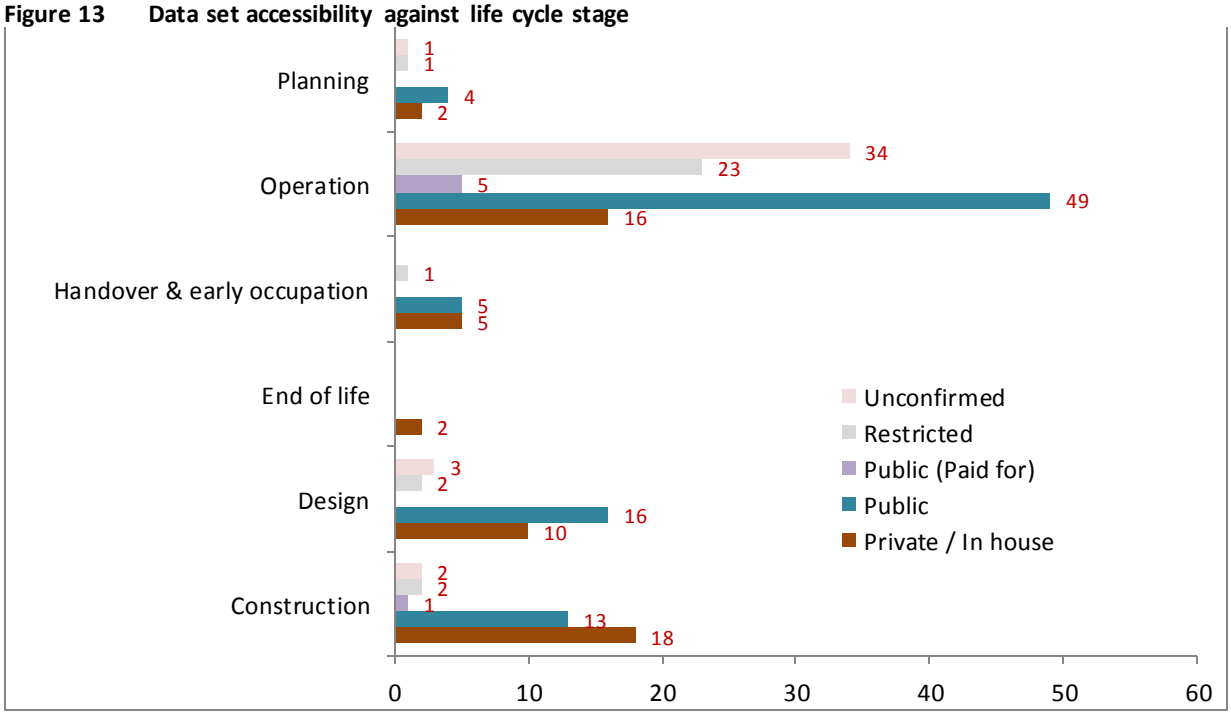
Scale: Smallest Circle = 1 dataset Largest circle = 19 datasets



4.7.4 Accessibility against life cycle stage

Figure 12 shows the pattern of data set accessibility for each of the life cycle stages.

Figure 13 Data set accessibility against life cycle stage

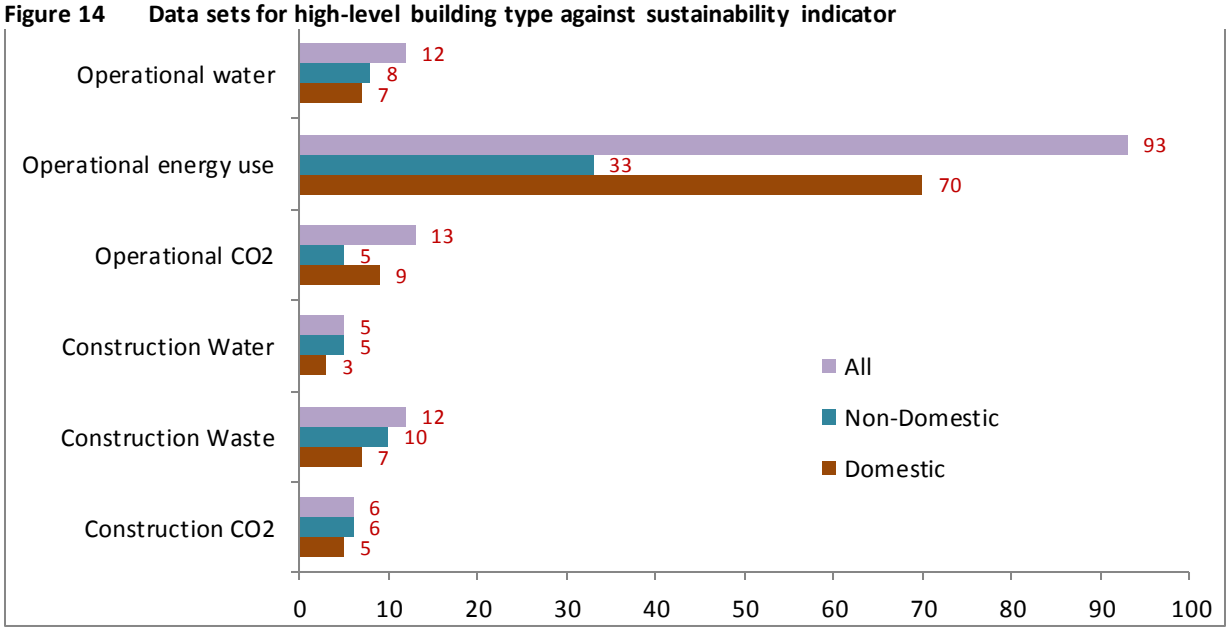


The results for each life cycle stage broadly reflect the overall distribution of the accessibility of the data. A notable exception is that the distribution for construction has a bias towards private data. Though this cannot be confirmed by the results this may reflect that a significant number of respondents were from construction companies and that building client / occupant organisations tend to be more directly involved in the planning, design, handover and operation of their buildings. It may also reflect that the dominant issues of operational energy currently is paid greater attention to at the planning, design and operational stages. Construction data, such as construction water, waste and CO_2 are typically paid more attention to by contractors than clients (perhaps due to the requirements of initiatives such as BREEAM and those run by WRAP).

4.7.5 Sustainability indicator against building type

Figure 13 shows that pattern of building type against the headline sustainability indicators. The numbers at the end of each bar indicate the number of data sets identified in each case. As a data set could be concerned with more than one indicator and more than one building type then the total across the chart is significantly higher than the number of data sets identified.

Figure 14 Data sets for high-level building type against sustainability indicator



For operational energy use and CO_2 there are significantly more domestic sets whilst for the other categories there are marginally more non-domestic sets. This is mostly likely to reflect the particular make up of respondents rather than infer general conclusions: Data sets provided by UCL had a bias towards domestic, operational energy. Survey respondents included several main contractors that predominantly work on non-domestic projects. No responses were obtained from large house builders.

5 Quality, Usability, Usefulness and Application Assessment

5.1 Method of assessment

Out of the 161 data sets concerned with sustainability in buildings, the research team identified 92 data sets to understand in more detail. For these data sets, the team wanted to know the quality of the data and how it has been processed into information, and also to understand the usefulness of the data in a variety of different ways.

5.1.1 Method of assessing quality

The quality of data or evidence was assessed by considering how it was collected and then processed, and the scrutiny applied to these stages. We asked contributors to the first survey to assess the data/evidence they had told us about against five headings:

- Data capture **standard**
- Data capture **consistency**
- **Transparency** of processing
- Type of **publication**
- Type of **peer review**

To enable contributors to make their assessment for each aspect of quality, three levels were available and some characteristics were suggested that would be typical for each level. These are presented below with the results in Table 3 Scoring matrix for data set quality

A numerical scale of levels (1-3) was chosen rather than descriptors such as *low*, *medium* and *high*. This was a conscious choice because in many cases we were asking the controller or producer of the data to rate their own dataset and it was felt that a low, medium, high scale would encourage respondents to mark high.

5.1.2 Method of assessing usefulness

In order to understand the usefulness of the data sets, we asked contributors to consider three aspects of the dataset: **usability**, **usefulness** and **application**. We were interested in how much extra work they needed to do in order to use the data/evidence (usability); how frequently they were making use of it (usefulness) and how wide that use was in the context of their own or others' work (application). Again, we offered a scale of scores and a description against each as a guide to the contributor. For ease of reference within this report, these are displayed with the associated results below in sections 5.4-5.6.

5.1.3 Presentation of the assessment

Once the data sets were all entered into an online database, we were able to create an online form that allowed questions about quality and usefulness to be answered. The contributor of each dataset was asked to complete the questionnaire only in connection with the data sets that he or she had originally contributed. This was a departure from the original research methodology, where it had been planned to ask all contributors to provide quality ratings against all the data sets they were aware of or had used. This would have been practical if there were just a few tens of data sets identified, but not with 161 data sets. Images of the online survey are shown in

Appendix B – Survey 2.

The survey asked for ratings of the current aspects of usefulness and also for ratings of the potential usefulness if certain barriers could be removed or overcome. Respondents were also asked to identify what those **barriers** were.

5.1.4 Additional rough assessment of data sets

The quality and usefulness assessments were completed for 49 of the 92 data sets that had been selected. It was acknowledged that this was a lower than ideal response rate. To obtain more information, we asked the project Peer Review group to make straightforward assessments of usefulness, choosing between low, medium or high, for all the data sets they had used or were aware of. This gave us some assessment of another 30 data sets.

5.2 Respondents to assessment survey

Table 2 shows the individuals who responded to the quality and usefulness survey.

Table 2 Respondents to the data quality and usefulness survey

Organisation	Individual	Position of individual
BAM Construct UK Ltd	Charles Law	Head of Environmental Management
BCO	Jenny Mac Donnell	Director of Research and Policy
Cardiff University	Ian Knight	Professor
Carillion plc	Quentin Leiper	Group Chief Engineer
Green Gauge Trust	Liz Reason	
Leeds Met University	Malcolm Bell	
Mace Ltd	Andrew Kinsey	Sustainability Manager
Oxford Brookes University	Rajat Gupta	Professor and Director of Oxford Institute for Sustainable Development
Skanska UK	Nigel Sagar	Senior Sustainability Manager
Willmott Dixon	Martin Ballard	Group Environment Manager
WRAP	Gareth Brown	Programme Area Manager
UCL	Megan McMicheal	
UCL	Rob Liddiard	
UCL	Simon Elam	
XCO2 Energy	Tom Kordel	Senior Energy Consultant

5.3 Data set quality

5.3.1 Proffered levels of quality

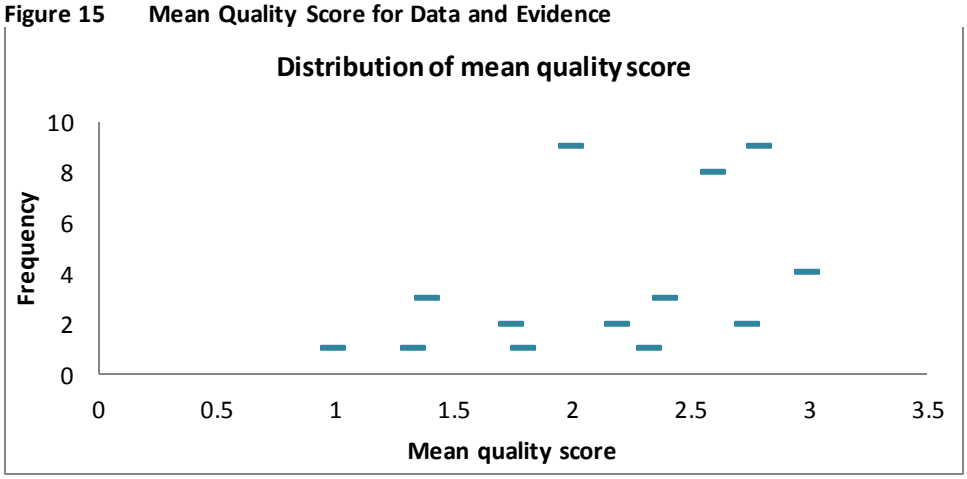
Table 3 Scoring matrix for data set quality

	Aspects of Quality				
Assessment	**Data capture standard**	**Data capture consistency**	**Transparency of processing**	**Type of publication**	**Type of peer review**
Level 3	Following an independent standard (e.g. ISO, BS, chartered institute)	Method defined pre-study and applied consistently throughout	Method of analysis or processing was published and peer reviewed	Results published and available (e.g. report, papers)	Suitable for an academic journal
Level 2	In-house standard but defined and published or available	Method evolved during early part of study, or standard not applied consistently	Method of analysis or processing was published but not peer reviewed, or only parts were published/ reviewed	Results were written up but retained in-house or restricted circulation, or only partially written up	Informal peer review (e.g. web forum)
Level 1	No known standard for capture, or standard not published	Method evolved during early & late stages of study	Method of analysis or processing has not been published	Results were not written up	No peer review
Unknown	It is not known whether a standard was applied or not	It is not known whether data was captured consistently or not	It is not known whether the processing or analysis method has been published or not	It is not known whether the results were written up or not	It is not known whether the results were peer reviewed or not

5.3.2 Summary of results

A mean quality score was calculated. Where an aspect of quality was indicated to be 'Unknown', the mean was not taken over this aspect. For example, the data from the City Form project was given ratings of 'Unknown' for both the data capture standard and its consistency, and ratings of 1, 2 and 3 for transparency of processing, type of publication and type of peer review respectively. Therefore the mean quality score is 2, as the average of the three returned ratings.

Figure 15 Mean Quality Score for Data and Evidence



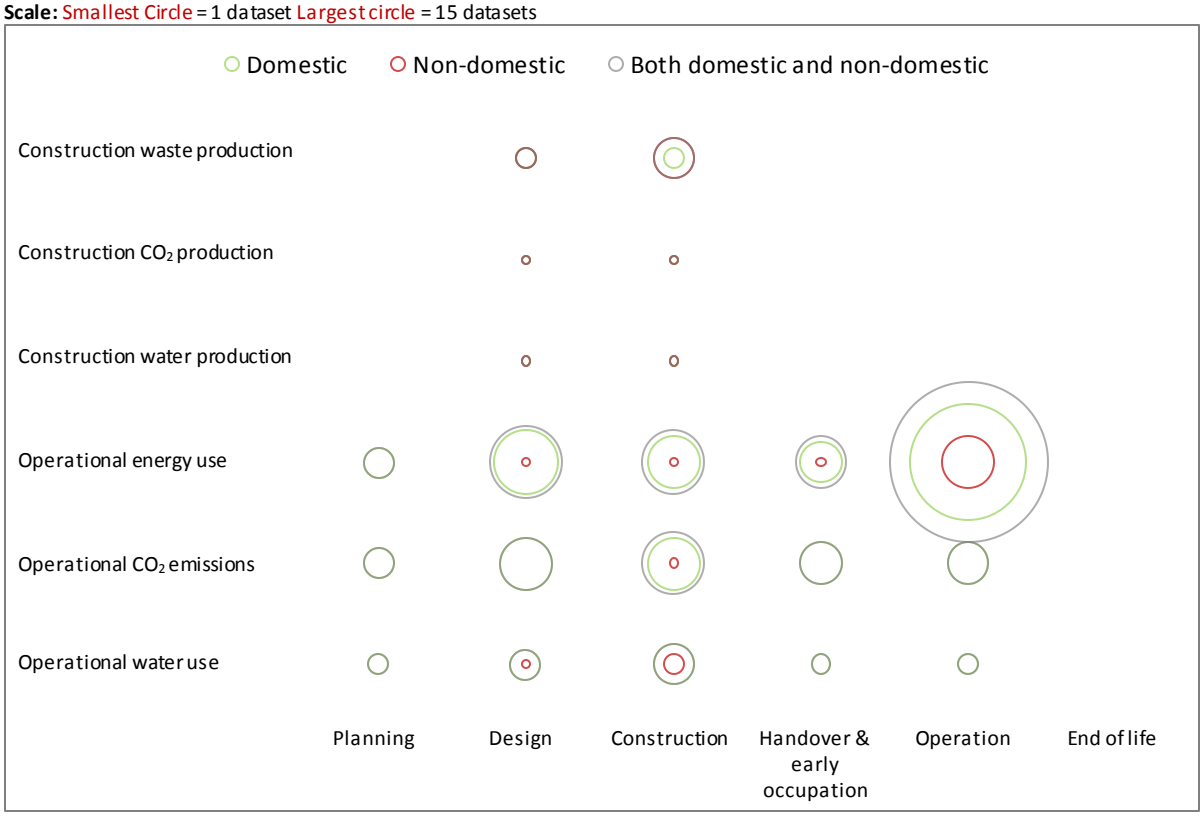
5.3.3 Comment

Quality is tending to be medium or high level. This is perhaps unsurprising, as we would expect people to be using information that they understand to be of good quality. A more rigorous study of the data landscape or a deep investigation into a few data or evidence sets might further explore these aspects of quality independently (we were asking the producers of the data and could not expect an entirely objective response).

5.3.4 Mapping high quality data sets

We have mapped the higher quality data sets for domestic and non-domestic buildings across life cycle stage and sustainability indicator. *High quality* was defined as achieving a mean quality score of 2.4 or greater. This could be achieved by gaining scores of two 3's and three 2's across the five aspects of quality.

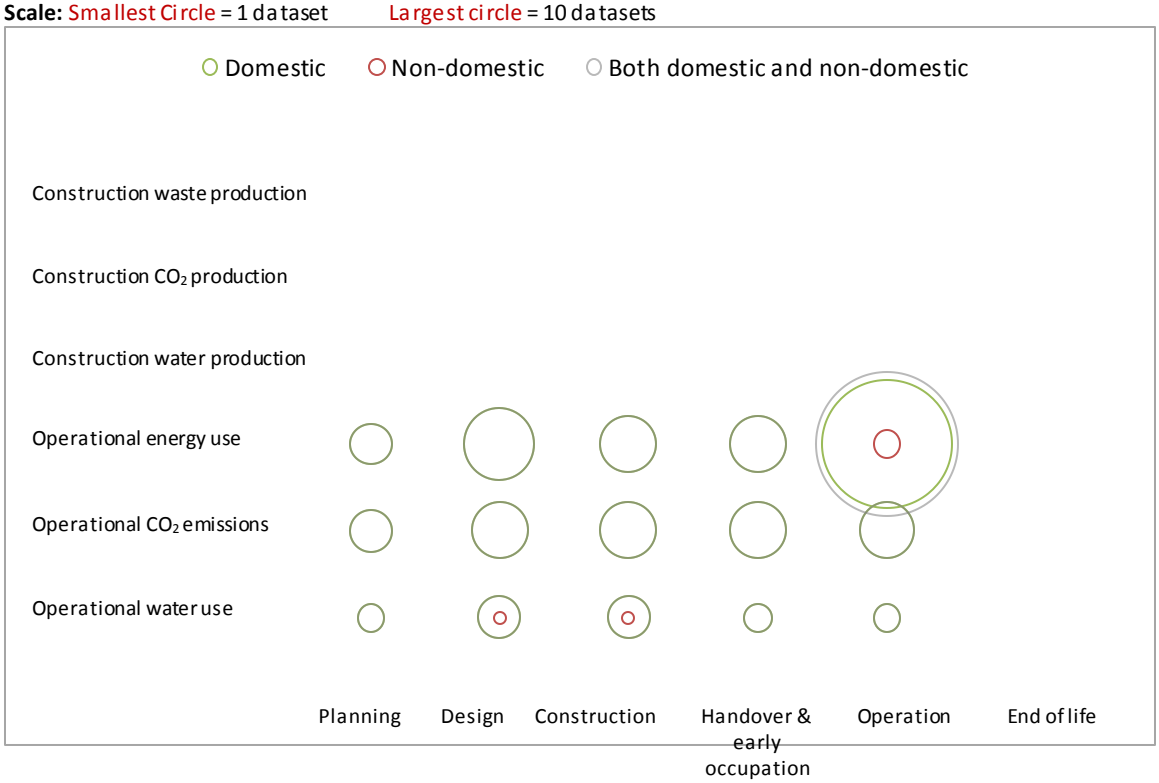
Figure 16 Mapping of high quality data sets
Scale: Smallest Circle = 1 dataset Largest circle = 15 datasets



5.3.5 Mapping high quality, accessible data sets

The mostly widely usefulness data sets will be those of good quality that are also accessible. Here we map all data sets that are deemed high quality (mean score 2.4) and also either available to the public for free or for a charge.

Figure 17 Mapping of high quality, accessible data sets
Scale: Smallest Circle = 1 dataset Largest circle = 10 datasets



5.3.6 Lessons

The map above shows where there is a paucity of data that is of a high quality. There is a greater number of gaps in this map than the map of all data sets. Caution should be taken when data of poor or unknown quality is used; misleading conclusions may be drawn.

Intervention might improve the cover of the data map with high quality data.

Intervention might make currently inaccessible data publicly available.

5.4 Usability

For the purposes of the current work, *usability* was defined as how readily the data could be put to use. The usability of a data set was considered poor if significant additional work, such as combining with other data (or information) or further analysis, was needed in order to make the data usable.

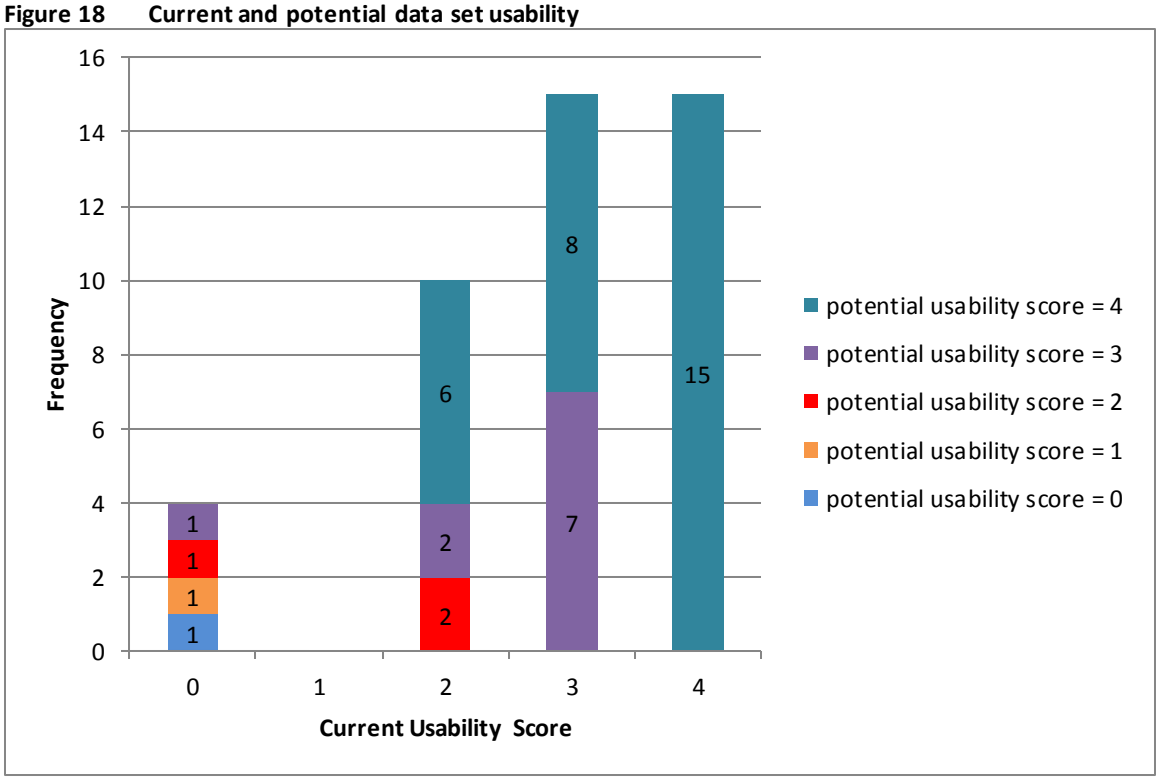
5.4.1 Assessment scale

Table 4 Assessment scale for usability

Level 4	Data/evidence is complete, and clear conclusions or actions have been derived from it
Level 3	Data/evidence is complete, and some conclusions or actions can be derived from it.
Level 2	Data/evidence is complete, and when further analysed or combined with other data or information, conclusions and actions can be drawn.
Level 1	Data/evidence is complete but no conclusions or actions can be derived from it.
Level 0	Data/evidence is not complete.

5.4.2 Summary of results

Figure 18 Current and potential data set usability



5.4.3 Comment

The top score for usability was level 4. We defined this as 'Data/evidence is complete, and clear conclusions or actions have been derived from it'.

The majority of the data sets assessed were said to be usable, that is the data or evidence is complete and either conclusions or actions to implement have been derived from it already, or can readily be arrived at.

5.4.4 Top usability scorers

Sets of evidence and data rated as reaching level 4 for usability were:

- City Form
- Commercial Buildings Energy Consumption Survey
- CREST One-minute Resolution Domestic Electricity Use Data, 2008-2009
- Energy Performance Certificate data
- Household Electricity Use Study
- IEE iSERVcmb Project. Use of sub-metered energy consumption for HVAC components in European Non-Domestic Buildings
- Leeds Met database of Co-heating
- MIMA projects
- Scottish Environmental Attitudes and Behaviours Survey 2008
- Skanska UK waste data
- Stamford Brook first study 2001 - 2008
- Temple Avenue existing dwellings
- Temple Avenue new build
- The Elm Tree Mews project
- Waste data on civils and building projects

5.4.5 Lowest usability scorers

Lowest usability scorers (rated at Level 0):

- Carbon emissions (construction operations)
- Construction – water use
- Recycled Content Database
- UKCG Construction, Demolition and Excavation Waste Data

5.5 Usefulness

We defined *usefulness* here as how frequently information or knowledge derived from the data is put into action or taken into consideration.

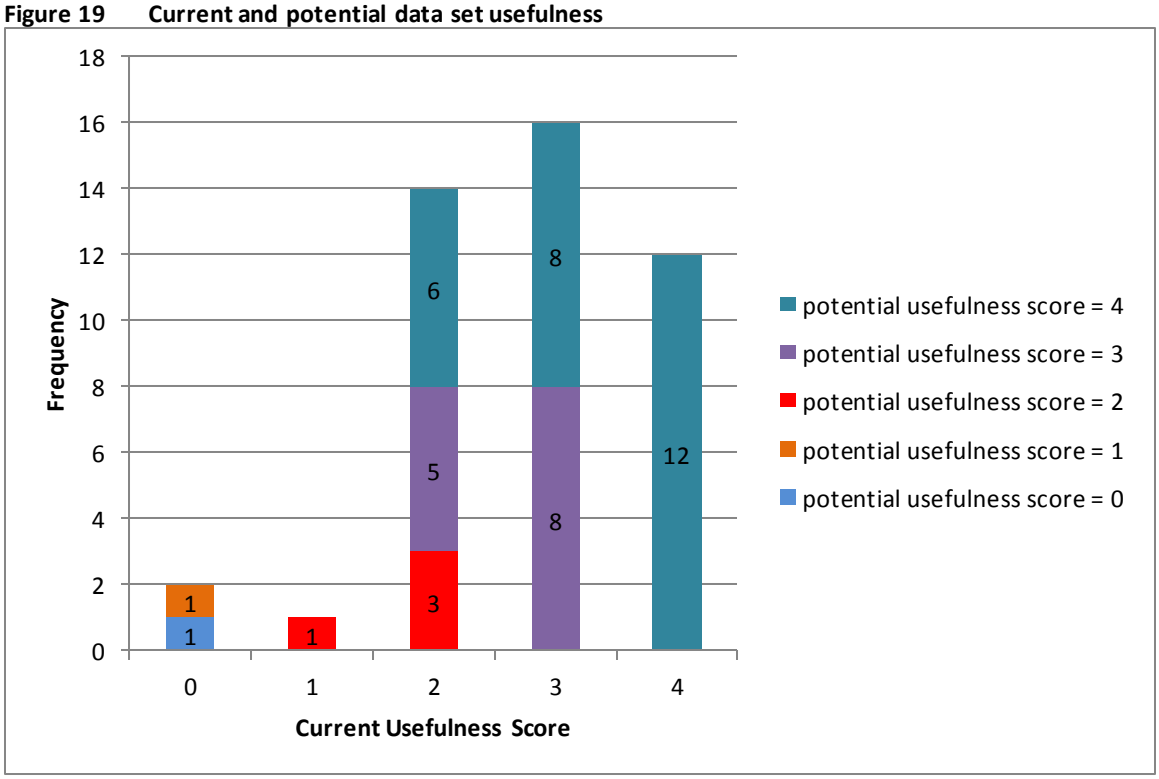
5.5.1 Assessment scale and summary of results

Table 5 Assessment scale for usefulness

Level 4	I always use the dataset for decision-making in relevant circumstances (need to have)
Level 3	I sometimes use the dataset for decision-making in relevant circumstances (nice to have)
Level 2	Data/evidence has raised my awareness of the issue and caused a related change in my behaviour (e.g. climate change data leads to using more energy efficient building design models)
Level 1	Data/evidence has raised my awareness of the issue but no change in behaviour (e.g. I know petrol consumption is lower at 56mph but I still drive at 70mph).
Level 0	No change to my behaviour

5.5.2 Summary of results

Figure 19 Current and potential data set usefulness



5.5.3 Comment

The bar chart shows that the vast majority of the data/evidence we were told about is raising awareness and has led to some degree of behaviour change, in some cases being frequently used to inform decisions.

5.5.4 Top usefulness scorers

- BAM Construction Energy data
- BAM Construction Ltd Construction, Demolition and Excavation Waste data
- BAM Construction Water use
- IEE iSERVcmb Project. Use of sub-metered energy consumption for HVAC components in European Non-Domestic Buildings
- Leeds Met database of Co-heating
- MIMA projects
- Skanska UK waste data
- Stamford Brook first study 2001 - 2008
- Temple Avenue existing dwellings
- Temple Avenue new build
- The Elm Tree Mews project
- Waste data on civils and building projects

5.5.5 Lowest usefulness scorers

Lowest usability scorers (rated at Level 0)

- City Form
- CREST One-minute Resolution Domestic Electricity Use Data, 2008-2009
- Sheffield Hallam University Energy Surveys

5.6 Application

We were interested in how widely the data, or information and knowledge derived from it, is being used. We refer to this aspect of usefulness as *application*.

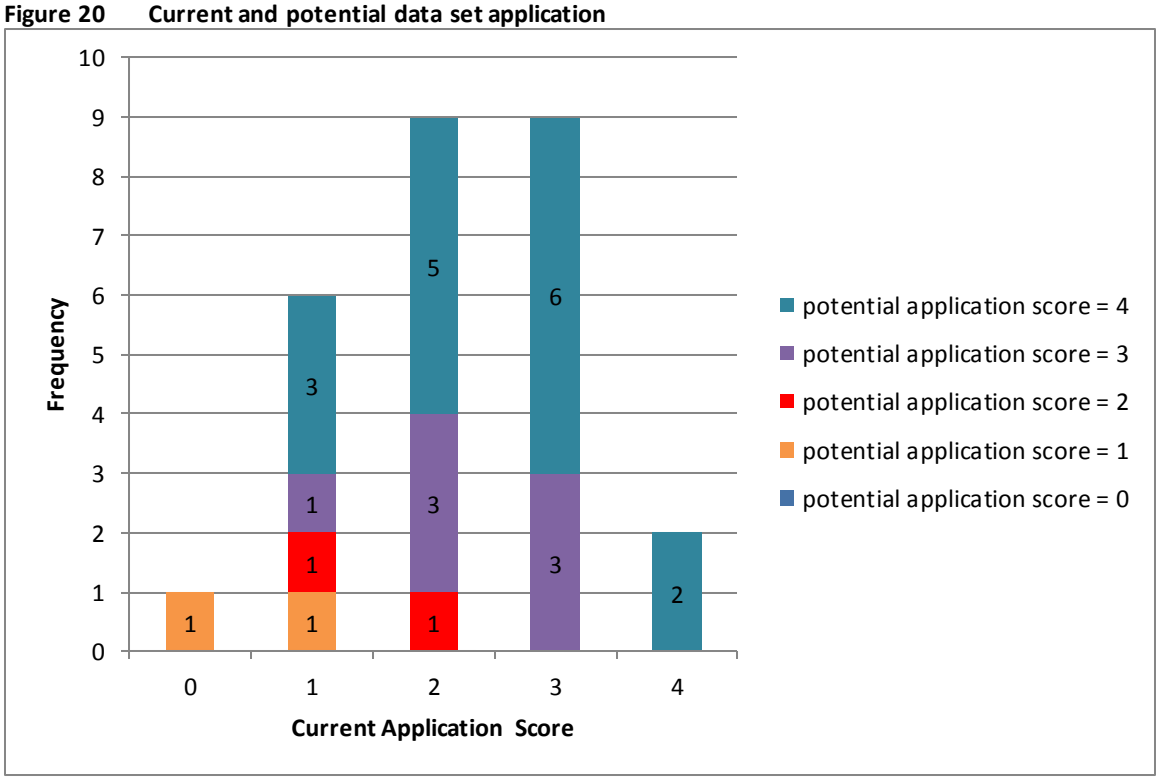
5.6.1 Assessment scale and summary of results

Table 6 Assessment scale for data set application

Level 4	Application is industry-wide.
Level 3	Application is organisation-wide.
Level 2	Data/evidence is used on all projects/activities I'm involved with (or on many relevant projects within my organisation).
Level 1	Data/evidence is used on some of my own projects/activities (or on a few relevant projects within my organisation).
Level 0	Data/evidence or information derived from it is not used.

5.6.2 Summary of results

Figure 20 Current and potential data set application



5.6.3 Comment

Application of the data/evidence shows a more even spread than the other aspects of usefulness. However, many sets are said to be in use across the industry.

The potential application of many data sets has been rated as level 4. This may be a case of optimism bias.

5.6.4 Top Scorers for Application

The following were rated as reaching Level 4 for application:

- BAM Construction Energy data
- BAM Construction Ltd Construction, Demolition and Excavation Waste data
- BAM Construction Water use
- BRE SMARTWaste Summary Data
- Business Case for Resource Efficiency
- DECoRuM carbon mapping data of houses
- Energy Performance Certificate data
- Homes Energy Efficiency Database
- Household Electricity Use Study
- IEE HARMONAC Project. Evidence of effectiveness of EPBD AC Inspections and the potential role of HVAC energy submetering
- IEE iSERVcmb Project. Use of sub-metered energy consumption for HVAC components in European Non-Domestic Buildings
- Leeds Met University database of Co-heating
- MIMA projects
- Net Waste Tool
- Resource Efficiency Benchmarks
- Stamford Brook first study 2001 - 2008

- Temple Avenue existing dwellings
- Temple Avenue new build
- The Elm Tree Mews project
- UKCG Construction, Demolition and Excavation Waste Data

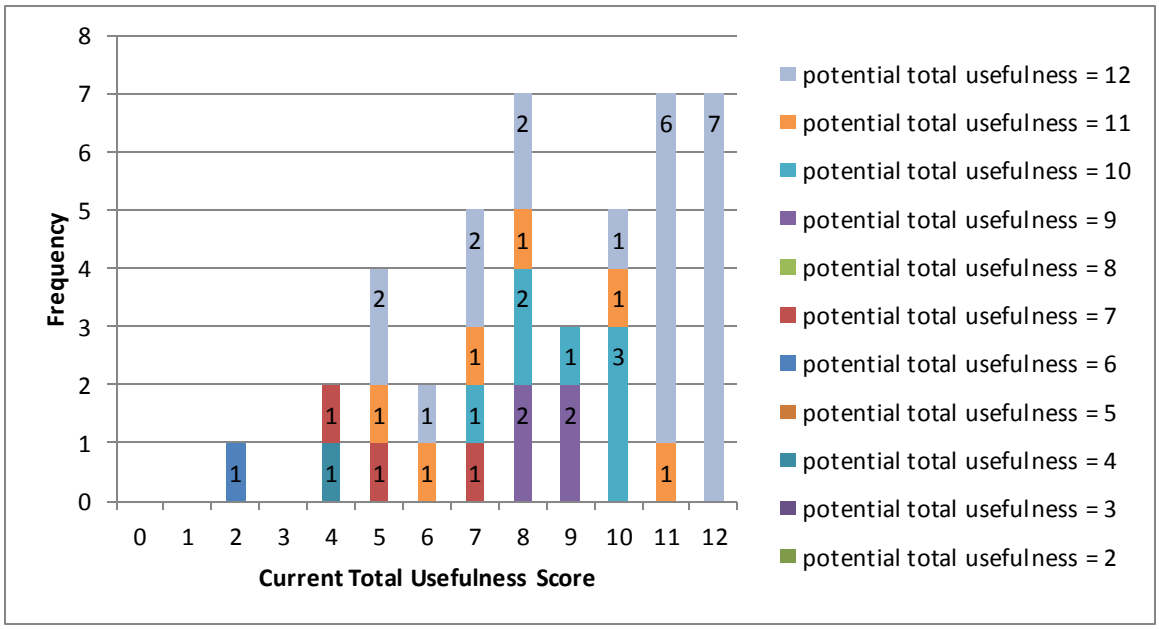
5.6.5 Lowest Scorers for Application

With lowest application scorers (rated at Level 0):

- City Form

5.7 Combined Usefulness Score

The total of the three aspects of usefulness (usefulness, usability and application, as defined in our assessment questionnaire) was taken for each data set and plotted.



5.7.1 Top-scoring Data Sets

The following scores 4 out of 4 on usability, usefulness and application:

- IEE iSERVcmb Project. Use of sub-metered energy consumption for HVAC components in European Non-Domestic Buildings
- Leeds Met database of Co-heating
- MIMA projects
- Stamford Brook first study 2001 - 2008
- Temple Avenue existing dwellings
- Temple Avenue new build
- The Elm Tree Mews project

None of these reach the top mean **quality** score (3). Amongst them, the top quality score is 2.8 and this is achieved by the following:

- IEE iSERVcmb Project. Use of sub-metered energy consumption for HVAC components in European Non-Domestic Buildings
- Stamford Brook first study 2001 - 2008
- Temple Avenue existing dwellings
- Temple Avenue new build
- The Elm Tree Mews project

5.8 Data Sets for Further Exploration

We have listed five data sets that have been assessed as of good quality and highly useful. As previously recorded, we were unable to poll a large number of data users in order to assess the data sets. Therefore, we can attribute no statistical rigour to our findings here. However, we note that of these five data sets, four are the result of intensive investigation of a relatively small sample of buildings. In order to understand the benefits and limitations of such data sets, we explore one of these high scorers in greater detail in our case studies in 7.6 Stamford Brook. Having also investigated some broader data sets and how they can be used, we will compare deep studies to broader studies in a subsequent section below (7.7 Broad data sets vs Deep investigations).

5.9 Rough Assessment of Usefulness

In order to increase the number of assessments of data and evidence sets, we asked our Peer Review Group to assess the usefulness of those, choosing between high, medium and low. Their assessment was compared to those of the other contributors.

The number of data sets with multiple assessments was 27. Of these, for 21 data sets the assessments differed.

It is acknowledged that the assessment of the data sets is based on very small samples (one assessment per data set) and that bias may be evident in these assessments. However, we believed that we could only receive a meaningful assessment from a contributor with familiarity of the data and some experience of using or collecting it. Time did not allow us to seek out users of the data who had not already come forward.

The Peer Review Group assessments have not been analysed further.

6 Potential and Barriers Analysis

6.1 Seeking the Most Under-exploited Data Sets

To investigate the most unexploited data and evidence, we looked at the **potential** scores for usability, usefulness and application (as defined in the previous sections 5.4, 5.5 and 5.6 respectively) and identified the sets given the **highest potential** scores (4 out of 4) but a current score that is lower. These might be considered as the data sets that could be the most effectively used but that needed additional work to achieve this.

- AECB Low Energy Buildings Database
- Electricity supplier meter data
- EVALOC - Monitoring data of energy performance of 50 owner occupier houses
- EVALOC - Quantitative and Qualitative Social surveys of 90 households across six low carbon communities
- Gas supplier meter data
- Greenhomes research, Greenoak Housing Association
- National Energy Efficiency Database (NEED)

To see what aspect needed most improvement, of those listed above, we then sought data/evidence with the **lowest current usefulness** (which was found to be 2). This yielded the following items:

- EVALOC - Quantitative and Qualitative Social surveys of 90 households across six low carbon communities
- Electricity supplier meter data
- Gas supplier meter data
- NEED

Considering the data sets with the **highest potential** in usefulness, usability and application, then looking for those with the **lowest current usability** (which was level 2), the results were:

- EVALOC - Quantitative and Qualitative Social surveys of 90 households across six low carbon communities
- EVALOC - Monitoring data of energy performance of 50 owner occupier houses
- Electricity supplier meter data
- Gas supplier meter data
- Greenhomes research, Greenoak Housing Association

Considering the data sets with the **highest potential** in usefulness, usability and application, then looking for those with the **lowest current application** (which was level 1), the results were:

- Electricity supplier meter data
- Gas supplier meter data

6.2 Seeking the Greatest Difference between Current and Potential Usefulness

We looked at the difference between current and potential scores for each aspect of usefulness (usability, usefulness and application as described in the previous sections 5.4, 5.5 and 5.6 respectively). We consider the data sets with the greatest difference in this section. We asked contributors to describe the barriers to reaching the potential score. These barriers might be interpreted as the most significant barriers that we were told about.

6.2.1 Potential Usefulness

We looked at the data sets with the greatest difference between current usefulness and potential usefulness (as judged by our contributors). The greatest increase in usefulness was just 2, and the data sets this was found in are presented along with the barriers described by the contributors:

- 3 **Airtightness** data from building regulations tests on new dwellings – barrier is "the availability and quality of data collection".
- 64 **Electricity supplier meter** data – barrier is "Access to the micro-data and metadata/documentation is heavily restricted which significantly reduces the potential utility of the dataset." Note, this is a **private** dataset.
- 14 **EVALOC - Quantitative and Qualitative** Social surveys of 90 households across six low carbon communities – our contributor told us that the data could be used to answer new research questions but that 'data curation' of the data set would be needed first.
- 65 **Gas supplier meter** data – barrier is "Access to the micro-data and metadata/documentation is heavily restricted which significantly reduces the potential utility of the dataset." Note, this is a **private** dataset.
- 44 **National Energy Efficiency Database** – barrier to reaching potential usefulness is described as, "less restricted access to NEED data would enhance the usefulness to a wider range of stakeholders.
- 25 **UKCG Construction, Demolition and Excavation Waste** Data – barrier is "Data needs to be submitted by all member organisations to a consistent level of quality."

6.2.2 Potential Usability

We looked at the data sets with the greatest difference between current usability and potential usability (as judged by our contributors). The greatest increase in usability was 3, and the only dataset this was found in was:

- 25 **UKCG Construction, Demolition and Excavation Waste Data** – barrier is "Data needs to be submitted by all member organisations to a consistent level of quality."

Next greatest increase was 2, found for the following:

- 64 **Electricity supplier meter** data – barrier is "Access to the micro-data and metadata/documentation is heavily restricted which significantly reduces the potential utility of the dataset."
- 28 **EPC Certificates for BAM Construction** Projects – barriers is "EPCs do not give enough detail to make clear conclusions"
- 14 **EVALOC – Quantitative and Qualitative** Social surveys of 90 households across six low carbon communities – our contributor told us that the data could be used to answer new research questions but that 'data curation' of the data set would be needed first.
- 13 **EVALOC – Monitoring** data of energy performance of 50 owner-occupier houses – our contributor told us that the data could be used to answer new research questions but that 'data curation' of the data set would be needed first.
- 65 **Gas supplier meter** data – barrier is "Access to the micro-data and metadata/documentation is heavily restricted which significantly reduces the potential utility of the dataset."
- 350 **Greenhomes research**, Greenoak Housing Association
- 32 **Mace carbon emissions** (construction operations) – barrier is "Data is not 100% complete due to difficulties in monitoring on sites where Mace do not have metered supplies (e.g. refurbishment of airports and work to fit-out existing buildings)."

6.2.3 Potential Application

We looked at the data sets with the greatest difference between current application and potential application (as judged by our contributors). The greatest increase in application was 3, and the data sets this was found in was:

- 3 **Airtightness** data from building regulations tests on new dwellings – barrier being "The lack of a mechanism of pulling the data together into a national data set. This could be done by government as part of the building regs reporting process but no one seems very interested in doing it."
- 64 **Electricity supplier meter** data – barrier being "Access to the micro-data and metadata/documentation is heavily restricted which significantly reduces the potential applicability of the dataset. Actual (i.e. not modelled) energy consumption data is vital to almost every research project with an energy element so the potential uses are virtually boundless. Not having access to disaggregated energy consumption data has been a real problem for many research projects investigating the socio-technical drivers of energy demand."
- 65 **Gas supplier meter** data – barrier is "Access to the micro-data and metadata/documentation is heavily restricted which significantly reduces the potential utility of the dataset."

6.2.4 Deeper Analysis – Case Studies

In section 7, we present case studies of data sets that through the analysis in this section we have identified as being noteworthy. They are data sets that could score highly given appropriate attention and intervention. Through the case studies, we seek the nature of those interventions and the opportunities and benefits to be got through those interventions.

- Underexploited Data: **EVALOC** – two data sets resulting from this research project were identified as having good potential for greater usability and usefulness.
- Unexploited Data: **NEED** – this data set is vast and has been identified as being potentially very useful; also it encompasses both the gas and electricity supplier meter databases
- **Energy supply meter data** – not assessed as a case study as much of the information can be accessed through NEED
- Unexploited Data: **Display Energy Certificates (DECs)** – this data set has been discussed by our Peer Review Group and identified for further comment. It represents a large number of non-domestic buildings.
- Unexploited Data: **Airtightness Test Results** – identified as potentially useful and known to be large, issues with and benefits of its further use have been discussed with both our Peer Review Group and Steering Group
- Exploited Data: **Stamford Brook** – this data set is representative of all the data sets that scored very highly for all aspects of data quality and usefulness (with the exception of IEE iSERVcmb Project). It is markedly different from the data sets covered by the other case studies in that it is data from a small sample, studied in great depth. Later, we discuss the differences between deep and broad data sets (7.7 Broad data sets vs Deep investigations)

6.3 Barriers described by our contributors

In both stages in our data collection, we asked contributors to tell us how the data/evidence they were telling us about might be better or more usefully used. In the second survey, we asked contributors to make an assessment of various aspects of the quality and usefulness of the data currently and what each aspect could potentially score. Above, we presented the responses to these questions. For each aspect, we asked the contributor to describe the barriers to reaching that potential usability, usefulness and application. Here we present all the barriers (or suggested interventions) submitted to us. Barriers have been grouped according to what stage in the data-to-change ladder they occur (please refer to Figure 21). Each barrier or invention was counted once for every data set it was suggested for. (A full transcript of barriers is included in Appendix F.)

Table 7 Reported Barriers and Possible Interventions

Barrier Or Requirement In Order To Meet Potential	Frequency With Which It Was Reported
Collection:	
Relies on users to supply or update data / automated data collection needed	5
Finer-grained data needs to be collected	6
Standardisation of collection method needed	5
More data points required	5
Funding for updates	6
Processing:	
Finer-grained data needs to be reported / made accessible	5
Access to benchmark data needed	4
Common method of reporting/publishing needed	1
Cost of access to data is prohibitively high / free access to data needed	1
Anonymise data in order to allow greater access for re-use	6
A single, national repository required	1
Detailed analysis needed	5
Cognition:	
Case studies of best practice based on evidence needed	1
Integration with decision-making process	6
Application:	
Awareness of data/evidence needs to be raised	3
Evidence points to actions that cannot yet be carried out	1

6.4 Commentary on the barriers identified preventing greater data set usefulness

6.4.1 Access to Data

Many of the barriers indicated relate to access to the data, either because the cost of access is high or the published data is aggregated or not of sufficiently low-level. There may also be issues of privacy and data protection considerations.

Clearly this can be a significant hurdle where the data set is acknowledged to be private, but for data sets that are already partially available to the public then this should be easier to overcome, provided the data collection protocols do not proscribe greater accessibility.

6.4.2 Completeness of Data Set

A barrier that was mentioned in several instances concerned the completeness of the data set. Particularly where data was being collected through a trade association, but essentially on a voluntary basis, it was seen that greater value would be gained from the data if more or all of the trade association's members contributed to the data set.

6.4.3 Aggregation of Data

For airtightness test results, there was the comment that data is being collected in a piecemeal manner and that data resided in several databases. The usefulness of the data would be much improved if a mechanism were put in place to collect this data together and there was a single, national gateway to it. This particular data set is the subject of a case study (7.3 Airtightness Test Results)

6.5 Lessons for the GCB

Various interventions may be needed to make data sets available in a usable state or reliable/representative of reality.

7 Case Studies

In this section, we look in more detail at some of the data sets identified through our analysis as being of particular interest. Some of these are data sets that could potentially be of great usefulness but where barriers exist and interventions are needed in order to fulfil that potential.

These unexploited data sets are examples of those where access is restricted and where better use could be made of the data if access was more open, or those which are less useful or usable than they could be because of the design and structure of the data.

One data set is scored highly in both assessment of its quality and usefulness – Stamford Brook studies. This is an example of a deep investigation of a small sample of dwellings and this makes it similar to some of the other high-scoring data sets. It has been studied in order to understand why it has scored highly in all aspects if usefulness.

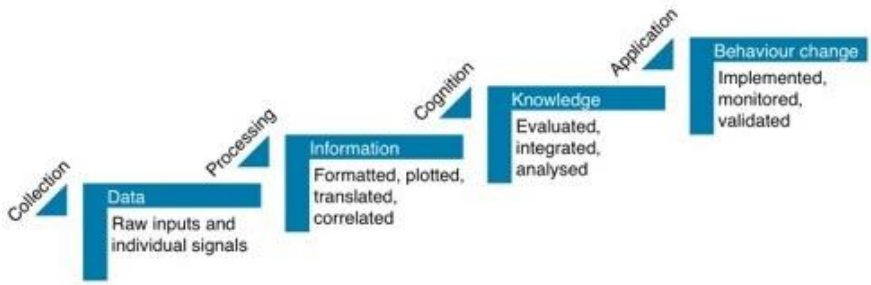
To put this part of the report in context, this section starts with an overview of the hierarchical theory of data, information and knowledge.

7.1 From data to behaviour change – The Knowledge Ladder

Existing theories of knowledge management are based on the idea of a hierarchy that starts with data, progresses through information and knowledge and ultimately ends up with wisdom[1]. This is usually represented as a pyramid with data at the bottom and wisdom at the top.

For this project, the standard hierarchy has been adapted to a ladder which emphasises both the states (data, information, etc) and the transitions between them, necessary to get from one state to the next. The topmost state of wisdom has also been replaced with behaviour change, as this is the ultimate objective that GCB is seeking to explore through this research project. The ladder, shown in Figure 21, therefore has seven different elements to it.

Figure 21 The steps between data and behaviour change



The eight elements in the knowledge ladder are summarised as follows:

1. **Collection** is the first transition – undertaking monitoring, surveying and other investigations to obtain data.
2. **Data** is the basic input – Data is usually characterised by its quantity, its scope, and the methods that have been used to collect it.
3. **Processing** is the second transition – analysis of data to extract patterns, summaries, results and conclusions. Processing is what turns quantities of data into a manageable amount of information.
4. **Information** is meaning extracted from data. This could be in the form of reports, academic papers, executive summaries.
5. **Cognition** is the third transition – translating information into a form that can be internalised by people, for example by putting it into their own personal contexts. Cognition

[1] Skyrme D.J. and Amidon D.M., *Knowledge management*, Institute of Management, Corby, 1997

is closely linked to learning and happens in different ways, for example by listening, reading, discussing, practising. People have different preferences for learning and some of the key models are linked to behavioural traits, for example the Kolb learning cycle.

6. **Knowledge** is the internalisation of information by people so that it has the potential to be acted on. Knowledge in itself is not the end of the journey since any number of barriers may prevent that action being taken.
7. **Application** is the fourth transition – putting knowledge into practice in a real situation. Many of the barriers to application may not have anything specific to do with the knowledge – the potential is there but other factors are getting in the way, such as lack of motivation, lack of appropriate incentives, habit.
8. **Behaviour change** is the successful application of knowledge that has been acquired. Behaviour change may be either conscious (the individual has to think before acting) or unconscious (the individual acts automatically). Unconscious behaviour change is more advanced and ultimately more successful, provided the behaviour is a desirable one.

While data, processing and information dissemination are essential starting points, none of these on their own will lead to change in behaviour which is essentially a human response based on capability, opportunity and motivation (see 9.1.1 'Influences on Behaviour'). In the context of sustainability in construction and the built environment, improvement comes from people behaving differently in their conscious and unconscious choices and decisions, including but not limited to reacting to new knowledge. This is the principal reason why this research project is considering not just the availability and quality of data and information, but also the usefulness of data and information in leading to knowledge acquisition and behaviour change.

Section 6 gives a broad picture of where unexploited data sets exist within those identified by the study. The following sub-sections look at some of those data sets in more detail and in doing so identify some ideas for how GCB might improve the exploitation of these data sets in particular and other unexploited data sets in general.

Another data set of interest is that coming from the Stamford Brook studies carried out by Leeds Metropolitan University Centre for the Built Environment (Leeds Met). This scored highly both for quality and usefulness. This is explored in the case study below.

In order deepen our understanding of the data sets considered noteworthy, we used a framework to plot out where and the nature of the interventions needed for that data set to reach its full potential in terms of usefulness. This framework is displayed in Appendix G Framework for Exploring Interventions.

7.2 EVALOC

7.2.1 Summary

EVALOC[2] is a 4-year research project led by Oxford Brookes University in collaboration with University of Oxford that is evaluating the impact and effectiveness of low carbon community projects in six selected case study locations. In particular EVALOC has been assessing and understanding the changes in energy use. The research team is made up of building science experts and social scientists. EVALOC received £1.14 million funding over four years from the ESRC-EPSRC Energy and Communities stream of the Research Councils UK Energy Programme from January 2011 to December 2014 (Grant reference: RES-628-25-0012). The case study communities were drawn from DECC's Low Carbon Communities Challenge.

The six communities being studied have a range of low carbon interventions characterized as:

[2] www.evaloc.org.uk

- **building demand** (e.g. fabric or system retrofits),
- **building supply** (e.g. solar thermal installations),
- **behaviour change** (e.g. smart meters) and
- **community renewables** (e.g. large scale PV).

The project is looking at the impacts of the LCCC projects at household and community levels. The project makes use of DECoRuM[3], an energy performance and carbon emissions mapping tool.

7.2.2 Collection and Data/Evidence

Quantitative data has been collected at the household level through energy surveys and trials of using energy and environmental displays, and on how personal social networks are used in connection with the goal of energy and carbon reduction.

Households were divided into three groups: Group A received a physical intervention, either demand or supply side; Group B received a behavioural intervention such as an energy display; Group C received no interventions and acted as a control group.

Three data sets have been reported to the GCB project:

1. **Energy performance data** of 50 owner occupied houses
2. **Quantitative and qualitative social surveys** of 90 households across the six case study communities
3. **Heating controls questionnaire survey**.

In data set 2, the 30 Group A participants had energy, environmental and interaction data collected every 5 minutes for a minimum of 12 months. For all three groups, occupancy feedback was gathered through semi-structured interviews, walkthroughs, social network analysis and heating control self-completion questionnaires. Group A also had activity logging sheets and thermal comfort diaries.

7.2.3 Processing and Information

The quality of the information processing, publication and peer review was judged by EVALOC's Principal Investigator to be at Level 3, representing a processing algorithm that was published and peer reviewed, results that were published and available, and peer review of results of a standard suitable for an academic journal.

Interim results have been presented to the Sustainable Building and Construction Conference held in Coventry on 3-5 July 2013 and to the PLEA Sustainable Architecture for a Renewable Future Conference held in Munich on 10-12 September 2013.

Six working papers and reports are available from the EVALOC website (www.evaloc.org.uk) as well as a refereed journal article[4]. One community event report is also listed but cannot be downloaded. Twenty conference and workshop presentations are listed on the EVALOC Publications page. In September 2012, a major international energy and communities conference was organised in Oxford which attracted 80 delegates (by invitation) from academia, policy-making, community groups, community practitioners and industry. A conference report and associated presentations are available on the EVALOC website[5]. .

Within the project, households taking part in the 6 case study projects have or are planned to have access to information about the thermal performance of their houses. This includes thermal images

---

[3] www.brookes.ac.uk/business-and-employers/new-technologies/decorum%C2%AE/

[4] Gupta, R., Barnfield, L., and Hipwood, T., (2014) Impacts of community-led energy retrofitting of owner-occupied dwellings, *Building Research & Information*, 42:4, 446-461, DOI: http://dx.doi.org/10.1080/09613218.2014.894742

[5] http://www.evaloc.org.uk/#!evaloc-international-energy--communities-conference/c2115

of their property and a web tool for Group A houses to see their energy use in near-real time and environmental conditions in 30-minute slices, and to benchmark their energy use against appropriate benchmarks.

Members of all 20 LCCC groups have had access to the EVALOC results and subsequently have been able to baseline their own emissions and assess the impacts of the LCCC in their own homes and communities.

Issues and Interventions

The mass of data could be used to answer further questions. For example, the research has looked at the use of mechanical ventilation. In order to do this, internal temperatures and data on window- and door-opening were recorded. It would be possible to ask additional questions about when windows, or doors, are opened; at what temperature this is done. However, in order to make the data usable for other researchers, the data would need to be prepared, cleaned and put alongside contextual data.

Principal Investigator on EVALOC, Prof Rajat Gupta, describes the preparation of data for re-use as ***data curation***. This would include **cleaning the data** and seeking out anomalies. These could occur, for example, where monitoring equipment has been switched off by an occupant creating zero readings and could led to misleading conclusions. Data curation would also involve recording **contextual information** and bringing together data sets. For example, the typography and occupancy of a dwelling, and the nature and timing of interventions would need to be recorded alongside the meter readings for that dwelling in order for the meter readings to be meaningful.

This data curation process would take a significant amount of time, currently not budgeted for in research grants. Rajat has considered including this time but is concerned that grant application assessors may not consider his proposal as good value for money unless they **understand the benefits of the curated data** for future/further use.

Rajat suggests that to ensure wide reusablity and consistency, **ground rules** should be established for the process and output.

7.2.4 Cognition and Knowledge

EVALOC is an action research project; the research team has organised 18 focus groups and 18 community events to talk about their work, as well as to collect data.

Community groups have had access to data and have used this to increase their outreach locally. This has included use of the DECoRuM maps to identify dwellings with potential need for retrofit measures, making personal visits to the residents and recruiting them for interventions. There was a 30% uptake, greatly higher than local authority schemes.

Policymakers are using the findings of EVALOC. The project is cited as an exemplar of University-Community collaboration in the current National Community Energy Strategy produced by DECC[6].

Consultants are not yet using the data. There has been reluctance on the part of community groups to share it with such organisations.

7.2.5 Application and Behaviour Change

As the EVALOC project has been concerned with evaluating some of the low carbon community projects already being funded by DECC, it was not expected that the project itself would directly lead to behaviour change. However, the quantity of data and information collected during EVALOC could provide a resource to others. Also, the relationships that the EVALOC project team has established

6 https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/275163/20140126Community_Energy_Strategy.pdf

with the case study communities could be continued through some follow-on work, which could be designed to reinforce any behaviour change that has occurred or to encourage behaviour change where this has not happened to the extent originally intended by the low carbon community projects.

7.2.6 Conclusions/Lessons

Data requires preparation in order to be made useful for addressing further research questions. This data curation includes cleaning data and recording it alongside contextual information.

Data curation needs to be funded. This could be done within the initial research grant or by a fund set aside for such activities (akin to the knowledge exchange projects).

Research councils could be lobbied by the GCB to recognise the value of data curation and to set aside funding for it.

Ground rules should be established for data curation in order that the data sets are widely reusable.

In the hands of trusted messengers (in this case, community members and university researchers), data is very persuasive and can engage house owners where other schemes have failed.

7.3 Airtightness Test Results

7.3.1 Summary

Airtightness testing is part of *Building Regulations* (*Part L*) compliance since 2006, and is mandatory for all non-domestic buildings, large extensions and on a sample of dwellings on every housing development. For both non-domestic buildings and domestic dwellings, an air permeability of no more than 10 $m^3/(h.m^2)$ must be achieved.

The procedure for airtightness tests is defined by the British Institute of Non-Destructive Testing (BINDT), while companies offering testing services are quality assured through membership of the Airtightness Testing and Measurement Association (ATTMA). Membership of ATTMA requires technical training to be in place to the satisfaction of United Kingdom Accreditation Service (UKAS).

The UKAS process checks the airtightness capabilities of companies applying to be members on application, but once accreditation is achieved there is no ongoing monitoring of service delivery. Around 18 companies offering national air tightness services are members of ATTMA, but there are many hundreds of smaller companies offering domestic airtightness tests who are only members of the BINDT.

7.3.2 Data collection and evidence

No national or accessible database currently exists. Results are held by testing organisations such as NHBC, BSRIA, Stroma and other ATTMA members. Much data that exists has been collected since 2006 but there are older test results that could be included for time evolution. Older data is held by BRE and CLG.

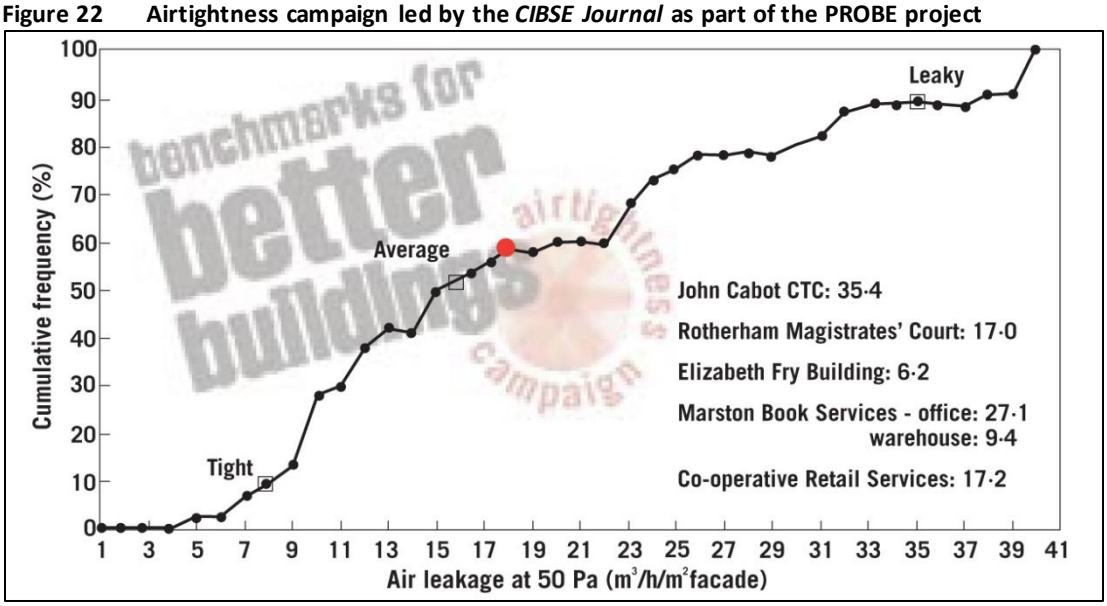
Companies delivering airtightness services gather and store their own data, and treat it as commercial in confidence. There is no sharing of data sets, and no concatenated airtightness data set, either for all buildings type or separated by typologies (for example to the categories adopted by Display Energy Certification).

However, there is a precedent for such a dataset. In 1997 BSRIA and the BRE collaborated on sharing data to create a benchmark dataset against which pressure tests conducted under the PROBE research project could be compared. The results of pressure tests were compared to the reference data set, and published bi-monthly in the CIBSE Journal – see Figure 18.

This collaboration and resulting publicity over the extent of the fabric air leakage problem led, in no small part, to airtightness testing being included in the *Building Regulations.* This is evidence that

given enough motivation and a shared cause, organisations can collaborate and share data to mutual benefit, and for the education of the construction industry and its regulators.

Figure 22 Airtightness campaign led by the *CIBSE Journal* as part of the PROBE project



7.3.3 Transparency of data

BSRIA's view is that a mandatory lodgement process for airtightness test results (similar to that adopted for Display Energy Certification) would be needed for data sharing to the extensive, effective, meaningful, and broad in scope. However, there is little enthusiasm in Government for such a move. It is unlikely that ATTMA would introduce a voluntary scheme in the absence of Government patronage.

Furthermore, airtightness data sharing and transparency would need to be supported by building clients. However, clients are not calling for it as they are largely only interested in achieving certification for their buildings, and not interested in sharing or communicating the reasons for an airtightness test failure to anyone else.

This is reinforced by the airtightness certification process (a system moderated by ATTMA), which provides clients with a form of reporting that is consistent and to an acceptable standard. Although the physics behind the testing process is arcane and not covered by any specific guid ance or standard procedures (airtightness testing is regarded by practitioners as more of a 'black art' than a science, particularly when testing complicated building geometries and building types), it is of no interest to clients nor to industry regulators.

7.3.4 Peer review

Peer review of airtightness results is regarded as a good thing in itself, as it would demonstrate that buildings are being improved as a consequence of the air leakage testing process. ATTMA is well placed to conduct peer review, but there is no call for it, neither by clients, by industry nor by the DCLG, the government department responsible for the *Building Regulations*.

One objection to peer review (and data transparency) is that it may be consequently easy to determine which organisation conducted which airtightness tests, and therefore work out market share. This information could be commercially risky, and therefore testing companies would avoid sharing such information.

Airtightness testing has appeared elsewhere in our database of data sets:

- Crest Nicholson are taking part in both the TSB BPE programme and the Zero Carbon Hub's 'Design versus As-Built' project. Both include airtightness testing

- Greenoak Housing Association has undertaken its own research project, *Greenhomes*. Airtightness testing was carried out on the dwellings studied in order to understand their energy efficiency and occupant comfort performance.

7.3.5 Barriers

Although measures are in place to ensure that the data resulting from the airtightness tests for *Part L* compliance should be of good quality, there are some reasons to suspect bias in the resulting databases. First, a sample of dwellings on a housing development is tested. It would be possible to select those most likely to meet minimum requirements to be tested. Therefore the resulting data set might not be representative of the whole development site.

Second, there is evidence that measures are taken to ensure the pass level is reached (e.g. use of sealant to reduce uncontrolled air leakage during the test) but that these do not necessarily have longevity. Therefore, the airtightness test results do not reveal the build quality or performance of the building fabric over the long term.

Suggested Interventions

- Contributing test organisations could be asked to **remove statistically spurious or otherwise suspicious results**.
- **Spot checks** could be carried out by a third party on tested housing sites.
- **Retesting** could be carried out on a number of buildings over the course of time to plot the deterioration of remedial actions on air leakage.
- Results would need to be **anonymised** to avoid commercial interests being compromised and to ensure broad sector support for such a database. Bring the key ATTMA members (BSRIA, BRE, NHBC and Stroma) together to agree a strategy to do this.
- **Government requires submission** of results to a national database as part of regulation.

7.3.6 Processing and Information

Some test results have been published, e.g. by the NHBC in their own publication [7]

Suggested Interventions

- Test results could be **published periodically** in order to track the performance of buildings and build quality.
- Using **measurements made over the course of time** on a sample of buildings, the typical deterioration of air permeability could be calculated and adjustments made to the database when needed, depending on the enquiry being made.

7.3.7 Cognition and Knowledge

Airtightness tests are being used in a number of projects to help understand the performance of buildings. Alongside co-heating tests, the results are indicative of the build quality and energy efficiency. Monitored energy use for an occupied building will be dependent on a number of factors including occupancy levels, behaviour, occupant comfort levels, heating (and cooling) systems performance, unregulated energy use, lighting and the building fabric and build quality.

Airtightness tests allow some disaggregation of the latter two effects from the rest.

7.3.8 Conclusions/Lessons

Airtightness data has the potential to be much more widely used than is possible at present. The data is currently dispersed amongst the airtightness operators, but could be collected or mirrored

[7] Air Leakage Testing and EPCs, Standards Extra, No 41, p 4, May 2008, National Home Building Council, Milton Keynes

centrally. This would not only provide a useful resource for the research community, it would also provide a helpful indicator of compliance with the relevant part of the Building Regulations.

GCB could lobby DCLG to legislate for this, although the existing costs to access EPC and DEC data must be avoided.

7.4 National Energy Efficiency Database, NEED

Here is an abbreviated version of thorough analysis. The full version of this case study can be found in Appendix H Full Case Study: Under-exploited data set: NEED.

7.4.1 Summary

NEED is the National Energy Efficiency Data-Framework, controlled by DECC. It brings together data sets of energy use, energy efficiency measures, property attributes and household characteristics in domestic and non-domestic buildings in Great Britain and contains records relating to over 27million buildings.

It contains data derived, largely, from administrative sources and it is quite widely used within DECC for policy development, analysis and evaluation.

While limited subsets of NEED data have been made available for a few DECC-mandated research projects in the past, there is currently no mechanism for accessing NEED data outside of DECC. However DECC intends to release anonymized NEED micro-data in the summer of 2014.

7.4.2 Data/Evidence and Data Collection

Data for NEED has been collected since 2000, although the framework itself has only been in place since 2009. NEED brings together data from four primary sources (three for non-domestic) each linked to an address "spine" (based on Ordnance Survey *AddressBase*):

- **Electricity and gas consumption** from energy suppliers
- **Property attributes** (type, age, size) from the Valuation Office Agency
- **Household characteristics** from Experian (occupying organization characteristics such as turnover, employment, SIC code, type of premises for non-domestic)
- **Energy efficiency measures** from the Homes Energy Efficiency Database from the Energy Saving Trust (only for domestic properties).

There is a proposal to add data from the EPC Register. This would add extra detail about building characteristics that is not provided by the VOA dataset.

NEED does not contain information relating to:

- **Very large consumers** of energy, such as power stations
- Properties in **Northern Ireland**
- **Non-metered fuels** such as oil and coal
- **DIY energy efficiency** measures that are not recorded on HEED.

Issues & Interventions

One of the most significant issues in NEED is the lack of standard methodologies for collecting data across, and sometimes within, the various data sets. For example, data within HEED were generated by energy supplier obligation schemes with no consistently applied standards of data collection or categorization.

While it is not feasible to ask the HEED data providers to correct their records, it may be possible to analyse the known issues and attempt to provide correction factors for key data variables. Alternatively, some data issues could be more easily clarified. For example, the issue with the weather adjustment of gas consumption is that the data provider does not provide complete information regarding its methodology. DECC or Ofgem could mandate that this be documented.

7.4.3 Data Processing

As can be seen from the diagram in Appendix H, the process to link and integrate the various data sets in NEED is complex.

Issues & Interventions

DECC provides useful summaries of its processing methodologies at the various stages of data processing but it often does not provide enough detail to be useful in informing detailed academic research. This could be overcome by persuading **DECC to produce more detailed descriptions of data processing methodologies**.

7.4.4 Information

Aggregated data (Excel) and reports are published on the NEED web pages at www.gov.uk/government/collections/national-energy-efficiency-data-need-framework.

In addition to the analyses that are contained in the regular NEED reports, the web site also contains a pivot-table creator to customise data views on domestic energy consumption. This allows two main variables to be analysed together (property age, property type, number of bedrooms, floor area, tenure, income, number of adults and region). Other parameters can also be set for the type of data to display (mean figures, median figures, or number of observations), the fuel to analyse (gas or electricity) and the year to show (2005 through to 2011). All the summary data that the table creator uses can also be downloaded as a spreadsheet file in csv format.

DECC has undertaken a consultation on making some NEED micro-data available. The proposals are for a limited anonymised data set of approximately 20,000 records to be available to anyone (expected end of May 2014), and for a larger data set of approximately 4 million records to be available through an end-user licence (expected by the end of July 2014).

Issues & Interventions

The critical issue here is the lack of access to NEED micro-data which may be partially resolved with the release of anonymized NEED data sets in the summer of 2014.

However it should be noted that a further intervention (the provision a non-anonymized or pseudo-anonymized dataset which would enable linking to other data sets) would significantly improve the usefulness and usability of NEED data.

7.4.5 Cognition and Knowledge

NEED has been used to generate specific reports on energy consumption and the impact of energy efficiency measures. These are published on the NEED web pages. Analyses carried out in 2012 and 2013 include

- Analysis on the impact of installing energy efficiency measures on a household's gas consumption
- Analysis of domestic gas and electricity consumption by property attribute and house hold characteristics
- Analysis of the quality of non-domestic energy consumption data.

Issues & Interventions

There are no significant issues with NEED publications. They are published as official statistics and so can be said to have undergone some form of peer review. Release of non-domestic NEED analysis would be welcomed.

7.4.6 Application and Behaviour Change

As previously noted, NEED has been well utilised within DECC for policy development, analysis and evaluation.

Issues & Interventions

The knowledge and insight gained from NEED has undoubtedly been applied in relation to policy evaluation and development. The only issue is that without access to data outside of DECC, the benefit of generating knowledge and behaviour change outside of a policy context is significantly reduced.

7.4.7 Conclusions/Lessons

NEED is a very valuable data resource that is currently only utilised within DECC but has the potential to be of great benefit to the research community and other stakeholders in the energy sector.

The key issues that should be addressed are:

- Appropriate access to NEED micro-data - going beyond the current proposal for two anonymized data sets.
- Uncertainties and data quality issues arising from inconsistent or non-existent data collection methodologies.
- Uncertainties arising from data linking and other data processing

There are also barriers to the suggested interventions:

- DECC does not own any of the data but has negotiated access to the data sets. In theory this access could be withdrawn at any time
- DECC is likely to be concerned with negative publicity generated if any NEED data is lost or stolen
- The data providers to NEED are inclined to only do the bare minimum required by the data sharing agreements
- DECC is constrained by decreasing resources. Any intervention is likely to require additional effort and funds, which could be difficult to obtain in the current economic climate.

7.5 Display Energy Certificate data

7.5.1 Summary

Display Energy Certificates have been compulsory for public buildings since 2008 – they were originally required for buildings with a floor area greater than 1000m^2 but are now required for buildings with a floor area greater than 500m^2.

7.5.2 Information

The DECs are lodged on the register administered by Landmark. Individual reports may be retrieved using the report number or the postcode of the building but these are only for personal use. There is no publicly available set of DEC data from the register administrators.

However, the Centre for Sustainable Energy has obtained three years of DEC data using Freedom of Information requests submitted to DCLG, for the years 2008, 2009 and 2010. The data set can be downloaded in Excel format from the CSE web site (www.cse.org.uk/resources/open-data/display-energy-certificate-data). Unfortunately, the DEC data available from CSE is not referenced back to the UPRN (the Unique Property Reference Number that provides a reference for each building in the UK). Therefore there is no easy way to compare DEC data from year to year. The only comparison that can be made is using postcode, address and occupier details. The DEC data has also been used to produce a Google Earth plug-in by Carbon Visuals which overlays CO_2 emissions per building on a UK map (www.carbonvisuals.com/work/google-earth-uk-public-buildings).

Issues & Interventions

Because DECs for buildings over 1000m^2 have to be renewed every year, this sub-set of the data has the potential to grow into a valuable time series data set. But the recast of the regulations establishing DECs created a disconnect between larger and smaller (500m^2 to 1000m^2) buildings where the DEC is valid for 10 years.

On its own, the DEC data set does not contain enough contextual information to make detailed comparisons between different buildings. For example there is no information about occupancy or operational hours of the surveyed building and these are details that can easily change over time. In the same way that NEED brings together different data sets to enable more sophisticated and helpful analysis, the occupancy and operational data could be obtained through a separate data set provided the two could be linked.

In addition, there are significant concerns within the research community about the cost of obtaining DEC data. Individual records can be downloaded from the Landmark database for free, but bulk data is charged for on a per-record basis. It is estimated that the cost to obtain the complete DEC data set would be in the order of £1million, which is clearly far beyond the reach of an academic institution.

7.6 Stamford Brook

7.6.1 Summary

Stamford Brook is a housing estate built on National Trust land at Dunham Massey, Cheshire. Partners on the scheme were the housebuilders, Redrow and Taylor Wimpey, and the National Trust. Together, they formed an Environmental Performance Standard for the new dwellings, with targets to exceed the requirements of the current building regulations. The Leeds Metropolitan University team carried out an action research project and was involved in the development from 2001 for more than eight years.

The energy part of the Environmental Performance Standard was for an improvement on both 2002 building regulations and the then future standards for 2006 building regulations.

This was a deep study of a small sample.

7.6.2 Collection and Data/Evidence

Throughout the planning, design and construction phases, the Leeds Met team was both observing and reporting on the progress of the development. Evidence collected was both quantitative and qualitative and included

- **airtightness** tests,
- **internal temperature**,
- **thermal imaging**,
- **co-heating** tests to measure whole house heat loss

Heat loss was found to be higher than expected. One house was studied in great detail and measurements and calculations found that the total house heat loss coefficient was higher by 60WK^{-1}. This represented a considerable difference from the predicted heat loss from the dwelling.

7.6.3 Processing and Information

Evidence was analysed alongside modelling results to reveal heat losses via airflow through the party wall cavity, thus bypassing insulation. Previously, it was assumed that heat loss through party walls was zero. Research was now showing that it was greater through the party wall than the external walls.

These finding were published in the journal *Building Services Engineering Research and Technology* in 2007[8]. Authors concluded that changes should be made to calculating processes for the U-values of dwellings (BS EN ISO 6946, the SAP 2005, BR 443, Dwelling CO_2 Emission Rate) and to documentation for accredited construction details to reflect the heat loss mechanism identified by their studies.

7.6.4 Cognition and Knowledge

There was existing evidence for the cavity wall bypass mechanism in the USA, albeit in a differing construction type, and a similar effect had been identified in the UK in the 1990s. However, calculating methods for dwelling heat loss were not reflecting this and no technical solutions had been developed and put into place. Following the publishing of the results of the Stamford Brook project, and further studies, changes were made.

This research was also significant in recognizing that the dwellings were losing more heat than designed for and in measuring this shortfall. Subsequently, this has become known as the *performance gap* and led to further studies, lively discussion within the housebuilding industry and policy change. In 2010, the Zero Carbon Hub reported on its evidence review of building performance research[9] and confirmed the significance of the discrepancy between building performance in actuality and as designed, and set out of plan of action for further research. A current major project undertaken by the Zero Carbon Hub follows on from this (see 8.3 Joining Data Sources – Zero Carbon Hub's Design vs As Built Project) and has sought to identify all sources of the performance gap, prioritise these for action, and now seeks solutions. This will inform Government and the whole industry.

7.6.5 Application and Behaviour Change

Building Regulations now assign a non-zero U-value to an unfilled cavity party wall ($0.5 Wm^{-2}K^{-1}$); this represents a significant penalty and potential to fail a SAP assessment. Design solutions and construction materials have been developed and tested to ensure an improved the thermal performance and a lower U-value.

The findings arguably raised awareness of many issues around the design of buildings, the supply chain, and the quality of building housebuilding and the impacts of these on the energy performance of our homes as a major source of CO_2 emissions. As well as the work of the Zero Carbon Hub, this has led to

- **further research** (including the Technology Strategy Board Building Performance Evaluation Programme (see 8.2 Creating a Large, Accessible Database – EMBED);
- **campaigning** (by, among others, the Good Homes Alliance and UKGCB)
- and **regulation change**.

7.6.6 Conclusions/Lessons

In-depth studies carried out on a small sample allow a flexible, adaptable research and this, in turn, can led to an understanding **not only of what** happens but also **why**. A number of studies can be brought together to create a statistically significant sample and led to robust evidence and deep understanding.

---

[8] *Evidence For Heat Losses Via Party Wall Cavities In Masonry Construction* BUILDING SERV ENG RES TECHNOL **May 2007** vol. 28 no. 2 **161-181**

[9] Carbon Compliance For Tomorrow's New Homes - A Review Of The Modelling Tool And Assumptions: Topic 4 - Closing The Gap Between Designed And Built Performance, August 2010, Zero Carbon Hub

7.7 Broad data sets vs Deep investigations

The data sets identified through this project represent a broad spectrum of sizes. Some data sets contain data relating to tens of thousands, hundreds of thousands or even millions of buildings while others contain data relating to just one or a few buildings. We identified the majority of the **best data sets**, in terms of our scores quality and usefulness, as being of the latter sort. We identified a number of data sets with **good potential** for being very useful with some intervention and these tended to be large, broad data sets.

Data sets can be characterized by their breadth and their depth. Typically, data that is broad tends to be shallow (collecting a limited number of items for many buildings) while data that is deep tends to be narrow (collecting a large number of items for a few buildings).

Examples of broad data sets include: electricity and gas meter data (over 35 million and 22 million meter points respectively), the Homes Energy Efficiency Database (over 15 million UK homes), the Energy Performance Certificate register (over 9 million certificate records).

Examples of deep data sets include: Leeds Met's studies of Stamford Brook, Elm Tree Mews and Temple Avenue (identified in the current piece of work), as well as Milton Keynes Energy Park trials (8 houses monitored for 2 years), Domestic energy use study (70 households), Domestic Energy Management System trial (6 households). Many deep data sets seem to have been generated out of individual research projects.

7.7.1 Pros of broad data sets

The advantages of broad data sets are:

- **Statistical significance**
- **Comparability across different characteristics**: regions, building types, occupant
- **Comparability over time** series (provided collection methods are compatible).

7.7.2 Cons of broad data sets

If broad data sets are also shallow, their disadvantages can be:

- **Little or no contextual detail** to support the core data
- **Lack of control** over data collection or knowledge that collection protocols have been followed
- Core data is almost always **quantitative** to minimize collection costs.

Some broad data sets are created to demonstrate compliance with legislation or adoption of government policies, and may not take account of the further uses to which the data could be put. For example, HEED records basic property characteristics at an address along with information about the heating system, insulation installed and any micro-generation technologies installed. On its own, HEED cannot say whether installation of specific energy efficiency measures has resulted in lower energy consumption. That information can only be provided if HEED is combined with meter reading data or similar energy consumption data. At present this is done through NEED (National Energy Efficiency Database Framework) which is a composite data set bringing together different data sources – see Section 7.5 for more information about NEED. Even then analysis of a single address would need to assume that the occupiers had similar lifestyle patterns before and after an energy efficiency installation.

7.7.3 Possible improvements to broad data sets

The principal concerns regarding the broad data sets identified in this project have been about lack of access to the detailed data and the lack of assurance and transparency that data has been collected or processed consistently. These concerns could be addressed through **mandated access and targeted investigations** to research and publish the collection and processing methodologies.

7.7.4 Pros of Deep data sets

The advantages of deep data sets are:

- **Lots of contextual detail**
- **Mixture of quantitative and qualitative data**, or a focus on qualitative data
- Data set often designed and **collected by a single team**, giving more confidence that collection is consistent.

In the case of the Stamford Brook study, only a few dwellings were studied. This allowed the investigation to be adapted throughout the study in order to find the causes for some of the initial, unexpected results. This led to the identification of the cavity party wall thermal bypass mechanism, which was significantly compromising the thermal performance of the buildings. Once this phenomenon was brought to light, there was impetus for further investigations and eventually important changes to building regulations, calculating procedures for thermal properties of whole buildings and a greater understanding of building performance.

7.7.5 Cons of Deep data sets

If deep data sets are also narrow, their disadvantages can be:

- **Lack of statistical significance** (i.e. the data set may itself be an outlier of the broader population)
- **Difficult to apply to other situations** because of the importance of the specific contextual information. For example if a study is looking at energy consumption in 10 identical dwellings, but focusing on the occupancy differences, then the basic building information may not have been captured at all. Collecting the building information could have allowed the data to be combined with results from other studies to produce a more statistically significant sample for other research.

7.7.6 Possible improvements to deep data sets

Multiple research projects sometimes appear to be collecting the same kind of detailed information. However, there is no guarantee that the studies can actually be compared or agglomerated. For example, if Study A reports energy consumption per square meter for one type of dwelling or occupant and Study B also reports energy consumption per square meter but for a different type of dwelling then these studies can only be compared if the unit area consumption calculation is the same, or can be made compatible from other data. This would require that **measurement protocols are compatible**, and project timescales are similar.

Another potential improvement would be for **research teams to collaborate more often** to harness opportunities of sharing of data by understanding what additional data items may be collected and used by others.

7.8 Recommendation for next step

Given the number of data sets that this study has identified, then one potential step towards finding out which publicly accessible data sets could be **combined into more useful composite data sets** would be to investigate their data structures in detail. Such an exercise should be able to determine the ease or difficulty of combining different data sets and in particular to explore the possibility of matching particular deep data sets with particular broad data sets to obtain a richer picture than either can give on their own.

8 Best Practice Case Studies

This section of the report sets out case studies where data collection, information processing and knowledge transfer have been successfully used to generate actions and change. Each case is framed by the knowledge staircase described above.

8.1 Using Data to Incentivise - Sustrans in Workplaces

8.1.1 Summary

Sustrans is a charity working to promote sustainable travel, predominately cycling and walking, to various communities and stakeholders. The organisation is responsible for the creation of the National Cycle Network. A notably successful project has been *Connect2 UK*, winning an Ashden Award in 2013. Connect2 UK extended the National Cycle Network so that is runs from housing to places of employment and community facilities such as schools and doctors' surgeries, the ambition being to replace essential journeys made by car with cycle journeys.

On its website, the charity cites an array of documented research. This informs its interventions and the overall targets that Sustrans has for its activities. This includes Sustrans' call for Government to invest in **doubling the number of journeys of fewer than 5 miles being by foot bike and public transport**.

In this case study, we consider the work that Sustrans undertakes with employers to encourage sustainable travel to/from and for work.

8.1.2 Collection and Data/Evidence

Sustrans reference a large body of academic research, a list of which is on their website: www.sustrans.org.uk/policy-evidence/related-academic-research. The coverage of this work is broad and includes environmental, health and economic benefits of cycling and walking, as well as investigations of the barriers to greater uptake of both modes.

Data is collected and used by the charity during its own activities. For example, on the National Cycle Network, monitoring is carried out (with automatic cycle counters) and user surveys undertaken. Using this data, Sustrans are able to estimate that 720,000 journeys per year that would have been made by car were made by bicycle or foot on the Connect2 UK routes and that 70,000 tonnes CO_2 per year are saved[10].

During the workplace projects, participants are incentivized to make their travel to and from, and for work more sustainable. Online tools allow them to enter information about their journeys and a competition means they take an interest in their own and everyone else's results. The data can also be used to assess the success of the activity for the employer and Sustrans.

8.1.3 Processing and Information

Academic research (with Sustrans work being the subject in some cases) has led to a number of key findings that Sustrans make use of including:

- There are numerous health benefits from walking and cycling to work[11]
- Cyclists report fewer sick-days than non-cyclists[11].
- Behaviour change can be achieved in workplaces using travel planning[11]

To quantify the benefits of its work, Sustrans has used the Department of Transport's WebTAG[12] appraisal methodology within which a monetary value is assigned to the social, economic, health

[10] Ashden Award Winner case study, www.ashden.org/winners/sustrans13

[11] www.sustrans.org.uk/policy-evidence/related-academic-research/sustainable-travel-everyday-destinations

[12] Transport Analysis Guidance - https://www.gov.uk/transport-analysis-guidance-webtag

and transport benefits derived from investments. An overall benefit:cost ratio (BCR) is then calculated. For the National Cycle Network, various benefits have been realized including generation of economic opportunities, provision of low carbon travel and as a means of education [13]

8.1.4 Cognition and Knowledge

Sustrans works with employers to increase active travel to workplaces and realize the benefits described by research, as well as reducing greenhouse gas emissions. They offer a number of interventions, which are usually deployed in combination. These include:

- The creation of personalized travel plans, enabling employees to have the best route in terms of time, distance, safety and CO_2 savings.
- Seeking travel champions. These are volunteers in the workplace that are able to offer leadership and encouragement to their colleagues.
- An online Sustainable Travel Challenge incentivizes active, e.g. for 2012 www.getmeactive.org.uk/content/about

8.1.5 Application and Behaviour Change

For the workplaces schemes, Sustrans is able to claim that for every 100 employees that they are able to engage in their programme, the employer typically saves £7500 a year and achieves CO_2 savings of 38 tonnes per year[14], there is a 6% reduction in commuting by car and improved well-being in 61 employees[15]

On the national scale, the charity is lobbying governments for investments in infrastructure that would allow a doubling of journeys by sustainable means. Melissa Henry (Sustrans) told us 'Evidence suggests that many more short journeys could be made by foot, bike and public transport rather than car. With the right investment in infrastructure and behaviour change, the evidence suggests that as many as 9 in 10 local journeys could be made by foot bike and public transport.'

8.1.6 Conclusions/Lessons

Data, in various interpreted forms, can be used to engage various stakeholders:

- Employers understand the cost benefits of encouraging sustainable travel.
- Employees understand the health and cost benefits to them of sustainable travel.
- A competition incentivizes choosing sustainable travel options.
- An online competition encourages submission of data (regarding the sustainable behaviour).
- Plotting progress gives confidence to funders and encourages further funding. The Ashden Award is based on proven success and it has been won by Sustrans for work on the National Cycle Route.

8.1.7 Also see

CarbonCulture[16] works with organisations to encourage sustainable behaviour in workplaces to reduce energy use and CO_2 production. They use online games to encourage employee participation and collect data. Data collection enables tracking of progress and savings can be measured. Organisations are provided with CarbonCulture webpages were data is reported in real time. Such transparency is significant for government participants including 10 Downing St, DECC and Foreign and Commonwealth Office. It enables them to share learning and demonstrate leadership.

---

[13] World Transport Policy and Practice, 2003, Volume 9, Number 1, pages 6-17

[14] Cutting Business Carbon Emissions Through Sustainable Travel, Sustrans

[15] www.sustrans.org.uk/policy-evidence/the-impact-of-our-work/workplace-personalised-travel-planning-derby

[16] https://platform.carbonculture.net

During the course of this project, we learnt from a contributor (Andrew Kinsey, Sustainability Manager, Mace) that **Mace** collect data on their construction projects for a number of reasons. One use is the rating of each project from A-G based on environmental performance. This are displayed internally and a league table is produced to inspire competition between project teams for better performance.

8.2 Creating a Large, Accessible Database – EMBED

8.2.1 Summary

EMBED is an interactive online hub for storing, analysing and sharing data from building energy monitoring studies (https://est.amee.com).

Data can be uploaded for private use or to make it public for partners and others to use.

The ambition is to create the largest and smartest repository for building and technology data from across the UK and the world.

It is supported by the Energy Savings Trust and Technology Strategy Board.

It was developed by AMEE (www.amee.com). Development has now been taken over by Masterdon C (www.mastodonc.com).

8.2.2 Data/Evidence

The existing platform allows building and portfolio managers to upload large quantities of data, including electricity and gas consumption, electricity generation, and environmental sensors including temperature, CO_2, and humidity.

EMBED also allows for contextual information surrounding a project to be uploaded, as well as test results, thermal images and other information considered relevant.

EMBED's current public data covers 69 properties and data from:

- EST's Energy Saving Trust's programme of field trials
- TSB's Retrofit for the Future programme
- TSB's Building Performance Evaluation programme

Any organisation can submit data

8.2.3 Processing and Information

Data is added using an application programming interface (API) or CSV spreadsheet file.

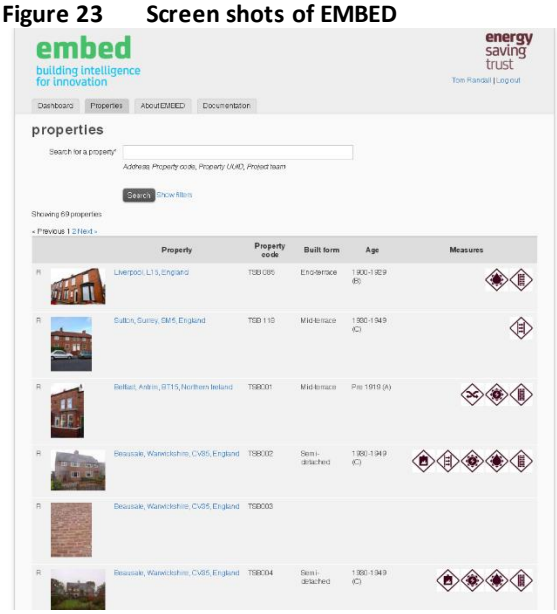
EMBED uses the AMON standard for the description and exchange of metering/monitoring device data (http://amee.github.io/AMON/). This ensures a common standard and allows the secure online transfer of live monitoring data from any device using the AMON open data format.

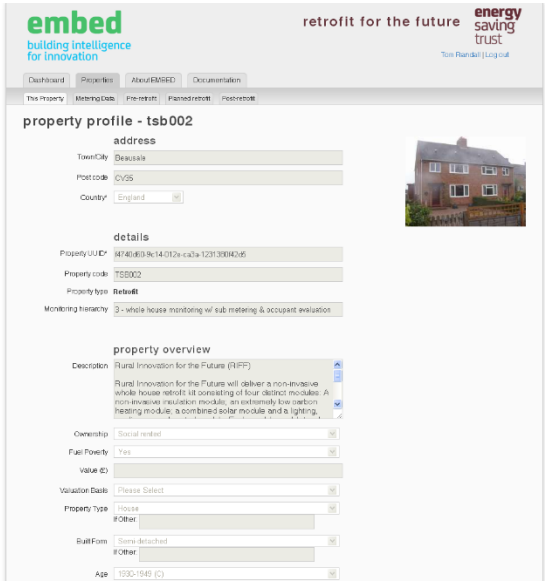
Masterdon C are currently making a number of updates and improvements to processing, including:

- faster and more error-proof data uploads
- automated detection of data quality issues, and flagging to users for correction
- automated calculation of key performance measures, including CO_2 emissions per m^2

These are due for completion May 2014.

Figure 23 Screen shots of EMBED





8.2.4 Cognition and Knowledge

Embed allows you to compare projects and performances E.g.

- Compare how different technologies are performing in different circumstances and learn from the individual projects.
- Search and filter for similar buildings and quickly visualise large quantities of data

Performance can be monitored to identify any emerging trends or issues.

8.2.5 Application and Behaviour Change

Embed currently has over 600 registered users. This includes both those simply wanting to look up data and those uploading data and using Embed for ongoing monitoring.

Know user types include, member of public, architects, engineers project managers, academics, developer and housing associations.

This suggests that Embed already facilitates behaviour change but a detailed study of the impact has not been undertake.

8.2.6 Conclusions/Lessons

EMBED would appear to share the ambition of the Green Construction Board to increase the effective use of data on the energy performance of buildings.

Obvious opportunities for the Green Construction Board are to support the development, population, promotion and use of Embed.

8.2.7 Contacts

Geoffrey Stevens, Technical Project Manager, Energy Saving Trust

geoffrey.stevens@est.org.uk 020 7654 2634

8.3 Joining Data Sources – Zero Carbon Hub's Design vs As Built Project

8.3.1 Summary

The Zero Carbon Hub is a not-for-profit organisation charged with informing Government and the housebuilding industry in order to achieve 2016 targets. A current project is a deep exploration of the *performance gap*. This is the evident difference between the anticipated energy performance of new homes as they are designed and as they are once built. With housing contributing a significant amount to CO_2 emissions and reduction in these emissions being vital to achieving national emissions targets, plus concerns regarding fuel prices, the actual emissions from housing must be understood and quantifiable.

8.3.2 Data/Evidence

In order to understand the causes of the performance gap, a review was undertaken on various pieces of evidence and data in this field. This comprise

- **Literature** - relevant work has come from academia, professional bodies, housebuilders and supporting trades, and included both publicly published and confidential industry research.
- **Housebuilding process review** covering sites of various kinds volunteered by the housebuilders – interviews, design reviews, observations on site.
- **SAP audits** – on site visits, SAP audits to reveal discrepancies in SAP assessments and differences between SAP assessments and observations on site
- **SAP assessor accreditation organisations questionnaire results** – to understand training (including ongoing training) and what is revealed by assessor audits and complaints to helplines.
- **SAP assessor questionnaire results** – to understand how they operate and where they believe the performance gap arises.

8.3.3 Processing and Information

Ten working groups, each considering a different aspect of housebuilding, compiled a list of issues that, from their experience, they believed were responsible for or influenced the extent of the performance gap. 55 issues were raised; these were grouped under the following themes:

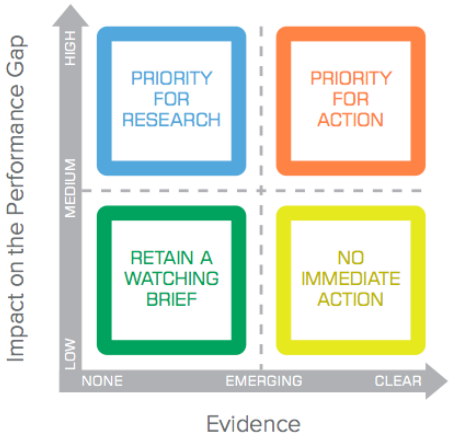
- Land Acquisition, Concept Design and Planning
- Detailed Design
- Procurement
- Construction and Commissioning
- Verification
- Testing
- Energy Tools and Conventions

Relevant literature was then reviewed, evidence for each issue tagged and recorded. Combining this information with results from the site visits, SAP audits and SAP questionnaires, a meta-data set was created. For each issue, the extent of its impact on the performance gap with determined, or estimated where the evidence was weaker.

8.3.4 Cognition and Knowledge

From the list of issues originally identified by the working groups, a list of priorities was created. This was based on the quality of evidence and the number of sources for each issue, and a qualitative assessment of the impact of the issue on the performance gap – as shown in Figure 24 Mapping issues to determine priority

Figure 24 Mapping issues to determine priority



8.3.5 Application and Behaviour Change

The list of priorities determines next steps for the Design vs As-Built project. 15 issues were identified as **priorities for action**. The final report from this project will recommend strategies for tackling these issues.

8.3.6 Conclusions/Lessons

Data and evidence from various sources and of various kinds can be combined to produce a usable meta-database.

Data from small studies, with samples that are not statistically significant in size, can be brought together to produce a robust body of evidence.

9 Interventions to Affect Behaviour Change

This section identifies different intervention types that could be used to promote the affective use of data sets. It identifies:

- An appropriate **behaviour change model**
- A comprehensive set of **interventions**
- How interventions **map against the Knowledge Ladder**
- How interventions relate to **potential policy**
- Interventions the **Green Construction Board may be able to implement**.

9.1 Identifying a Comprehensive Behaviour Change Framework

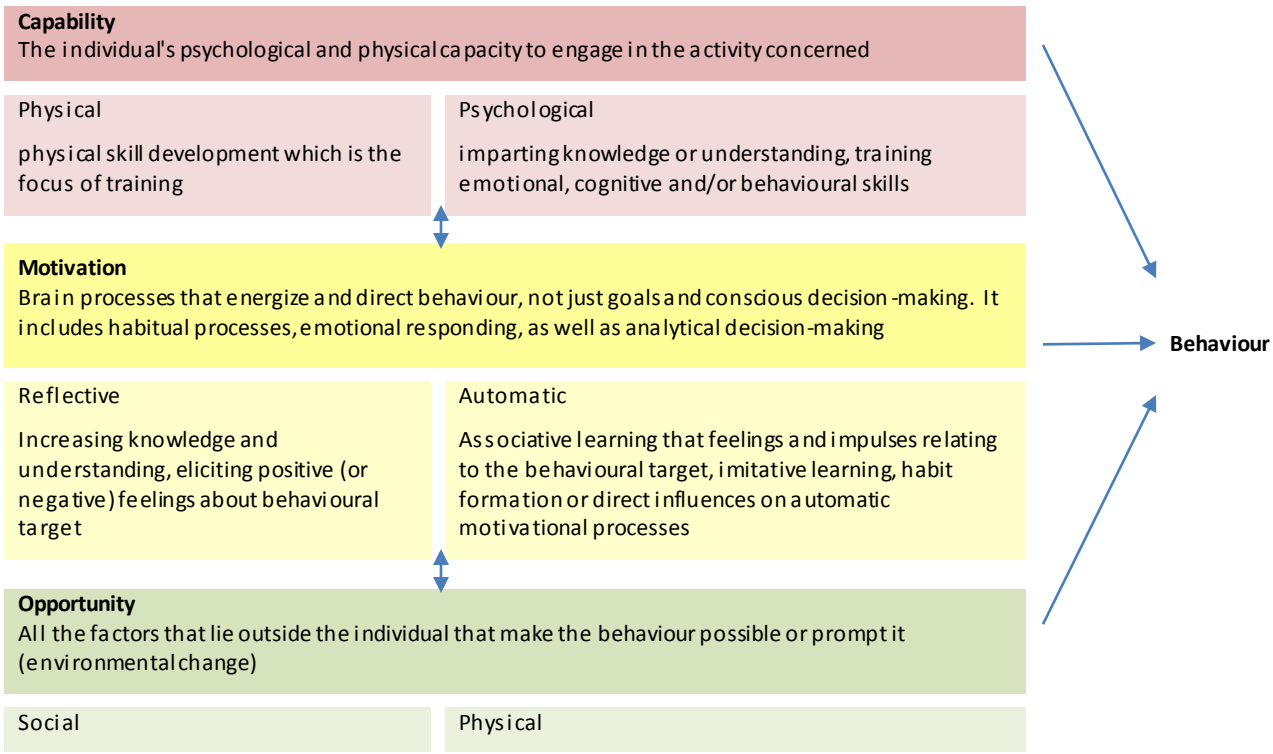
Prior to exploring different intervention types a behaviour change framework was identified that is considered comprehensive in its coverage of different behaviour change mechanisms. The framework includes 9 unique intervention types and also cross-references these with 7 policy categories.

The 'Behaviour Change Wheel'[17] was developed following the review of 19 established behaviour change frameworks, including the Institute for Government MINDSPACE framework and DEFRA's 4E model.

9.1.1 Influences on Behaviour

The wheel identifies 3 fundamental influences on behaviour each with 2 sub-types:

Figure 25 COM – B Framework of Behaviour Influences



[17] The behaviour change wheel: A new method for characterising and designing behaviour change interventions, Susan Michie, Maartje M van Stralen and Robert West, www.implementationscience.com/content/6/1/42

9.1.2 Interventions

The Behaviour Change Wheel also identifies nine unique intervention types and maps these against the behaviour change influences.

- **Interventions are defined** in Table 8, along with examples.
- **Interventions are mapped against behaviour influences** in Table 9.

Table 8 Behaviour Change Intervention Types

Intervention	Definition	Example
Education	Increasing *knowledge* or understanding	• Qualifications • Media communications
Persuasion	Using communication to induce positive or negative feelings or stimulate action	• Lobbying • Media communications • Case studies
Incentivisation	Creating expectation of reward	• Awards
Coercion	Not *prevented* from a behaviour but creating expectation of *penalty* or *cost*	• Regulation • Standards • Contract requirements
Training	Imparting *skills*	• CPD • Qualifications
Restriction	*Prevented* from a behaviour. Using rules to reduce the opportunity to engage in the target behaviour	• Regulation • Standards • Contract requirements
Environmental restructuring	Changing the physical or social context	• Processes
Modelling	Providing an example for people to aspire to or imitate	• Case studies
Enablement	Increasing means/reducing barriers to increase capability or opportunity[1]	• Funding

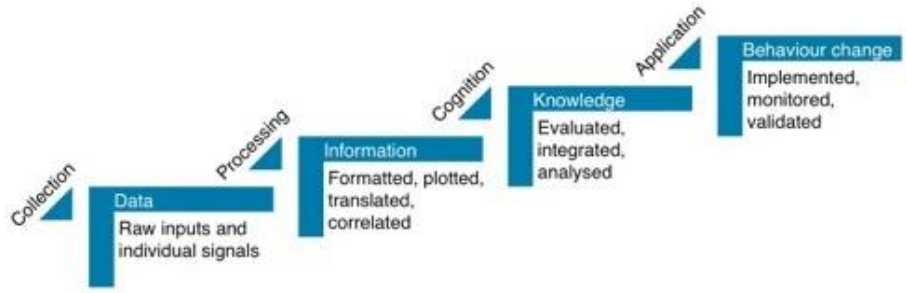
Table 9 Interventions mapped against Behaviour Change Influences

	Intervention Function								
Influence on behaviour	Education	Persuasion	Incentivisation	Coercion	Training	Restriction	Environmental restructuring	Modelling	Enablement
Capability Physical					✓				✓
Psychological	✓				✓				✓
Opportunity Social						✓	✓		✓
Physical						✓	✓		✓
Motivation Reflective	✓	✓	✓	✓					
Automatic		✓	✓	✓			✓	✓	✓

9.2 Mapping the Behaviour Change Framework & Interventions against the Knowledge Ladder

To establish how different intervention types may be used to promote the use of data sets, the nine intervention types identified above were compared against the Knowledge Ladder set out in section 7.1 and represented here:

Figure 26 The Knowledge Ladder



The steps and transitions of the ladder were first compared against the influences on behaviour. This comparison took into account that the three transitions 'processing', 'cognition' and 'application' were processes and the first three steps were objects. The final step in the ladder is itself termed behaviour change, so to avoid confusion this was omitted from the mapping.

- **Capability** is the ability to **do** something and therefore relates to transitions
- **Motivation** is acting, automatically or reflectively to something **learnt / experienced**.
- **Opportunity** can be both the opportunity to **do** *or* **learn** something.

The mapping is shown in Table 10.

Table 10 Behaviour Change Influences mapped against the Knowledge Ladder

Model of Behaviour Source:	Collection	Data Raw inputs and individual signals	Processing	Information Formatted, plotted translated, correlated	Cognition	Knowledge Evaluated, integrated, analysed	Application	Behaviour change Implemented, monitored, validated
Capability Physical	✓		✓		✓		✓	
Capability Psychological	✓		✓		✓		✓	
Motivation Reflective	✓	✓	✓	✓	✓	✓	✓	
Motivation Automatic	✓		✓		✓		✓	
Opportunity Social	✓	✓	✓	✓	✓	✓	✓	
Opportunity Physical	✓	✓	✓	✓	✓	✓	✓	

The relationships established between influences on behaviour and interventions (Table 9) were then used to translate the mapping between behaviour and the knowledge ladder shown in (Table 10) into a mapping between interventions the Knowledge Ladder.

Table 11 Interventions mapped against the Knowledge Ladder

Intervention	Collection	Data Raw inputs and individual signals	Processing	Information Formatted, plotted translated, correlated	Cognition	Knowledge Evaluated, integrated, analysed	Application	Behaviour change Implemented, monitored, validated
Education	✓	✓	✓	✓	✓	✓	✓	
Persuasion	✓		✓		✓		✓	
Incentivisation	✓		✓		✓		✓	
Coercion	✓		✓		✓		✓	
Training	✓		✓		✓		✓	
Restrictions	✓	✓	✓	✓	✓	✓	✓	
Environmental restructuring	✓	✓	✓	✓	✓	✓	✓	
Modelling	✓		✓		✓		✓	
Enablement	✓		✓		✓		✓	

9.3 Interventions resulting from Sustainable Construction Data

Following the establishment of a comprehensive behaviour change framework, generic intervention types and their relationships to the Knowledge Ladder, the following table sets out examples of intervention to promote behaviour change based on high quality sustainable construction data.

Table 12 Interventions mapped against the Knowledge Ladder

Intervention	Collection	Data Raw inputs and individual signals	Processing	Information Formatted, plotted translated, correlated	Cognition	Knowledge Evaluated, integrated, analysed	Application	Behaviour change Implemented, monitored, validated
Education	**Promote** existence of collection techniques in **CPD, qualifications & media comms**	Available data included in **CPD & qualifications**	**Promote** existence of processing techniques in **CPD, qualifications & media comms**	Available information included in **CPD & qualifications**	**Promote** existence of cognition techniques in **CPD, qualifications & media comms**	Available information included in **CPD & qualifications**	**Promote** existence of applicaiton techniques in **CPD, qualifications & media comms**	
Persuasion	**Advocate & highlight benefits** of collection techniques in **communications**		**Advocate & highlight benefits** of processing techniques in **communications**		**Advocate & highlight benefits** of cognition techniques in **communications**		**Advocate & highlight bene**fits of application techniques in **communications**	
Incentivisation	**Awards** for innovative collection		**Awards** for innovative processing		**Awards** for innovative cognition		**Awards** for innovative application	
Coercion	**Require the use** of specific collection in **contracts, standards & qualifications**		**Require the use** of specific processing in **contracts, standards & qualifications**		**Require** the use of specific cognition in **contracts, standards & qualifications**		**Require** the use of specific application in **contracts, standards & qualifications**	
Training	Collection techniques included in **CPD & qualifications**		Processing techniques included in **CPD qualifications**		Cognition techniques included in **CPD & qualifications**		Application techniques included in **CPD & qualifications**	
Restrictions	**Prevent the use** of poor quality collection in **contracts, standards & qualifications**	**Prevent the use** of poor quality data in **contracts, standards & qualifications**	**Prevent the use** of poor quality processing in **contracts, standards & qualifications**	**Prevent the use** of poor quality information in **contracts, standards & qualifications**	**Prevent the use** of poor quality cognition in **contracts, standards & qualifications**	**Prevent the use** of poor quality knowledge in **contracts, standards & qualifications**	**Prevent the use** of poor quality application in **contracts, standards & qualifications**	
Environmental restructuring	Processes **promptedin processes**	Available data **prompted in processes**	Processes **promptedin processes**	Information **prompted in processes**	Cognition **prompted in processes**	Knowledge **prompted in processes**	Application **prompted in processes**	
Modelling	**Case studies** on effective collection		**Case studies** on effective processing		**Case studies** on the effective cognition		**Case studies** on the effective application	
Enablement	**Funding** for innovative collection		**Funding** for innovative processing		**Funding** for innovative cognition		**Funding** for innovative application	

9.3.1 Policy Categories

The work behind the Behaviour Change Wheel also identifies 7 unique *policy category types,* Table 13 and then sets out how they relate to the 9 intervention types identified, Table 14.

Table 13 Policy categories

Policies	Definition	Examples
Communication/ marketing	Using print, electronic, telephonic or broadcast media	Conducting mass media campaigns
Guidelines	Creating documents that recommend or mandate practice. This includes all changes to service provision	Producing and disseminating treatment protocols
Fiscal	Using the tax system to reduce or increase the financial cost	Increasing duty or increasing anti-smuggling activities
Regulation	Establishing rules or principles of behaviour or practice	Establishing voluntary agreements on advertising
Legislation	Making or changing laws	Prohibiting sale or use
Environmental/ social planning	Designing and/or controlling the physical or social environment	Using town planning
Service provision	Delivering a service	Establishing support services in workplaces, communities etc.

Table 14 Policy categories mapped against the intervention functions

	Intervention Function								
Policy category:	Education	Persuasion	Incentivisation	Coercion	Training	Restriction	Environmental restructuring	Modelling	Enablement
Communication/ marketing	✓	✓	✓	✓				✓	
Guidelines	✓	✓	✓	✓	✓	✓	✓		✓
Fiscal			✓	✓	✓		✓		✓
Regulation	✓	✓	✓	✓	✓	✓	✓		✓
Legislation	✓	✓	✓	✓	✓	✓	✓		✓
Environmental/ social planning							✓		✓
Service provision	✓	✓	✓	✓	✓			✓	✓

9.4 Illustrations of Interventions using another Impact Study - ISSUES[18]

ISSUES (Implementation Strategies for Sustainable Urban Environment Systems) was part of the EPSRC's Sustainable Urban Environment Programme[19] its aim was:

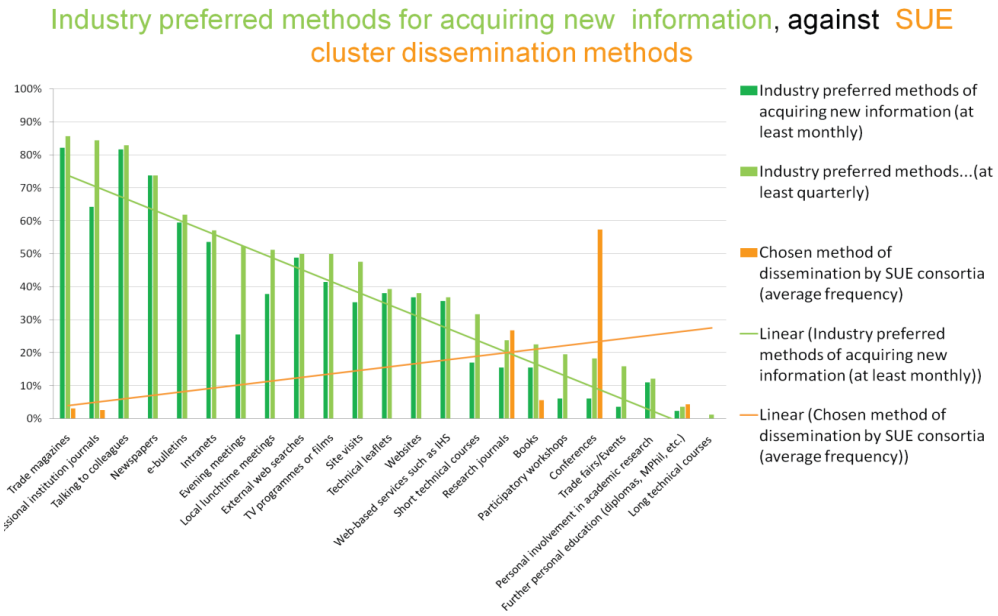
> *'to ensure that the findings from research work carried out by the SUE consortia are understood and **used by policy makers, practitioners and other end-users**'*

The following are list of what they identified as significant *helps* or *hindrances* to the use of research findings by potential end users, and suggestions of potential interventions:

Helps

- **Early engagement** with end users in the research cycle
- **'Co-creation'** of knowledge between academia & industry
- **People exchange**
- **Dissemination beyond traditional academic routes** (see figure)

Figure 27 Industry methods for acquiring information Vs traditional academic routes



Potential Interventions:

Capability

- **Education, Training** in how to make research relevant and accessible to end users. *E.g. course modules.*

Motivation:

- **Education, Persuasion, Modelling** communicating the value of adopting these approaches *E.g. Course modules, internal comms, articles*
- **Incentivise** these approaches *E.g. Funding requirements and competitions*
- **Coercion** requiring their use *E.g. Funding requirements*

Opportunity

- **Environmental Restructuring** increasing the mixing of research and end users *E.g. facilitating co-location (Knowledge Transfer Partnerships), events.*

[18] www.urbansustainabilityexchange.org.uk/

[19] www.epsrc.ac.uk/research/ourportfolio/themes/engineering/activities/sue/Pages/environment.aspx

Hinders

- **Some funding and evaluation structures** E.g. the Research Excellence Framework 2014[20] used to assess UK University research only allocates 20% of the assessment to 'Impact' and some of the evidence allowed does not confirm significant impact; citation, debate, attendance at events.

Potential Interventions:

Motivation:

- **Incentivise, Coercion** changing funding & evaluation from a hindrance to rewarding or requiring affective approaches

9.5 Illustration of Interventions using Identified Data Sets

This section illustrates the behaviour change intervention framework by applying it to data sets identified by the surveys.

The illustrations consider both:

- **Interventions resulting a complete journey from data set to behaviour change** Stamford Brook (Party Walls) and EVALOC
- **Interventions to address identified barriers and potentially increase the impact of a data set** National Energy Efficiency Database and EVALOC

9.5.1 Interventions that did result in behaviour change: Stamford Brook (Party Wall Insulation)

In the specific case of the identification of the party wall bypass and solutions the Stamford Brook data sets illustrate the complete journey from data collection to behaviour change, as summarised in the following table.

The table is a simplification of the work undertaken; there were 3 iterations of collection, processing and cognition to acquire more accurate & detailed knowledge of issue.

It is also worth noting that there was a degree of serendipity; after the initial test results identified the possibility of a thermal bypass the research project was reoriented to allow the issue to be investigated in more detail with a further round of testing. The additional work was essential in confirming that the effect was a real one and establishing its likely scale. It was this phase that provided sufficient confidence in the physics to be able to instigate regulation and behaviour change. However, this second phase was only possible because of short falls elsewhere in the research programme. This released the necessary resources and took the research team on a different journey, a journey that ended up exerting considerable influence on the detailed design of new dwellings and increased the industry's understanding of thermal bypassing in general.

[20] www.ref.ac.uk

Why the data set completed the journey to behaviour change

Many other data sets result in knowledge that highlights the potential benefit of a behaviour change but do not result in that change. The main reasons that the Stamford Brook work, in relation to the party wall bypass, managed to complete the journey to behaviour change, compared to other datasets, appear to be:

- **Enablement:** The established reputation of the researchers
- **Environmental Restructuring:** The opportunity to have an audience with Government & the Building Regulations Team
- **Education:** Of Government & the building regulations team sufficient to understand the findings
- **Persuasion:** Of Government that the problem should be addressed and the development of workable solution concept
- **Coercion:** The ability of Government to address the findings through building regulations, penalising those who do not insulate the party wall

The Nature of the Behaviour Change

It may be worth noting that creating behaviour change through coercion, albeit with tacit acceptance of industry, meant that change occurred but not through a fundamental change in attitude of industry.

Stamford Brook and the broader performance gap issue

In addition to identifying the party wall issue the Stamford Brook research also identified a range of other issues that contributed to a **'performance gap'** between design intent and operational reality.

The work has been influential in highlighting the broader performance gap issue and creating change in the form other research projects (such as the Technology Strategy Board Building Performance Evaluation studies and research undertaken by the Zero Carbon Hub).

Table 15 Stamford Brook Behaviour Change Interventions

Model of Behaviour Source:	Collection	Data Raw inputs and individual signals	Processing	Information Formatted, plotted translated, correlated	Cognition	Knowledge Evaluated, integrated, analysed	Application	Behaviour change Implemented, monitored, validated
Capability	**Education:** Building Physics **Training:** Building monitoring		**Education:** Building Physics Statistical analysis		**Education:** Building Physics Statistical analysis		**Training:** In the insulation of party walls	
Motivation	**Education**: Knowledge of issues **Incentivisation:** Potential benefit to reputation & future research	**Education:** Comprehension of results		**Education:** Comprehension of results		**Education:** Understanding by industry and Gov of results **Persuasion:** Of funders of need for additional data collection **Persuasion:** Of Gov to address the problem	**Coercion:** B Regs results penalised if party wall not insulated	**Party wall bypass solutions applied** **(Better thermal performance)**
Opportunity	**Enablement:** Established Reputation Funding				**Modelling:** Insulated party walls at Stamford Brook & Bradford demonstrate what can be achieved.	**Enablement:** Established reputation **Environmental Restructuring:** Ability to have an audience with Gov and B. Regs team		

9.5.2 Interventions that have and could result in increased knowledge and behaviour change: EVALOC

Data sets derived from the EVALOC project were identified as **under-exploited** data sets. In assessing DECC's Low Carbon Community Challenge (LCCC) projects, data collected includes

- **Energy performance data** of 50 owner occupied houses
- **Quantitative and qualitative social surveys** of 90 households across the six case study communities
- **Heating controls questionnaire**.

Analysis of this data has allowed for the creation of various academic research publications and conference presentations. As researchers worked with the communities involved in the studied neighbourhood projects, these community groups have been enabled to recruit more participants and encourage greater involvement with the Low Carbon Community group.

Successes

EVALOC was designed as an action research project; researchers interacted with the LCCC community groups, as well as studying them. Events were organised to engage residents in the LCCC projects, carry out surveys (to collect data) and share findings. Community members were able to use information generated by the researchers to increase participation in their activities. For example, maps from DECoRuM were used to find out which local homes were most likely to be suitable and in need of energy efficiency improvements. Groups' members then targeted these dwellings to persuade residents to participate in the LCCC. Uptake of retrofit measures proved to be significantly higher through this method than via local authority schemes. The interventions involved in this successful behaviour change are mapped below **Table 16.**

Barriers to Further Success

The principal researcher on the academic team identified a number of barriers to greater use of the extensive data sets that he and his colleagues have created. The data sets would need to be prepared for re-use:

- **Raw data needs to be cleaned**, that means identifying anomalies and known inconsistencies (such as when equipment has been accidentally been switched off)
- **Contextual information needs to be recorded** by those working closely at data collection stage.

Our contributor, Prof Rajat Gupta, referred to these activities collectively as *data curation*.
In order for this to happen, the following barriers need to be addressed:

1. **Funding** is needed. Prof Gupta estimates that on this 3-year project, an extra 3 months would be needed to carry out the data cleaning and recording of contextual information. The benefits of data curation in order to allow re-use would need to be promoted to funders and the wider academic community in order to gain good take-up of the process.
2. **Ground rules** would need to be established in order to make sure data curation was consistent and data re-use widely applicable. The establishing of ground rules would need to be funded and the ground rules themselves need publicising and communicating.

Table 17 suggests potential interventions to progress the EVALOC data sets along the knowledge ladder. The barriers identified are those that restrict the re-use of **data**, therefore interventions are only applicable in the lower steps of the ladder.

Our contributor suggested reprocessing and new interpretation of the data sets could lead to new findings (information and then knowledge). We are not speculating here on either the nature of further findings or their use to promote behaviour change. We do not, therefore, suggest any interventions for further up the knowledge ladder.

Table 16 EVALOC Behaviour Change Interventions Undertaken

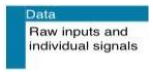
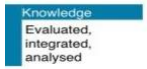
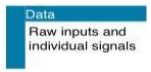
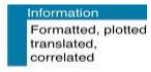
Model of Behaviour Source:								
Capability			**Education**: researchers skilled at processing and interpreting qualitative and quantitative data		**Education**: increasing knowledge of LCCC participants of climate change, energy efficiency measures and renewable energy		**Enablement**: community members able to locate potential energy efficiency opportunities within neighbourhood and target homeowners.	
Motivation					**Enablement**: data provided to allow individual residents and community groups to benchmark their energy efficiency **Education**: increasing knowledge of LCCC participants of climate change, energy efficiency measures and renewable energy		**Modelling**: success stories, energy savings shared **Persuasion**: sharing success stories	
Opportunity	**Enablement**: DECC LCCC funding to community groups **Enablement**: research council funding to study 6 LCCC groups (EVALOC) **Enablement**: researchers skilled at collecting data and feedback from LCCC participants	.						

Table 17 Potential Interventions to address barriers to the further impact of EVALOC

Model of Behaviour Source:								
Capability	**Education:** ensure researchers understand the ground rules of data curation (cleaning raw data).		**Education:** ensure researchers understand the ground rules of data curation (recording contextual data and making it available to others alongside raw data)					
Motivation	**Persuasion:** promote the opportunities of making data re-usable **Persuasion**: funders need to be lobbied to fund data cleaning. **Persuasion**: promote the benefits of data cleaning to research community. **Persuasion**: lobby for the creation and use of ground rules for data curation (cleaning raw data) to ensure consistency.		**Persuasion:** promote the opportunities of making data available to others for re-use. **Persuasion**: funders need to be lobbied to fund recording of contextual information (linking data sets appropriately. **Persuasion**: promote the benefits of recording contextual information and making it available to others to research community. **Persuasion**: lobby for the creation and use of ground rules for data curation (recording contextual information).					
Opportunity	**Enablement**: fund time to clean raw data. **Environmental restructuring**: make data cleaning a standard part of the academic research process.	.	**Enablement**: fund time to record contextual data. **Environmental restructuring**: recording of contextual information and making that available with raw data to become a standard part of the academic research process.					

9.5.3 Interventions that could result in increased knowledge and behaviour change: NEED

NEED (National Energy Efficiency Database) was identified as an under-exploited data set. It brings together data from:

- **Electricity and gas consumption** from energy suppliers
- **Property attributes** (type, age, size) from the Valuation Office Agency
- **Household characteristics** from Experian
- **Energy efficiency measures** from the Homes Energy Efficiency Database (HEED) from the Energy Saving Trust

Government uses it extensively but the following barriers to its greater impact have been suggested. If addressed, NEED data could generate extensive additional knowledge and affect more extensive behaviour change (detail is provided in Appendix H):

1. **Inconsistent collection methodologies** e.g. by energy companies for EEC, CERT & CESP energy efficiency schemes. (Nb. the work and disruption required means recollection is not considered feasible)
2. **Potential systemic skews**: Developed using data collected for other purposes
3. **Mismatches** between time data was collected
4. **Opaque weather adjustments** applied by Xoserve
5. **Opaque data set linking** processes applied by DECC
6. **Lack of access to disaggregate data** (this is due to be partially addressed in summer 2014 with general access to a smaller anonymised data set and licenced access to a larger data set

The following table suggests potential interventions to progress the NEED data set along the knowledge ladder.

It illustrates that identified barriers are all related to access and processing of raw data.

It implies that it is unknown whether, if these barriers were addressed, new knowledge would be acquired – though professional judgement would suggest it is likely.

It also illustrates that further interventions would be required to result in behaviour change. Potential behaviour change interventions are numerous and would be dependent on the nature of the knowledge acquired. Therefore they have not been suggested here (generic examples are given in Table 12).

Table 18 Potential Interventions to address barriers to the further impact of the NEED data-set

The barriers addressed, listed 1-6 above, are in brackets

Model of Behaviour Source:	Collection	Data: Raw inputs and individual signals	Processing	Information: Formatted, plotted translated, correlated	Cognition	Knowledge: Evaluated, integrated, analysed	Application	Behaviour change: Implemented, monitored, validated
Capability			**Training (1,2,3):** Of DECC staff in new analytical methods					
Motivation			**Coercion (4):** Require Xoserve to publish weather adjustment process					
Opportunity		**Education (6):** Provide greater access to raw data.	**Education (5):** DECC publish data linking process **Enablement (1,2,3):** Fund researchers to apply additional analytical techniques					

9.6 Possible Behaviour Change Interventions for GCB to adopt or promote

9.6.1 GCB's Terms of Reference

The Green Construction Board's terms of reference allow it to undertake the following core activities. Details were provided by Gilbert Barron, GCB:

1. **Provides co-ordinated Leadership**

   *The GCB brings Government and senior figures from across the construction industry together. From the Government side, BIS, DCLG, DECC, DEFRA and IUK are all represented. It is **seen as the key board overseeing green construction**, and as such, is responsible for driving this element in the industrial strategy for construction, Construction 2025.*

   ***The Infrastructure Carbon Review** is a good example of leadership. It was jointly developed with IUK – but initiated by the GCB – and sets out a series of actions for government, clients and suppliers, and has the potential to reduce up to 24 million tonnes of carbon from the construction and operation of the UK's infrastructure assets by 2050. **The Routemap for the Built Environment** is another example – to show what is required from the construction industry to achieve 80 per cent reduction in greenhouse gas emissions from 1990 levels by 2050.*

2. **Highlights required new knowledge or research**

   ***Greening the Industry** has five subgroups (Water; Waste; Carbon; Materials and Biodiversity) and has **prepared data, guidance, and best practice** in each of these areas. This one may also be relates to (6) below where working groups have themselves commissioned work (e.g. the Buildings WG material to highlight the performance gap).*

3. **Promote achievements**

   *The Promotions working group has taken this work forward by **developing case studies** and other **promotional material**, and **GCB representation** at a number of events – including a GCB stand at last year's Ecobuild and work with UKTI abroad. The Buildings and Infrastructure working groups have also identified exemplar or "**Focus" projects** which are near to market, or already live, and which are believed to demonstrate the business case for green construction.*

4. **Works on Government's role as a construction client**

   *The GCB influences central Government's role as a client through **The Chief Construction Advisor** – Peter Hansford. He sits on the GCB but is also Chair of the Government Construction Board, which is made up of the 7 main construction spending Departments, and is there to oversee their role as customer.*

5. **Address any matters agreed by the Co-Chairs**

   The agenda is set by the co-chairs (through officials in the case of the Minister), so yes.

6. **Commission research / reports**

   The main example would probably be the development and launch of the low carbon Ro utemap for the Built Environment carried out by Arup. The GCB's working groups have also each commissioned a number of research projects (3 by the Knowledge and Skills working group).

7. **Provides Policy advice**

   As per the examples given above.

9.6.2 Interventions that could be promoted by the GCB

- **All Intervention types can be advocated via policy** The GCB's role as policy advisors means they are able to advocate all policy categories. This intern means that the GCB may advocate all the intervention types indirectly via Policy.
- **Some intervention types can be implemented by the GCB directly** In the judgement of the authors the GCB's scope means it is able to implement the following interventions directly:
    - Education
    - Persuasion
    - Modelling

10 Recommendations

10.1 Improving the Data Set Landscape:

10.1.1 Creating New Data Sets Where None Currently Exist

- **All six key sustainability indicators** are represented through data sets
- **All six principal life cycle stages** are also represented, even though sometimes the number of data sets identified was small.

Opportunity: This means that there is at least some raw material to work with, but that the indicators or life cycle stages least well represented **could be strengthened through new data sets being generated**, for example operational CO_2 emissions at design stage are underrepresented. See Figure ES3, above.

Note: At present all indicators and life cycle stages have been treated as equally important, but given the findings of this study **GCB could consciously decide where their priorities lay**. If these priority areas coincide with small numbers of data sets, then work to promote, encourage or require new data collection could be started.

10.1.2 Improving Access to Existing Data Sets

- **Access is clearly** an issue with the data landscape with many private or restricted-access data sets.
- **It is likely that there are many more private data sets** than this study was able to identify.

Opportunity: There could therefore be great potential in **arranging greater access to data**, so that it could be more widely used, so that it could be combined with other data sets and so that more helpful information, guidance and knowledge could be obtained from it. In particular, **access to micro-data concerning electricity and gas meter readings** was cited as a significant barrier. There are likely to be substantial **data protection, privacy & commercial sensitivity issues** to overcome before such a wide-ranging data set could be made generally available and the GCB could assist in addressing these. Other relevant examples are Display Energy Certificates (DECs) and airtightness test results.

Opportunity: A subset of improving access to data sets is the issue of **collating existing disparate data sets**. For example, this is the situation believed to exist regarding airtightness measurement data. At present there is no requirement for organisations measuring airtightness to lodge their results on a central register, and imposing such a requirement would add costs to the surveys.

10.1.3 Improving the Quality of Data Sets

- **Gaps exist in the map of high-quality data sets**, as illustrated in Figure ES3 above.

Opportunity: Improvements can be made to the quality of data by ensuring peer-reviewed, published **standards are agreed upon and used consistently** for data collection and processing. This would include the **recording of contextual information**. This could broaden the use of the data beyond the original impetus for its collection. For example, both the EVALOC and NEED databases have potential to answer new questions.

10.1.4 Aggregation and Cross-Referencing Of Data Sets

Opportunity: Given the number of data sets that this study has identified, then one potential step towards finding out which publicly accessible data sets could be **combined into more useful composite data sets** would be to **investigate their data structures in detail**. See NEED.

Such an exercise should be able to determine the ease or difficulty of combining different data sets and in particular to explore the possibility of matching particular deep data sets with particular

broad data sets to obtain a richer picture than either can give on their own. See the Stamford Brook studies.

10.1.5 Improving Usability of Data Sets

Opportunity: Various interventions may be needed to make data sets available in a usable state or reliable/representative of reality to address aggregation, completeness and contextualisation.

10.1.6 Improving Reusability of Data Sets

Opportunity: This could be enabled by recording contextual information and cleaning the data. This could broaden the use of the data beyond the original purpose for its collection.

10.2 Further Investigation of Identified Data Sets

Of the data sets identified, the following are ones that analysis suggested could score highly given appropriate attention and intervention.

- **NEED** this data set is vast and has been identified as being potentially very useful; also it encompasses both the gas and electricity supplier meter databases
- **Airtightness Test Results** identified as potentially useful and known to be large, issues with and benefits of its further use have been discussed with both our Peer Review Group and Steering Group.
- **Display Energy Certificates (DECs)** this data set has been discussed by our Peer Review Group and identified for further comment. It represents a large number of non-domestic buildings.
- **EVALOC** two data sets resulting from this research project were identified as having good potential for greater usability and usefulness.
- **Stamford Brook** this data set is representative of all the data sets that scored very highly for all aspects of data quality and usefulness (with the exception of IEE iSERVcmb Project). It is markedly different from the data sets covered by the other case studies in that it is data from a small sample, studied in great depth.
- **Energy (gas and electricity) meter data** NB: not assessed as a case study as we have assumed much of the same information can be accessed through NEED

10.3 Promoting Collaboration

- **Lack of research collaboration** was identified as a significant barrier to research resulting in behaviour change.

Opportunity: GCB could encourage industry and academia to collaborate more often to harness opportunities of sharing of data by understanding what additional data items may be collected and used by others.

10.4 Adopt and maintain the online database

- **A database of 150 data sets has been created** (http://goo.gl/W4uGe3)

Opportunity: GCB could adopt, maintain and develop this database that has been established by the research team as part of this project. Or encourage an organisation to adopt/develop a framework or database of datasets.

10.5 Promote awareness of Data Sets

- **There is a general lack of awareness** of many useful data sets

Opportunity: Promote awareness within the built environment community of all the data sets so far identified, and encourage the community to add new data sets to the database. Or encourage an organisation to carry out this promotion and awareness-raising.

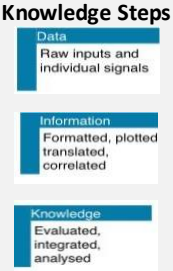
10.6 Apply the behaviour change framework to identify appropriate interventions

The behaviour change framework can be used to consider specific interventions that could enable data sets to contribute towards behaviour change.

Intervention Types and Examples

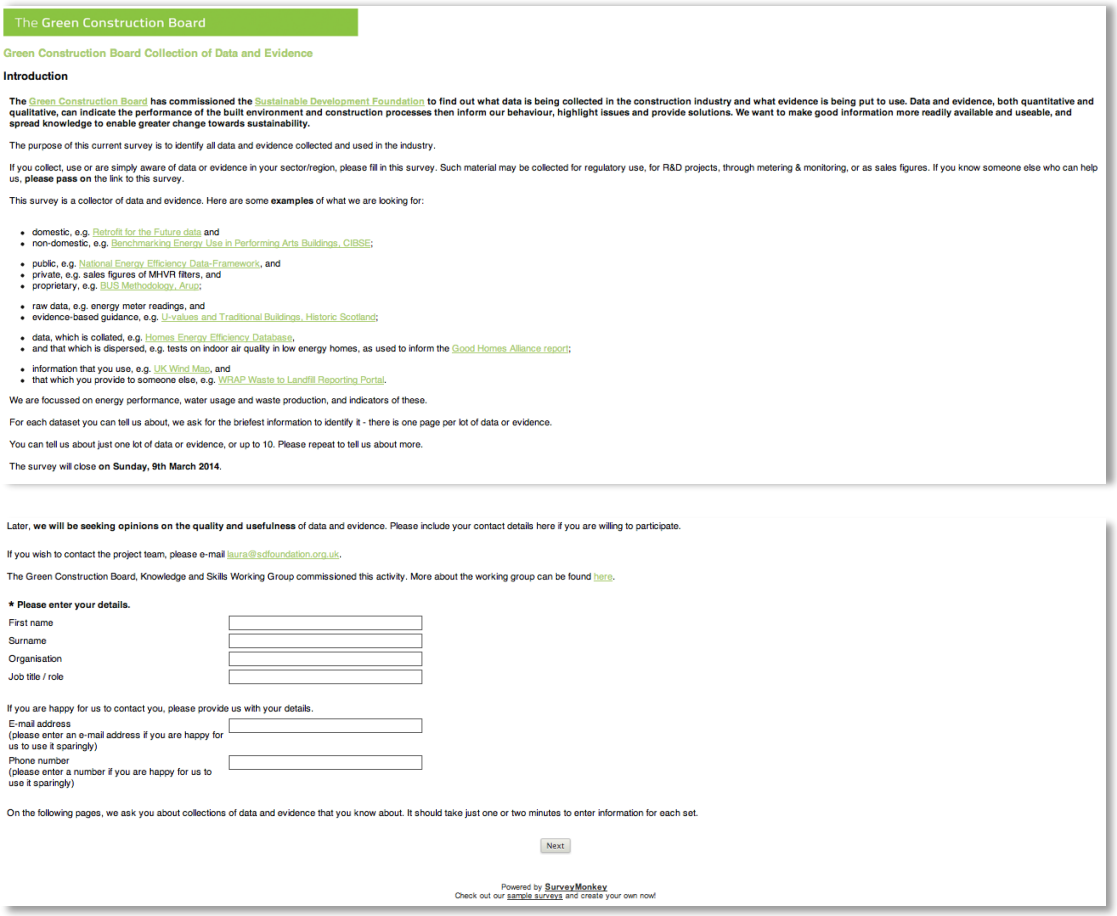
Intervention	Definition	Example
Education	Increasing knowledge or understanding	• CPD • Qualifications • Media communications
Persuasion	Using communication to induce positive or negative feelings or stimulate action	• Lobbying • Media communications • Case studies
Incentivisation	Creating expectation of reward	• Awards
Coercion	Creating expectation of punishment or cost	• Regulation • Standards • Contract requirements
Training	Imparting skills	• CPD • Qualifications
Restriction	Using rules to reduce the opportunity to engage in the target behaviour (or to increase the target behaviour by reducing the opportunity to engage in competing behaviours)	• Regulation • Standards • Contract requirements
Environmental restructuring	Changing the physical or social context	• Processes
Modelling	Providing an example for people to aspire to or imitate	• Case studies
Enablement	Increasing means/reducing barriers to increase capability or opportunity[1]	• Funding

Application of Interventions

Influence on Behaviour:	Intervention Type:	Knowledge Ladder point of Application: Knowledge Transitions	Knowledge Steps
			
CAPABILITY			
Physical & Psychological	Training Enablement	✓	
Psychological	Education	✓	✓
OPPORTUNITY			
Social & Physical	Restriction Environmental restructuring Enablement	✓	✓
MOTIVATION			
Reflective & Automatic	Persuasion Incentivisation Coercion	✓	
Reflective	Education	✓	✓
Automatic	Environmental restructuring Modelling Enablement	✓	

11 Appendix A – Survey 1

Figure 28 First page of online survey, hosted by Survey Monkey



The Green Construction Board

Green Construction Board Collection of Data and Evidence

Introduction

The Green Construction Board has commissioned the Sustainable Development Foundation to find out what data is being collected in the construction industry and what evidence is being put to use. Data and evidence, both quantitative and qualitative, can indicate the performance of the built environment and construction processes then inform our behaviour, highlight issues and provide solutions. We want to make good information more readily available and useable, and spread knowledge to enable greater change towards sustainability.

The purpose of this current survey is to identify all data and evidence collected and used in the industry.

If you collect, use or are simply aware of data or evidence in your sector/region, please fill in this survey. Such material may be collected for regulatory use, for R&D projects, through metering & monitoring, or as sales figures. If you know someone else who can help us, **please pass on** the link to this survey.

This survey is a collector of data and evidence. Here are some **examples** of what we are looking for:

- domestic, e.g. Retrofit for the Future data and
- non-domestic, e.g. Benchmarking Energy Use in Performing Arts Buildings, CIBSE;
- public, e.g. National Energy Efficiency Data-Framework, and
- private, e.g. **sales figures of MHVR filters**, and
- proprietary, e.g. BUS Methodology, Arup;
- raw data, e.g. energy meter readings, and
- evidence-based guidance, e.g. U-values and Traditional Buildings, Historic Scotland;
- data, which is collated, e.g. Homes Energy Efficiency Database,
- and that which is dispersed, e.g. tests on indoor air quality in low energy homes, as used to inform the Good Homes Alliance report;
- information that you use, e.g. UK Wind Map, and
- that which you provide to someone else, e.g. WRAP Waste to Landfill Reporting Portal.

We are focussed on energy performance, water usage and waste production, and indicators of these.

For each dataset you can tell us about, we ask for the briefest information to identify it - there is one page per lot of data or evidence.

You can tell us about just one lot of data or evidence, or up to 10. Please repeat to tell us about more.

The survey will close **on Sunday, 9th March 2014**.

Later, **we will be seeking opinions on the quality and usefulness** of data and evidence. Please include your contact details here if you are willing to participate.

If you wish to contact the project team, please e-mail laura@sdfoundation.org.uk.

The Green Construction Board, Knowledge and Skills Working Group commissioned this activity. More about the working group can be found here.

*** Please enter your details.**

First name

Surname

Organisation

Job title / role

If you are happy for us to contact you, please provide us with your details.

E-mail address
(please enter an e-mail address if you are happy for us to use it sparingly)

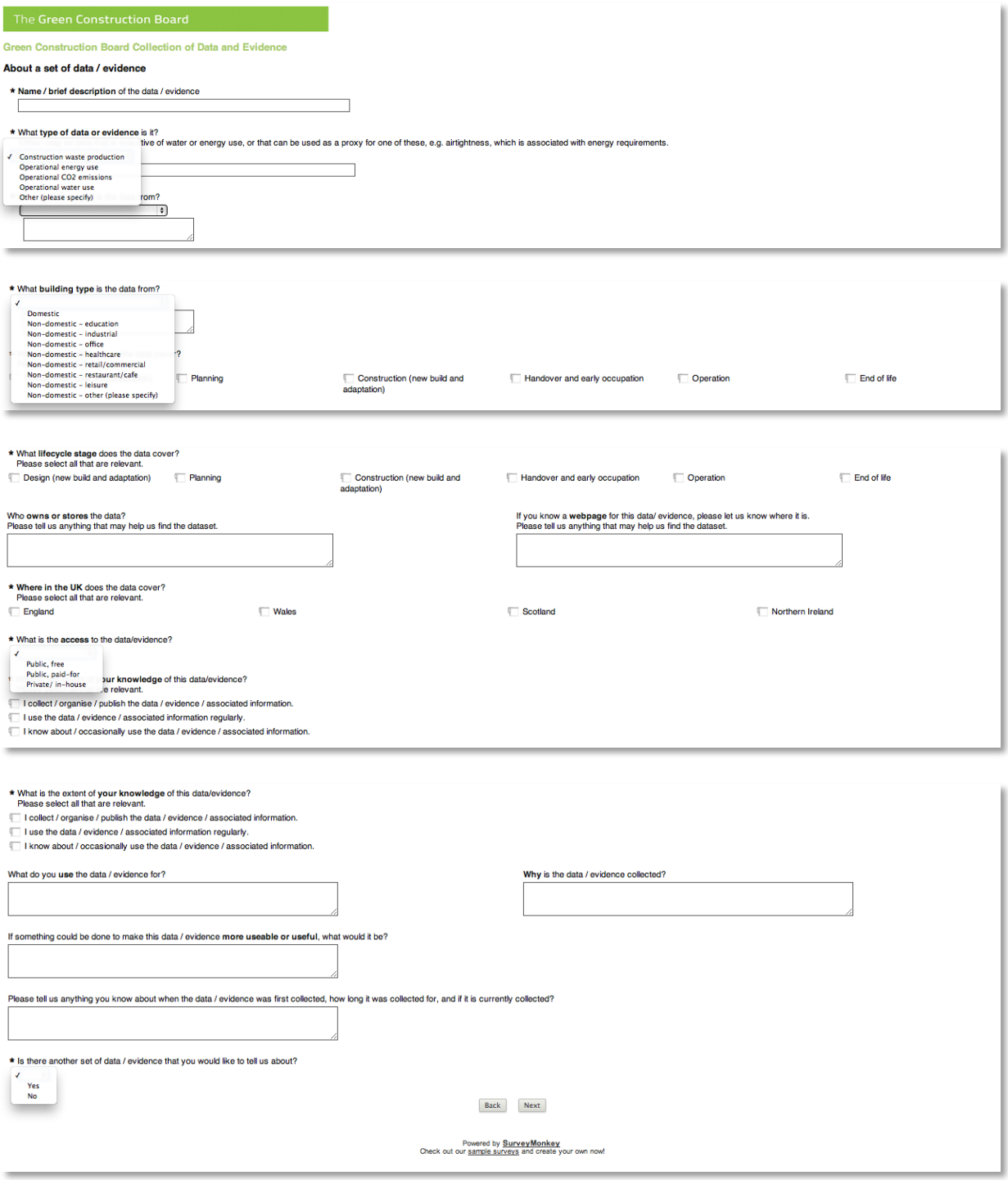
Phone number
(please enter a number if you are happy for us to use it sparingly)

On the following pages, we ask you about collections of data and evidence that you know about. It should take just one or two minutes to enter information for each set.

Next

Powered by SurveyMonkey
Check out our sample surveys and create your own now!

Figure 29 Images of second page of online survey.
In order to show drop-down lists, sections of page are duplicated.



This second page was repeated if the respondent wanted to impart information on additional set of data or evidence. We allowed them to tell us about a maximum of 10 sets of data/evidence.

12 Appendix B – Survey 2

Figure 30 Opening page for second survey

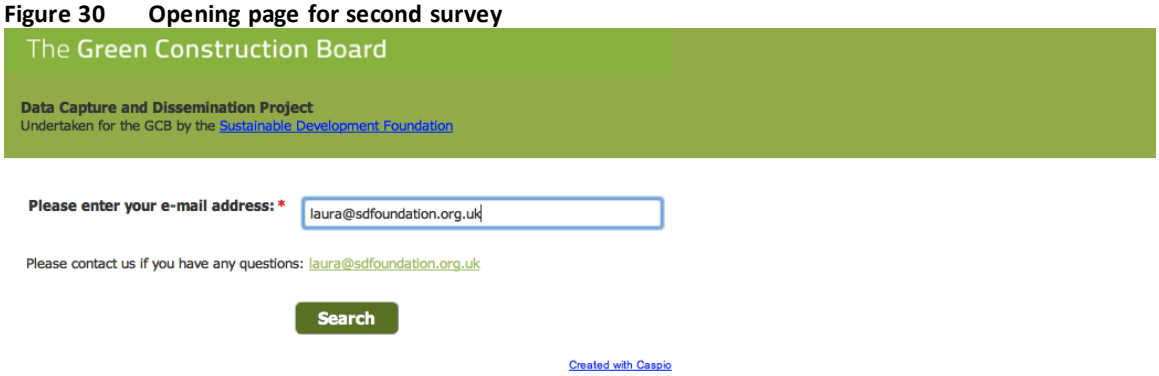


Figure 31 Second page, having entered e-mail address

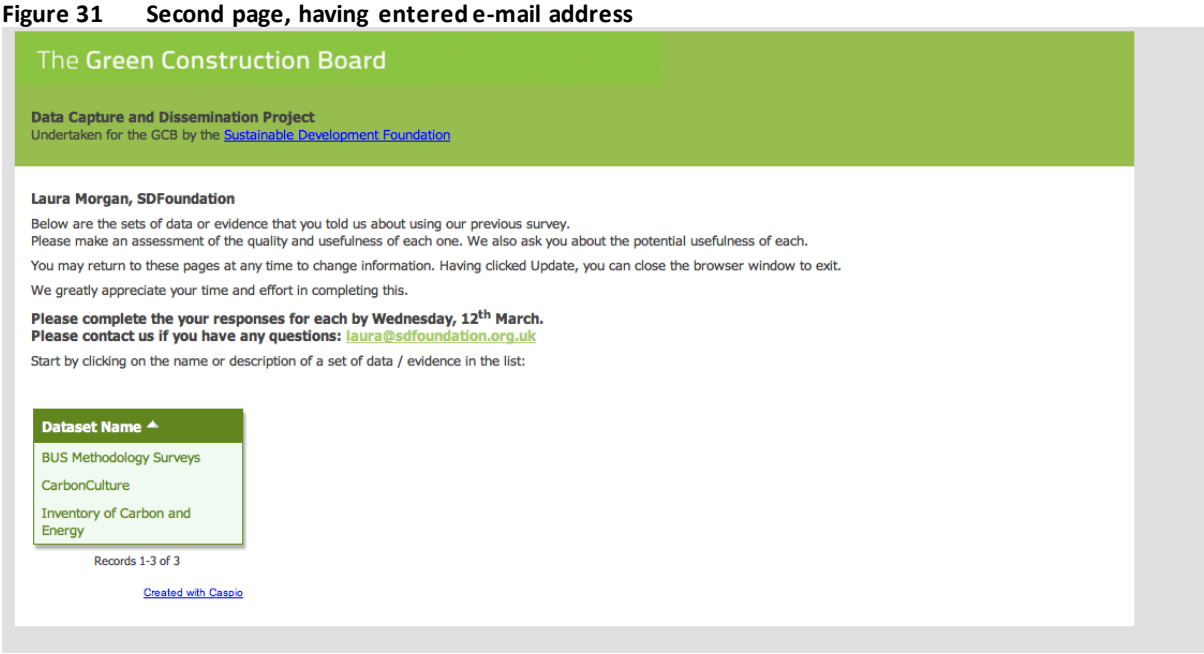


Figure 32 Data entry page for second survey

The Green Construction Board

Data Capture and Dissemination Project
Undertaken for the GCB by the Sustainable Development Foundation

CarbonCulture

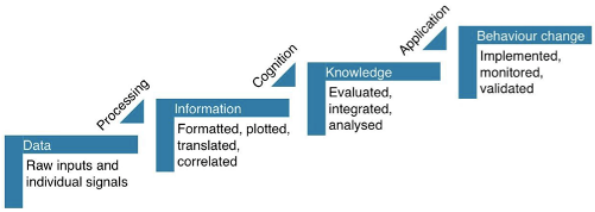
Please use this form to tell us about the quality of this set of data / evidence and its usefulness.

Please complete by Friday, 21st February.

Q1 Accessibility/Availability
We asked about availability of the data/evidence previously. We have since added a further category. Please update this information if you would like to change your response.

Data Availability Public, Free

You may wish to refer to the diagram to the right. This represents the relationship between data/evidence, information, knowledge and behaviour change. Q2 relates to the data and information end of the diagram, while Q3, Q4 and Q5 relate to the knowledge and behaviour change steps.



Q2 Please make an assessment of the QUALITY of the data/evidence here.
Please consider the quality of the data capture, the way in which it is/was processed into information, and the manner in which the information has been published.
Mark each of the 5 aspects of quality as reaching Level 1-3, or Unknown.
Use the descriptions here under each aspect as a guideline to help you make your assessment.

	Aspects of Quality				
Your Assessment	**Data capture standard**	**Data capture consistency**	**Transparency of processing**	**Type of publication ***	**Type of peer review ***
Level 3	Following an independent standard (e.g. ISO, BS, chartered institute)	Method defined pre-study and applied consistently throughout	Method of analysis or processing was published and peer reviewed	Results published and available (e.g. report, papers)	Suitable for an academic journal
Level 2	In-house standard but defined and published or available	Method evolved during early part of study, or standard not applied consistently	Method of analysis or processing was published but not peer reviewed, or only parts were published/ reviewed	Results were written up but retained in-house or restricted circulation, or only partially written up	Informal peer review (e.g. web forum)
Level 1	No known standard for capture, or standard not published	Method evolved during early & late stages of study	Method of analysis or processing has not been published	Results were not written up	No peer review
Unknown	It is not known whether a standard was applied or not	It is not known whether data was captured consistently or not	It is not known whether the processing or analysis method has been published or not	It is not known whether the results were written up or not	It is not known whether the results were peer reviewed or not

* Where a data source has given rise to a variety of reports, papers etc with different types of publication and/or peer review, please assess as Level 2 for these aspects.

2a Data capture standard Unknown **2b Data capture consistency** Unknown

2c Transparency of processing Level 1 **2d Type of publication** Level 2 **2e Type of peer review** Level 2

2f Additional Comments on Quality

Q3 Usability
Please select from the following for the **current** and **potential** usability of the data / evidence

Level 4	Data/evidence is complete, and clear conclusions or actions have been derived from it
Level 3	Data/evidence is complete, and some conclusions or actions can be derived from it.
Level 2	Data/evidence is complete, and when further analysed or combined with other data or information, conclusions and actions can be drawn.
Level 1	Data/evidence is complete but no conclusions or actions can be derived from it.
Level 0	Data/evidence is not complete.

3a Current level of usability * Level 3 **3b Potential level of usability** * Level 4

3c Please tell us what the barriers are to reaching this potential level of usability

3d Any other comments on the usability of this data/evidence

Q4 Usefulness
Please select from the following for the **current** and **potential** usefulness of the data / evidence

Level 4	I always use the dataset for decision-making in relevant circumstances (need to have)
Level 3	I sometimes use the dataset for decision-making in relevant circumstances (nice to have)
Level 2	Data/evidence has raised my awareness of the issue and caused a related change in my behaviour (e.g. climate change data leads to using more energy efficient building design models)
Level 1	Data/evidence has raised my awareness of the issue but no change in behaviour (e.g. petrol consumption lower at 56mph but still drive at 70mph).
Level 0	No change to my behaviour

4a Current level of usefulness * Level 1 **4b Potential level of usefulness (please consider uses of the data other than initial intentions for it)** * Level 4

4c Please tell us what the barriers are to reaching this potential higher level of usefulness, and briefly describe any potential new uses of the data For example, where it could be combined with other data.

Availability of all data.

4d Any other comments, for example, why your organisation uses the data/evidence or what other uses there may be for the data/evidence other than initial intentions for it

Q5 Application
Please select from the following for the **current** and **potential** level of application of the data / evidence

Level 4	Application is industry-wide.
Level 3	Application is organisation-wide.
Level 2	Data/evidence is used on all projects/activities I'm involved with (or on many relevant projects within my organisation).
Level 1	Data/evidence is used on some of my own projects/activities (or on a few relevant projects within my organisation).
Level 0	Data/evidence or information derived from it is not used.

5a Current level of application * Level 0 **5b Potential level of application** * Level 4

5c Please tell us what the barriers are to reaching this potential level of application

5d Any other comments, for example, approximate number of individuals in your organisation using the data / evidence

Please click Update to submit your responses for this data / evidence.

Update Back

Record 2 of 3

Created with Caspio

13 Appendix C – List of data sets identified

The data sets identified during the project are listed in the order they were added to the database. Note that the UIDs of the data sets are not continuous. This list omits the data sets identified by CEE but which were judged not directly relevant to sustainability in construction/buildings.

Table 19 List of relevant data sets identified during the project

Dataset_Name	Dataset_Abbr	UID	Survey_Monkey_UID	UCL_UID
UK's Housing Energy Fact File		1	MS1	
Household Electricity Survey		2	MS2	
Airtightness data from building regulations tests on new dwellings		3	MS3	
Leeds Met database of Co-heating		4	MS4	
Resource Efficiency Benchmarks		5	MS5	
Business Case for Resource Efficiency		6	MS6	
WRAP Reporting Portal		7	MS7	
Net Waste Tool		8	MS8	
Recycled Content Database		9	MS9	
BRE SMARTWaste Summary Data		10	MS10	
EVALOC heating controls questionnaire		11	MS11	
DECoRuM carbon mapping data of houses		12	MS12	
EVALOC - Monitoring data of energy performance of 50 owner occupier houses		13	MS13	
EVALOC - Quantitative and Qualitative Social surveys of 90 households across six low carbon communities		14	MS14	
Construction Sector Water data (SFfC)		15	MS15	
Data about energy use in the heritage miners cottage recently acquired at Cwmdare		16	MS16	
KESS PHD research using thermography		17	MS17	
Energy consumption in new build homes		18	MS18	
thermal performance of new homes		19	MS19	
Supply chain knowledge and competence		20	MS20	
2012 Construction Sector Carbon data		21	MS21	
Willmott Dixon EDS system		22	MS22	
Skanska UK waste data		23	MS23	
BAM Construction Ltd Construction, Demolition and Excavation Waste data		24	MS24	
UKCG Construction, Demolition and Excavation Waste Data		25	MS25	
BAM Construction Energy data		26	MS26	
BAM Construction Water use		27	MS27	
EPC Certificates for BAM Construction Projects		28	MS28	
Construction waste management data		29	MS29	
Responsible sourcing of construction materials.		30	MS30	

Dataset_Name	Dataset_Abbr	UID	Survey_Monkey_UID	UCL_UID
Construction - Water use		31	MS31	
Carbon emissions (construction operations)		32	MS32	
Landlord-procured energy consumption across our managed property portfolio		33	MS33	
Landlord-procured water consumption across our managed property portfolio		34	MS34	
Landlord-managed waste/recycling across our managed property portfolio		35	MS35	
IEE HARMONAC Project. Evidence of effectiveness of EPBD AC Inspections and the potential role of HVAC energy submetering		36	MS36	
IEE iSERVcmb Project. Use of sub-metered energy consumption for HVAC components in European Non-Domestic Buildings		37	MS37	
Waste data on civil and building projects		38	MS38	
Julie's Bicycle		39	MS39	
Queenborough and Rushenden		40	MS40	
ITP Hotel benchmarks		41	MS41	
Occupier density data		42	MS42	
Homes Energy Efficiency Database	HEED	43		1
National Energy Efficiency Database	NEED	44		2
English Housing Survey	EHS	45		3
English House Conditions Survey	EHCS	46		4
Survey of English Housing	SEH	47		5
Scottish Housing Conditions Survey	SHCS	48		6
Home Energy Efficiency Scheme Wales	HEES	49		7
Fuel poverty statistics	N/A	50		8
Domestic energy use study: why do comparable households use different amounts of energy	Hi-Lo	51		9
Energy Follow Up Survey to the English Housing Survey	EFUS	52		10
Green Deal Segmentation	GD Segmentation	53		11
Homeowners willingness to take up low carbon heat	LowCarbonHeat WTP	54		12
What's in it for me	WIIFM	55		13
Customer Led Network Revolution	CLNR	56		14
GB Electricity Demand Project		57		15
Display Energy Certificates	DECs	58		16
Energy Systems Modelling Environment	ESME	59		17
British Household Panel Survey	BHPS	60		18
Carbon Reduction in Buildings (CaRB) home energy survey (HES)	CaRB	61		19
Energy Demand Research Project	EDRP	62		20

Dataset_Name	Dataset_Abbr	UID	Survey_Monkey_UID	UCL_UID
Warm Front	Warm Front	63		21
Electricity supplier meter data	Electricity meter data	64		22
Gas supplier meter data	Gas meter data	65		23
Energy Performance Certificate data	EPC	66		24
Human Factors in Domestic Gas Consumption in South-East England, 1981-1985 (Human Factors)		67		25
Hull Low Energy Housing Project: Social Survey, 1981	N/A	68		26
Monitored Domestic Energy Use Data Archive, 1973-1983	N/A	69		27
Domestic Energy Management System trial	N/A	70		28
CREST One-minute Resolution Domestic Electricity Use Data, 2008-2009	CREST Data	71		29
Family Expenditure Survey	FES	72		30
Central Heating Survey	N/A	73		31
Irish Smart Meter Trials	N/A	74		32
CSE Lower Super Output Area dataset	CSE LSOA	75		33
CSE Fuel Poverty Data sets	N/A	76		34
Energy Company Obligation elgibility data	ECO eligibility data	77		35
CSE DIMPSA dataset	DIMPSA	78		36
Residential Monitoring to Decrease Energy Use & Carbin Emissions in Europe	REMODECE	79		37
Household Electricity Use Study	HES, HEUS	80		38
Barriers to take up of district heat	DistrictHeatBarriers	81		39
Green Deal Willingness to pay	GDWTP	82		40
Green Deal consumer demand	GDDemand	83		41
Retrofit for the Future	RfF	84		42
EMBED		85		43
Carbon Trust Micro-CHP Accelerator Trial		86		44
TSB Domestic Energy Management System trial	Maximising wind energy to power homes	87		45
BREDEM validation data		88		46
Pennyland Project		89		47
Linford Field Trial		90		48
York Energy Demonstration Project	The Green-house Programme	91		49
Milton Keynes Energy Park Courtyard Houses	(MKEP Courtyard Houses)	92		50
Milton Keynes Energy Park 1990 Project	MKEP 1990	93		51

Dataset_Name	Dataset_Abbr	UID	Survey_Monkey_UID	UCL_UID
Milton Keynes Energy Park 2005	MKEP 2005	94		52
City Form		95		53
Census		96		54
General Household Survey	GHS	97		55
British Household Panel Survey	BHPS	98		56
Understanding Society	UK Household Longitudinal Study (UKHLS))	99		57
Living with Environmental Change - Envirobase	LWEC Envirobase	100		58
Cold Comfort : A National Survey of Elderly People in Cold Weather, 1991		101		59
General Lifestyle Survey	GLF	102		60
Health Survey for England		103		61
The UK 2000 Time Use Survey		104		62
The Multinational Time Use Survey	MTUS	105		63
Health and Lifestyle Survey	HALS1, HALS2	106		64
Northern Ireland Health and Social Wellbeing Survey 1997		107		65
Scottish Environmental Attitudes and Behaviours Survey 2008		108		66
Scottish Health Survey		109		67
Social Variations in Health in Early Old Age : Investigation of Precursors in a 60 Year Follow-up Study, 1997-2004		110		68
Survey of Public Attitudes and Behaviours toward the Environment		111		69
Eurobarometer		112		70
Valuation Office Agency Rating List	VOA RList	117		75
Valuation Office Agency Summary Valuations	VOA SMV	118		76
Non-domestic Building Stock model	NDBS	119		77
Sheffield Hallam University Energy Surveys	SHU	120		78
Electronic Property Information Mapping Service	ePIMS	121		79
Commercial Buildings Energy Consumption Survey	CBECS	122		80
Experian consumer information (B2C)		123		81
Experian business information (B2B)		124		82
CarbonBuzz		125		83
Leicester City Council Project		126		84
Carbon Trust Advanced Metering project		127		85
Energy performance of air-conditioning systems in offices		128		86
BRE - Energy-related Environmental Issues Programme - Avoiding or minimising the use of air-		129		87

Dataset_Name	Dataset_Abbr	UID	Survey_Monkey_UID	UCL_UID
conditioning				
Photovoltaics in Buildings	Buildability of Photovoltaic Systems, BPS	130		88
Probe		131		89
Essex County Council's Energy Database		132		90
NHS Hospitals		133		91
Energy Management Programme - Northern Ireland		134		92
BRE Environmental Assessment Model	BREEAM	143		101
Inter-Departmental Business Register	IDBR	144		102
Hospital Estates and Facilities Statistics	HEFS	145		103
Prisons energy data		146		104
Estates Management Statistics: Environmental Information	EMS	147		105
VOA Council Tax Property Attributes	VOA CTPA	148		106
Solid Wall Insulation Field Trial		149		107
Probabilistic future weather files	PROMETHEUS weather files	150		108
Digest of UK energy statistics	DUKES	151		109
Housing in Cambridge - A computerised catalogue of a sample of British house plans	Cambridge stock	152		110
BRE Smartwaste system		323		
WRAP Waste to Landfill Reporting Portal		324		
Low Energy Buildings Database		331		
Stamford Brook first study 2001 - 2008		332		
The Elm Tree Mews project		333		
Temple Avenue existing dwellings		334		
Temple Avenue new build		335		
MIMA projects		336		
CSE Display Energy Certificate Data		338		
CSE GB household emissions dataset		339		
CSE Energy Company Obligation (CSCo) data in a usable format		340		
CSE Heating and housing census data		341		
CSE GB postcodes off the mains gas grid		342		
NHF Count Us In		343		
SHIFT Sustainable Homes Index For Tomorrow		344		
Domestic In-Use Building Performance Evaluation		345		
Energy Performance Certificate outputs		346		
Inventory of Carbon and Energy	ICE	347		

Dataset_Name	Dataset_Abbr	UID	Survey_Monkey_UID	UCL_UID
CarbonCulture		348		
BUS Methodology Surveys		349		
Greenhomes research project covering evaluation of three schemes of 38 houses built 2005-2009		350		

Table 20 List of other data sets identified during the project

Dataset_Name	Dataset_Abbr	UID	Survey_Monkey_UID	UCL_UID
OS AddressBase	OSAB	113		71
OS MasterMap		114		72
UK Administrative data		115		73
Meteorological data	Weather data	116		74
Satellite AIS Data	S-AIS	135		93
NEA	NEA	136		94
Port characteristics	ports	137		95
Time charter and spot charter fixtures	sin	138		96
Technical characteristics of global fleet	wfr	139		97
Port to port distances	p2p	140		98
Eurostat Trade data	eurostat	141		99
Comtrade Trade data	comtrade	142		100
DfT Transport statistics		153		111
National Travel Survey	NTS	154		112
Road Casualties in Great Britain	RCGB	155		113
Annual Bulletin of Transport Statistics for Europe and North America		156		114
World Road Statistics		157		115

14 Appendix D Accessibility of Data sets Tabulated

Table 21 Numbers of accessible data sets by life cycle stage and sustainability indicator

	Planning			Design			Construction			Handover & early occupation			Operation			End of life		
	Dom	Non-Dom	All types	Dom	Non-Dom	All types	Dom	Non-Dom	All types	Dom	Non-Dom	All types	Dom	Non-Dom	All types	Dom	Non-Dom	All types
Construction waste production	0	0	0	4	4	4	4	5	4	0	0	0	1	1	1	0	0	0
Construction CO2 production	0	0	0	1	1	1	2	2	2	0	0	0	0	0	0	0	0	0
Construction H2O production	0	0	0	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0
Operational energy use	3	0	0	7	1	7	6	1	6	5	0	5	17	5	19	0	0	0
Operational CO2 emissions	3	0	3	5	1	5	5	1	5	4	0	4	6	2	7	0	0	0
Operational H2O use	2	0	2	4	2	4	4	2	4	2	0	2	3	2	4	0	0	0

14.1.1 Access level and data quality against building type and life cycle stage

The table below breaks down the sub-groups of data sets. For those sub-groups that are not too numerous, it shows the individual Access and Data Quality scores for each data set as a pair of numbers. Higher numbers indicate more open access and higher quality. The key is as follows:

Access score (scale of 1 to 4)	Mean quality score (scale of 1 to 3)

Where either figure is blank, this indicates that no information was returned against the data set in question. For sub-groups that are too numerous to indicate each data set individually, a summary is given.

Table 22 Accessibility & quality breakdown of data sets against building type & life cycle stage

Building type Stage	Planning		Design	Construction	Handover		Operation		End of life	
Domestic	4 4 4 4 2 1	 2.8 2.8 2.8	24 data sets 15 public 1 restricted 5 private 10 with quality >= 2.4	27 13 public 3 restricted 10 private 11 with quality >= 2.4	4 4 4 4 2 1 4	2.8 2.8 2.8 2.8	81 36 public 2 public paid 16 restricted 7 private 12 with quality >= 2.4		1	
Non-domestic (not stated)	1		13 8 public 1 restricted 4 private 6 with quality >= 2.4	19 8 public 1 public paid 10 private 10 with quality >= 2.4	1 1	2 2.8	30 6 public 3 public paid 7 restricted 3 private 4 with quality >= 2.4		1	
Arts							4 1 2	2.75		
Aviation				1	2		4 2	2.75		

Building type Stage	Planning	Design	Construction	Handover	Operation	End of life
Data Centre			1 2			
Education		1	1 1 1 2.2 2	1 1 3	4 4 1 1 2 3 2.75	1
Fire Station					4 2 2.75	
Healthcare			1 1 2.2 2		4 4 2 2.75	
Hotel					4 1 2 1.33 2.75	
Industrial			1		1 1 1	
Leisure					4 2 2.75	
Library					2 2.75 4	
Office			1 1 1 2.2 2	1 3	4 1 1 1 1 2 3 2.75	
Police Station					4 2 2.75	
Prison					4 2 2.75	
Restaurant						
Retail			1 1 2		4 1 1 1 2 2.75	

Some observations on the results from this table:

- Access to a data set is more likely to have been assessed than the quality of that data set
- Many of the non-domestic data sets which are categorized by building type are private or in-house data sets, with an access score of 1
- There is no particular correlation between the accessibility of a data set and its quality. It might have been supposed that publicly available data sets would be consistently higher quality than private data sets but this is not supported by the findings
- The Office and Education sectors have data which is private but which is high quality

15 Appendix E High Quality Data Sets Frequency Tabulated

Table 23 Number of Data Sets for Combinations of Life cycle Stage & Sustainability Indicator

	Planning			Design			Construction			Handover & early occupation			Operation			End of life		
	Dom	Non-Dom	All types	Dom	Non-Dom	All types	Dom	Non-Dom	All types	Dom	Non-Dom	All types	Dom	Non-Dom	All types	Dom	Non-Dom	All types
Construction waste production	0	0	0	2	2	2	2	4	4	0	0	0	0	0	0	0	0	0
Construction CO_2 production	0	0	0	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0
Construction H_2O production	0	0	0	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0
Operational energy use	3	0	3	6	1	7	5	1	6	4	1	5	11	5	15	0	0	0
Operational CO_2 emissions	3	0	3	5	0	5	5	1	6	4	0	4	4	0	4	0	0	0
Operational H_2O use	2	0	2	3	1	3	4	2	4	2	0	2	2	0	2	0	0	0

16 Appendix F Verbatim record of responses

Suggestions for improving the use and usefulness of data sets, and of the barriers to usefulness, usability and application of data sets.

Contributor	ID	Data set	Improve Use And Usefulness	Usefulness Barriers	Usability Barriers	Application Barriers
Malcolm Bell	3	Airtightness data from building regulations tests on new dwellings	A national airtightness database should be provided form Building Control records. This should be a regulatory requirement placed on government.		The lack of a mechanism of pulling the data together into a national data set. This could be done by government as part of the building regs reporting process but no one seems very interested in doing it.	Availability and transparency
Malcolm Bell	4	Leeds Met database of Co-heating	This is largely a private data set but a number of publications contain the main test results and some publications carry head line results from some 30 or more tests.	Needs to have the whole set available for publication this needs agreement of the data owners/funders		
Gareth Brown	5	Resource Efficiency Benchmarks	Funding for regular updates.	Data of this nature cannot provide definitive outcomes and should be used as a decision support tool when exploring options.	Data of this nature cannot provide definitive outcomes and should be used as a decision support tool when exploring options.	
Gareth Brown	6	Business Case for Resource Efficiency	Funding for updates	Further examples would need to be provided to attain level 4.		
Gareth Brown	7	WRAP Reporting Portal	Funding for regular updates to functionality and to interrogate captured data	Data collection is reliant upon regular user input.	Data collection is reliant upon regular user input.	Data collection is reliant upon regular user input.

Contributor	ID	Data set	Improve Use And Usefulness	Usefulness Barriers	Usability Barriers	Application Barriers
Gareth Brown	8	Net Waste Tool	Funding for regular updates		The tool's dataset provides a contribution to outcomes that must consider a range of issues and data sources.	
Gareth Brown	9	Recycled Content Database	Funding for regular updates	Data must be read objectively and utilised as appropriate to specific circumstances therefore the extent of resulting action is variable.	The dataset relies upon organisations submitting data about products. Therefore, more data by more suppliers can always be provided.	The dataset relies upon organisations submitting data about products. Therefore, more data by more suppliers can always be provided.
Gareth Brown	10	BRE SMARTWaste Summary Data	Funding for regular updates	Data will only inform one aspect of much broader decisions.	Data must be read objectively and utilised as appropriate to specific circumstances therefore the extent of resulting action is variable.	
Rajat Gupta	12	DECoRuM carbon mapping data of houses	It could be used for targeted retrofitting activity		Useful for understanding the potential of applying of best practice energy efficiency measures and low/zero carbon systems	
Martin Ballard	15	Construction Sector Water data (SFfC)			Consistency of data capture points (meter and sub-meter referencing) to enable assessment of use by construction activity Assignment of consumption by project phase (Excavation, demolition, construction, commissioning)	
Julia Plaskett	18	Energy consumption in new build homes	publish the results anonymously and by building/ project type for others to mine.			

Contributor	ID	Data set	Improve Use And Usefulness	Usefulness Barriers	Usability Barriers	Application Barriers
Steve Cook	21	2012 Construction Sector Carbon data	Not all companies measure the same footprint boundary, or in the same way so comparisons can be difficult			
Nigel Sagar	23	Skanska UK waste data	Work is underway with the UKCG to develop robust baseline waste generation figures (t/m2 floor area) to use in the industry. Develop case studies to feed back to designers as to where waste minimisation can be achieved at the design stage.	Cost data is not complete. Working with waste contractors to see if we can "synchronise" their data with our waste database = more efficient entering of data.	Data is entered on a project by project basis but client data collection requirements do not always align with the methods we use.	Cost data is not complete. Some data is entered "bulked up" into monthly entries. We rely on projects assuring the data - a Skanska UK level of assuarnce is based on a sample audit of the data.
Charles Law	24	BAM Construction Ltd Construction, Demolition and Excavation Waste data	We would like to collect more information on the composition of the mixed C&D waste skips, but time constraints currently do not allow for this.		Data for the contents of mixed waste skips is not readily available and would be time consuming to complete	
Charles Law	25	UKCG Construction, Demolition and Excavation Waste Data		Data set is not complete, see above.	Data needs to be submitted by all member organisations to a consistent level of quality.	
Charles Law	26	BAM Construction Energy data			Although gathered at a site level, data does not include for example consumption of individual machines as this would be too time consuming and not cost effective.	

Contributor	ID	Data set	Improve Use And Usefulness	Usefulness Barriers	Usability Barriers	Application Barriers
Charles Law	27	BAM Construction Water use	Currently covers all water use from the site. We would like to develop the system to break down the use into varius areas: Temporary accommodation, site, commissioning, etc.		Site level data, but not down to indiviual items of use.	
Charles Law	28	EPC Certificates for BAM Construction Projects			EPCs do not give enough detail to make clear conclusions.	
Andrew Kinsey	29	Construction waste management data	Electronic chain of custody - automated data collection would make it easier to collect. accessible benchmark data would make it more useful	Better integration with procurement processes and informed decision making.	data is reasonably complete. We are working towards changes to practice and procedure based on the data collected. the barrier is time to review and understand the data and persuading people to change practices (eg adoption of plastic rather than timber pallets to avoid waste on site, which some do not agree with, and requires the supply chain to change their approach).	Does not include projects where Mace act as project managers (ie we are not principal contractors), also is limited to the UK only.

Contributor	ID	Data set	Improve Use And Usefulness	Usefulness Barriers	Usability Barriers	Application Barriers
Andrew Kinsey	31	Construction - Water use	comparison with reliable industry benchmarks - kpis vs spend are variable depending on nature of projects	Data needs to be complete and more granular. There are costs and practical difficulties in monitoring to make it more useful. We have observed one project with 100 times the use of others which caused an investigation and discovery of a large leak.	Monitoring data is incomplete as some sites have difficulty measuring it. Where it is not possible to measure (eg when working on one floor of a multi tenanted building) we have made estimates based on the numbers of people on site	As above, ability to monitor on all sites is a limiting factor.
Andrew Kinsey	32	carbon emissions (construction operations)	Comparison with industry benchmarks, more automated means of collection of data	We have used monitoring data and energy profiling to reduce cost of energy.	Data is not 100% complete due to difficulties in monitoring on sites where Mace do not have metered supplies (eg refurb of airports and work to fitout existing buildings).	We do not collect data on projects/programmes where Mace are project managers. The data is collated on Mace construction projects where we act as principal contractors
Tom Kordel	40	Queenborough and Rushenden		I wouldn't use the data set but some of the findings are useful and have changed the way I would carry out such a study in the future.	The quality of the refurbishment was not high and therefore results are mixed. The quality of some of the monitoring was also flawed.	
Tom Kordel	41	ITP Hotel benchmarks	Make the data set more open so that we can understand how big the data pool was and from what countries the data came from.	Knowledge of the source data and how the benchmark was derived	Contact the ITP and discuss the methodology behind and future use of the ITP hotel energy benchmark and its source	Currently only available via paid for report. Would need to be freely downloadable for higher application

Contributor	ID	Data set	Improve Use And Usefulness	Usefulness Barriers	Usability Barriers	Application Barriers
Jenny Mac Donnell	42	occupier density data		The data is useful in informing the future design of buildings. Barriers to making it immediately effective are the design features in existing buildings which cannot accommodate higher densities.		
Simon Elam	43	Homes Energy Efficiency Database			Access to address-level data is heavily restricted due to data sharing agreements and DPA. Complex analysis of HEED data is problematic due to the complexity of unpicking the issues around multiple data sources and the uncertainties that arise from this.	
Simon Elam	44	National Energy Efficiency Database		Less restricted access to NEED data would enhance the usefulness to a wider range of stakeholders.	Less restricted access to NEED data would obviously enhance the utility to a wider range of stakeholders.	Less restricted access to NEED data would enhance the potential application to a wider range of stakeholders.
Simon Elam	64	Electricity supplier meter data		Access to the micro-data and metadata/documentation is heavily restricted which significantly reduces the potential utility of the dataset. Actual (i.e. not modelled) energy consumption data is vital to almost every research project with an energy element so the potential uses	Access to the micro-data and metadata/documentation is heavily restricted which significantly reduces the potential utility of the dataset.	Access to the micro-data and metadata/documentation is heavily restricted which significantly reduces the potential applicability of the dataset. Actual (i.e. not modelled) energy consumption data is vital to almost every research project with an energy

Contributor	ID	Data set	Improve Use And Usefulness	Usefulness Barriers	Usability Barriers	Application Barriers
				are virtually boundless. Not having access to disaggregated energy consumption data has been a real problem for many research projects investigating the socio-technical drivers of energy demand.		element so the potential uses are virtually boundless. Not having access to disaggregated energy consumption data has been a real problem for many research projects investigating the socio-technical drivers of energy demand.
Simon Elam	65	Gas supplier meter data		Access to the micro-data and metadata/documentation is heavily restricted which significantly reduces the potential utility of the dataset. Actual (i.e. not modelled) energy consumption data is vital to almost every research project with an energy element so the potential uses are virtually boundless. Not having access to disaggregated energy consumption data has been a real problem for many research projects investigating the socio-technical drivers of energy demand.	Access to the micro-data and metadata/documentation is heavily restricted which significantly reduces the potential utility of the dataset.	Access to the micro-data and metadata/documentation is heavily restricted which significantly reduces the potential application of the dataset. Actual (i.e. not modelled) energy consumption data is vital to almost every research project with an energy element so the potential uses are virtually boundless. Not having access to disaggregated energy consumption data has been a real problem for many research projects investigating the socio-technical drivers of energy demand.

Contributor	ID	Data set	Improve Use And Usefulness	Usefulness Barriers	Usability Barriers	Application Barriers
Simon Elam	66	Energy Performance Certificate data		By far the biggest barrier to the usefulness of the data is the high cost of purchasing bulk EPC data (10p per record or aprox £1m for all UK records)		By far the biggest barrier to the application of the data is the high cost of purchasing bulk EPC data (10p per record or aprox £1m for all UK records)
Simon Elam	80	Household Electricity Use Study		The only barrier (if it is one) to the current usefulness is that it is a new data source and so has yet to be fully exploited and the findings fully absorbed by all relevant stakeholders.		Similar to Q4, the only barrier (if it is one) to the current and future application of this data / evidence is that it is a new data source and so has yet to be fully exploited and the findings fully absorbed by all relevant stakeholders
Megan McMicheal		City Form			The consortia responsible for this data have published a number of journal articles and a book chapter (see http://oisd.brookes.ac.uk/researchfindings/07-02.html).	The data is based on case studies of three cities, so can not represent the general population.
Liz Reason	331	AECB Low Energy Buildings Database	Make it much more widely known and elicit more data sets.	If more projects and in-use energy data were entered into this relatively simple database it would provide some real insight into what works and what doesn't on retrofit projects.	We need more projects entering data and probably a process of establishing its quality	More people should know about this database and be incentivised to add more buildings to it. Perhaps we can find a way of celebrating every time a project is entered and celebrate more when they provide the in-use data.

Contributor	ID	Data set	Improve Use And Usefulness	Usefulness Barriers	Usability Barriers	Application Barriers
Malcolm Bell	332	Stamford Brook first study 2001 - 2008	There is a vast store of archived data that is secondary to the main published material. This could be mined for further insights and understandings			
Malcolm Bell	333	The Elm Tree Mews project	Data set has unpublished archive that could be mined for further insight and understanding.			
Malcolm Bell	334	Temple Avenue existing dwellings	Data mining of archive			
Malcolm Bell	335	Temple Avenue new build	Data archive could be mined for further insight			
Malcolm Bell	336	MIMA projects	Needs to be released by MIMA			
Adrian Buffery	350	Greenhomes research project covering evaluation of three schemes of 38 houses built 2005-2009	A shared method for formatting and publishing in-use building research that would be of practical value to commissioners in improving quality.	There is a fundamental difficulty with small-scale research in the field given the high level of variability in domestic setting. Combining many data sets should allow some meta confidence!	Some data already analysed and used in service/design development, full data and analysis due later this year.	Resources to publish and promote effectively in collaboration with relevant parties.

17 Appendix G Framework for Exploring Interventions

Table 24 Framework for Interventions to improve the Quality or Usefulness of a Data Set

		Access		Quality scores (each 1-3)										Usefulness scores (each 0 – 4)					
Dataset (Project unique ID)	**Scope/ population**	**Accessibility score (1-4)**	**Intervention**	**Data capture score**	**Intervention**	**Capture consistency**	**Intervention**	**Transparency of processing**	**Intervention**	**Publication**	**Intervention**	**Peer review**	**Intervention**	**Usability score (0-4)**	**Intervention**	**Usefulness score (0-4)**	**Intervention**	**Application score (0-4)**	**Intervention**
Airtightness test results (3)	New dwellings	1 (private/ in-house)		3		3		1		Unknown		1		2 (potential 3)		2 (potential 4)		1 (potential 4)	
Display Energy Certificates (58) **(based on discussion 08/04/2014, not survey assessment)**	Public buildings with useable space greater than 500 square metres	2 (restricted)	Make bulk data free / low cost	3		2		2	Raw data is not available – make available by …?	3 for individual DECs, 0 for bulk data									
National Energy Efficiency Database (NEED)	Energy use and energy efficiency in dom and non-dom buildings in GB variety of sources	2 (restricted)		1		1		1		2		2		3 (potential 4)		2 (potential 4)		2 (potential 4)	

18 Appendix H Full Case Study: Under-exploited data set: NEED

(National Energy Efficiency Database)

Summary

NEED is the National Energy Efficiency Data-Framework, controlled by DECC. It brings together data sets of energy use, energy efficiency measures, property attributes and household characteristics in domestic and non-domestic buildings in Great Britain and contains records relating to over 27million buildings.

It contains data derived, largely, from administrative sources and it is quite widely utilised within DECC for policy development, analysis and evaluation. The stated aims of NEED are to:

- develop, monitor and evaluate key policies;
- identify energy efficiency potential which sits outside the current policy framework;
- develop a greater understanding of the drivers of energy consumption; and
- gain a deeper understanding of the impacts of energy efficiency measures

(DECC, 2013)

While limited subsets of NEED data have been made available for a few DECC-mandated research projects in the past, there is currently no mechanism for accessing NEED data outside of DECC. However DECC intend to release anonymized NEED micro-data in the summer of 2014 which will remove some, but not all, of the barriers to greater utilisation of the data.

Data/Evidence and Data Collection

Data for NEED has been collected since 2000, although the framework itself has only been in place since 2009. NEED brings together data from four primary sources (three for non-domestic) each linked to an address "spine" (based on Ordnance Survey *AddressBase*[21]):

- Electricity and gas consumption from energy suppliers
- Property attributes (type, age, size) from the Valuation Office Agency
- Household characteristics from Experian (occupying organization characteristics such as turnover, employment, SIC code, type of premises for non-domestic)
- Energy efficiency measures from the Homes Energy Efficiency Database from the Energy Saving Trust (only for domestic properties).

There is a proposal to add an additional primary dataset, namely data from the EPC Register (controlled by DCLG and operated by Landmark). This would add extra detail to HEED (Homes Energy Efficiency Database) as the EPC register contains additional building characteristic data that is not provided by the VOA dataset.

NEED does not contain information relating to:

- Very large consumers of energy, such as power stations
- Properties in Northern Ireland
- Non-metered fuels such as oil and coal
- DIY energy efficiency measures that are not recorded on HEED.

Issues & Interventions

One of the most significant issues in NEED relates to a lack of standard methodologies for collecting data across, and even within, the various data sets.

[21] AddressBase is an Ordnance Survey software product that provides spatial coordinates, Unique Property Reference Number (for cross-referencing with other databases) and high-level building function.

Unlike data generated by surveys specifically developed with clear (research) objectives (e.g. EHS (the English Homes Survey)), NEED has been developed using data sets derived from processes quite incidental to the statistical analysis for which they are currently being utilised. This can lead to systemic biases or skews, which should not occur in surveys using best practice methodologies.

For example, EEC, CERT and CESP data in HEED was generated by energy supplier obligation schemes which were often pursued with a least-cost approach to delivery and with no consistently applied standards for collecting data. This results in highly inconsistent categorizations of data even within individual source files as different installers will use different classifications for data variables e.g. different age bands were used for property age, different bandings for (pre-installation) depth of loft insulation etc. Although there was an attempt to improve the consistency of data collection with the signing of the Voluntary Data Sharing Agreement in 2012, it did not affect data collected before that time (the bulk of EEC and CERT data).

VOA data are generated primarily to determine the rateable value of buildings while Annualised Quantity (AQ gas consumption) data are used to estimate gas consumption for the upcoming gas year and to apportion gas transportation costs between shippers.

The weather adjustment process applied to AQ gas consumption records is notably opaque and thus increases the difficulty in utilising NEED data in research projects.

The potential skews in the individual data sets could be compounded when the data sets are linked in NEED e.g. due to differing match rates/quality when address-matching and issues with the validity of integrating data collected at different points in time.

The data was collected from many different sources. Data therefore may not be directly comparable unless it is harmonised prior to analysis (which is likely to involve significant effort).

With a nod to Donald Rumsfeld, the issues above are some of the "known unknowns"/uncertainties. Clearly is not feasible to ask the data providers to attempt to rectify these issues affecting millions of records collected over many years. However it may be possible to conduct a detailed analysis of all known issues and attempt to quantify the problems and potentially provide, for example, correction factors for key data variables. It should be noted, however, that such a process will involve significant effort with uncertain results (many data sources go back over a decade and were provided with little or no documentation. Therefore it may be almost impossible to shed any further light on data collection issues in many cases).

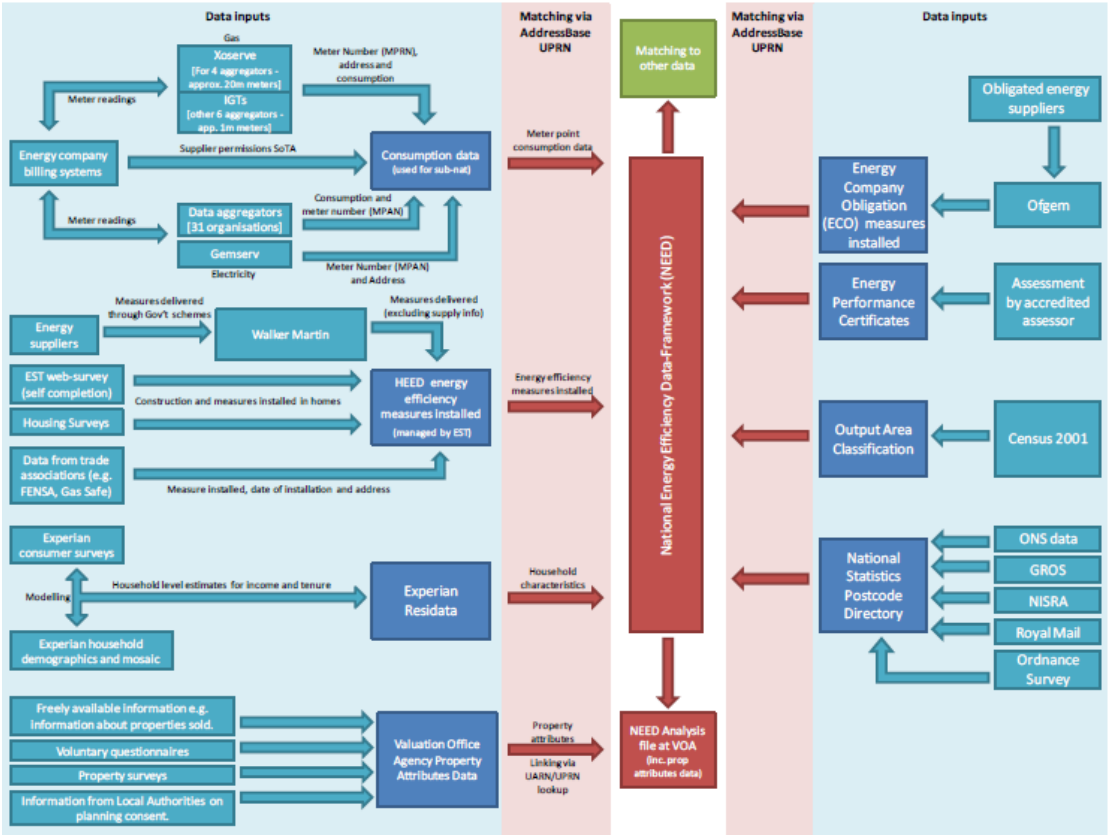
However some data collection issues could be clarified relatively easily. The issue with the weather adjustment of gas consumption is simply that the data provider (Xoserve) does not currently provide complete and transparent information regarding its methodology for this process. DECC or Ofgem could/should mandate the provision of complete and transparent documentation of this process.

If we return to Rumsfeld then the other issue is the "unknown unknowns". With the inherent complexity in integrating so many data sets, each with its own process for data collection, there are likely to be many other issues hidden in the data that could bias results or skew analysis. A thorough investigation to uncover these issues would be highly advisable.

Data Processing

As can be seen from the diagram below, the process to link and integrate the various data sets in NEED is quite complex.

Figure 33 Process map for linking data sets within NEED



(DECC, 2013)

Issues & Interventions

DECC provides useful summaries of its processing methodologies at the various stages of data processing but it often does not provide enough detail to be useful in informing detailed academic research.

The obvious example here would be the process by which the data sets are linked using address-matching processes to a data spine (Ordnance Survey AddressBase). Basic details of match rates are provided but information on false positives and false negatives are desirable to enable analysis of potential systemic bias resulting from the address matching process.

Information

Aggregated data (Excel) and reports are published on the NEED web pages at https://www.gov.uk/government/collections/national-energy-efficiency-data-need-framework

In addition to the analyses that are contained in the regular NEED reports, the web site also contains a pivot-table creator to customise data views on domestic energy consumption. This allows two main variables to be analysed together (property age, property type, number of bedrooms, floor area, tenure, income, number of adults and region). Other parameters can also be set for the type of data to display (mean figures, median figures, or number of observations), the fuel to analyse (gas or electricity) and the year to show (2005 through to 2011). All the summary data that the table creator uses can also be downloaded as a spreadsheet file in csv format.

DECC has undertaken a consultation on making some NEED micro-data available. The proposals are for a limited anonymised data set of approximately 20,000 records to be available to anyone (expected end of May 2014), and for a larger data set of approximately 4 million records to be available through an end-user licence (expected by the end of July 2014).

Issues & Interventions

The critical issue here is the lack of access to NEED micro-data which may be partially resolved with the release of anonymized NEED data sets in the summer of 2014.

"The release of anonymised data could and should prove to be an invaluable resource for academic research - including the development and refinement of housing stock and energy models, the development of new analysis techniques, and the development of holistic approaches to understanding energy use and behaviours in the UK housing stock. The availability of a dataset combining energy consumption, property characteristics, energy efficiency interventions and other data at the building level will facilitate research into the interactions of energy use drivers that would be not be viable using aggregated statistics or other data sources.

The data could be used across a wide range of current and future research projects either as a primary data source or to contextualize the analysis of smaller, more focused data sets.

Examples of research projects that are likely to benefit from the use of the anonymised NEED dataset are the development of a holistic end use energy demand data model; investigation into the prevalence of high/low domestic energy users; and examining the trends of energy demand among certain household characteristics over the period of data."

(Elam, 2014)

However it should be noted that a further intervention (the provision a non-anonymized or pseudo-anonymized dataset which would enable linking to other data sets) would significantly improve the usefulness and usability of NEED data.

"If a further dataset was made available under special/secure license terms which included high resolution geographic identifiers then the value of the NEED data would grow enormously. This dataset would be extremely beneficial for projects investigating the relationship between energy demand across buildings and transport; three dimensional modelling of building stock; the impact of building permeability on energy use; the relationship between temperature, dwelling characteristics and energy demand; using NEED and EPC database to examine the relationship between ratings and actual energy use; drawing a sample from NEED to represent solid wall properties to contextualize the Solid Wall Field trials and monitoring; examining clusters of low energy users in more detail, particularly those areas that were built to a historically high standard; and examining the effect current (past) energy use has on decision-making to undertake energy efficiency retrofits."

(Elam, 2014)

Cognition and Knowledge

NEED has been used to generate specific reports on energy consumption and the impact of energy efficiency measures. These are published on the NEED web pages at https://www.gov.uk/government/collections/national-energy-efficiency-data-need-framework. Analyses carried out in 2012 and 2013 include

- Analysis on the impact of installing energy efficiency measures on a household's gas consumption
- Analysis of domestic gas and electricity consumption by property attribute and household characteristics
- Analysis of the quality of non-domestic energy consumption data.

Statistical releases

- National Energy Efficiency Data-Framework (NEED) report: Summary of analysis 2013 Part 2
- National Energy Efficiency Data-Framework (NEED) table creator
- National Energy Efficiency Data-Framework (NEED) report: Summary of analysis 2013 Part 1
- National Energy Efficiency Data-Framework (NEED) report: Summary of 2012 analysis

Issues & Interventions

There are no significant issues with NEED publications. They are published as official statistics and so can be said to have undergone some form of peer review. Release of non-domestic NEED analysis would be welcomed.

Application and Behaviour Change

As previously noted, NEED has been well utilised within DECC for policy development, analysis and evaluation.

"The impact of NEED successfully achieving these aims leads to benefits for Government, business and households. NEED has already supported a number of DECC policies, with important consequences. For example, The Green Deal. NEED has been used to understand the reduction in consumption (and resulting reduction in energy bills) for households installing energy efficiency measures. To date NEED has looked at savings from a number of measures, including cavity wall insulation, loft insulation, installation of condensing boilers and solid wall insulation.

Evidence from NEED showed that the savings observed in real households are lower than the technical estimates previously assumed. Having this evidence meant DECC could use improved estimates of savings in the Green Deal and ECO which has given consumers a better indication of likely energy savings resulting from having measures installed.

NEED has also had a smaller, but still significant, part to play in a range of other DECC policies, for example, the Renewable Heat Incentive and Fuel Poverty. Data on consumption by property attributes, including the distribution of households consumption, has been used to help DECC understand the likely under or over payment if payments for the renewable heat incentive were to be based solely on property attributes available in NEED. It has informed Fuel Poverty analysis so there is a better understanding of actual consumption for different types of properties and households and therefore a better understanding on how policy options will impact on different households. Having this information enables DECC to provide better value for money and understand better the impacts of policy options, for both DECC and consumers.

NEED has also helped DECC understand where further research should be focused. It provides high level results which have highlighted a need for further investigation, for example understanding why households which appear the same in physical property attributes use varying amounts of energy, and understanding how the use of heating controls impacts on the amount of energy households are using.

These examples of policy support would not have been possible without the use of property level data available as a result of the NEED project. No other source currently available can provide the extent of information provided by NEED and the matching at property level enables a deeper understanding of the distributions and impacts on different types of properties.

In future it is anticipated that the role of NEED in development and evaluation of DECC policy will continue to grow. In the short term it will contribute to the evaluation of CERT and CESP – looking at the savings from measures installed under these schemes. In the medium term, it is a core component of the evaluation of the Green Deal and Smart Meters. The property level data only available through NEED is essential to these programmes to be able to isolate the impacts of specific policies. To undertake specific research with the same power would be prohibitively expensive.

NEED will also continue to support other DECC policies, such as Heat, Fuel Poverty and ECO. It will also support the development of new policies, helping to identify the potential focus of future policy, particularly in the non-domestic sector where DECC's evidence base and policy development is less advanced.

Externally, the publication of anonymised and aggregated information derived from NEED is valued by a range of stakeholders, with energy suppliers seeing it is as beneficial to have a trusted

independent source, which can help them show the benefits of getting homes insulated. Academics use the published outputs to feed into and validate related research.

Alternatives to NEED include surveys and technical monitoring trials. These are important and compliment the data in NEED, but cannot alone provide the evidence required to support DECC policy. Technical trials give detailed insight into why things are happening, but cannot lead to the population wide conclusions that can be drawn from NEED. Technical trials can also mean that those participating in the trials may change behaviours as a result of knowing they are a participant and large trials are extremely expensive.

While survey data could provide an alternative to the data used in NEED, it would be costly to collect data which allows the detailed analysis that can be done using NEED. Sample sizes would need to be very large (many times larger than the current sample for the English Housing Survey) and the burden on households and businesses would be significant. The approach of combining existing data sources is also consistent with the Government's ambitions to make more use of administrative data." (DECC, 2013)

Issues & Interventions

The knowledge and insight gained from NEED has undoubtedly been applied in relation to policy evaluation and development. The only issue is that without access to data outside of DECC, the benefit of generating knowledge and behaviour change outside of a policy context is significantly reduced.

Conclusions/Lessons

NEED is a very valuable data resource that is currently only utilised within DECC but has the potential to be of great benefit to the research community and other stakeholders in the energy sector.

The key issues that should be addressed are:

- Appropriate access to NEED micro-data - going beyond the current proposal for two anonymized data sets.
- Uncertainties and data quality issues arising from inconsistent or non-existent data collection methodologies.
- Uncertainties arising from data linking and other data processing

It should also be noted that there are barriers to the suggested interventions.

- DECC is inclined to proceed cautiously in relation to opening up access to NEED data. DECC does not own any of the data in NEED but has merely negotiated access to the various data sets. Theoretically this access could be withdrawn at any time (with the energy consumption data owned by the energy suppliers being of most concern).
- It is likely that DECC is concerned with the negative publicity that would likely be generated should any NEED data be lost or stolen.
- The data providers to NEED are disinclined to do anything above and beyond the bare minimum required by the terms of the data sharing agreements. Even the bare minimum is done with noted reluctance in many cases. This will impede most interventions to improve the utility of NEED.
- DECC, like most organizations relying on public funding, are battling with ever decreasing resources. Any intervention will require additional effort and funds and some will require significant additional resources that will be difficult to obtain in the current economic climate

References

DECC. (2013). *NEED Privacy Impact Assessment*. (URN 13D/199). Department of Energy & Climate Change.

Elam, S. (2014). Response to NEED consultation on anonymised data