eprints@whiterose.ac.uk
https://eprints.whiterose.ac.uk/

Article

An Expert View on Data and Modelling for Planning Domestic Retrofit

Marianna J. Coulentianos [1,*], Danielle Abbey [2], Christy Tsz So [1] and Wil O. C. Ward [2,3]

1   School of Design and Creative Arts, Loughborough University, Loughborough LE11 3TU, UK
2   Department of Civil and Structural Engineering, University of Sheffield, Sheffield S10 2TN, UK;
    dabbey1@sheffield.ac.uk (D.A.); wocw2@leicester.ac.uk (W.O.C.W.)
3   School of Computing and Mathematical Sciences, University of Leicester, Leicester LE1 7RH, UK
*   Correspondence: m.coulentianos@lboro.ac.uk

Abstract: The transition to Net Zero rests partly on the widespread adoption of energy-efficient retrofit measures for domestic dwellings. The scale of retrofit efforts is extensive, as up to 80% of the UK's domestic housing stock for 2050 has already been built. To address the scope, data and models will play a crucial role in informing design decisions and optimising retrofit strategies. While new methods and tools for data and modelling in retrofit continue to be developed, the perspectives of professionals using these tools on their quality remain mainly absent from discussion across academia and practice. This study investigated the experiences and perceptions of data and modelling from professionals working in the planning stages of domestic retrofit, serving as a needs-finding exercise driving retrofit planning. Through semi-structured interviews and qualitative coding, the results highlight a critical trade-off between precision, confidence, and the burden of data collection. These findings underscore the need to balance precision, ease of use, and adaptability in data and modelling retrofit tools. Issues around data availability and wider access to data and modelling results across stakeholders emerged as a missed opportunity.

Keywords: domestic retrofit; data; modelling; interviews

1. Introduction

The necessity and scale of the challenge of domestic retrofit in the UK required to meet national targets for Net Zero is well established [1,2]. Residential dwellings contribute up to 20% of UK energy-related carbon emissions [3]. These dwellings will form 80% of the total domestic stock in 2050, meaning that to reach net-zero emissions, they will need to be retrofitted alongside other measures, such as decarbonising the power grid [1]. Furthermore, the motivations behind retrofit are not solely carbon-related: energy efficient retrofit interventions can reduce fuel costs and increase health and wellbeing [4].

The optimal retrofit pathways for reaching Net Zero targets are diverse and vary for each dwelling depending on factors such as building geometry; thermal properties; systems and controls; building use; and ventilation [5]. The driving factors behind retrofit decisions can differ, including homeowner and occupant preferences; timing; financial cost; professional standards; and the availability of services and resources [6–10]. Many of these decisions are informed by large datasets which are then fed into modelling tools, designed to support retrofit decision making. For example, Patterson et al. (2023) describe the use of surveys as essential for improving energy performance in retrofit scenarios and building consistent quantitative baselines of building stock [11]. The scale of the solutions required to retrofit the housing stock on a national scale means that the tools that drive decision-making need to be fit for purpose, and the data that feed into these tools need to be reliable, accurate, and robust to ensure the efficiency of, and confidence in, affected interventions.

Amid increasing enthusiasm for leveraging data and modelling across the built environment, particularly utilising the advances in big data [12], there is evidence of overuse

and overreliance on particular sources of data in planning retrofit interventions, despite their known or suspected pathologies [13]. One example is Energy Performance Certificates (EPCs), produced using an adapted version of a simulation model called the Standard Assessment Procedure (SAP) and used to report estimated annual energy consumption, which is widely used in policy and decision-making practices, despite being considered not-fit-for-purpose by the UK Green Buildings Council [2].

While many efforts are ongoing to increase the reliability of data and modelling for retrofit [5,14,15], there are few studies investigating the data- and modelling-related experiences and needs of stakeholders working in the retrofit industry, including those involved in day-to-day decision making on retrofit projects that span multiple dwellings. Furthermore, studies have explored the wider trends in energy efficiency research, such as [16]. However, they do not focus directly on issues concerning data and modelling. The research presented in this paper aims to help bridge this gap by investigating the data and modelling needs of stakeholders working in domestic retrofit planning and early stages of implementation. Designed as a needs-finding exercise, a series of semi-structured interviews were conducted with professionals involved in different roles and stages in the domestic retrofit process to explore their experiences with data and models in the retrofit planning stages and understand their needs. Common themes across different actors are identified and opportunities for future work are highlighted to develop robust data- and modelling-related tools for retrofit.

The remainder of the paper is structured as follows: Related work is introduced in Section 2. The research methodology is presented in Section 3, with information about the participants and the data collection and analysis processes which followed a qualitative research process of thematic analysis. In Section 4, participants are further characterised, and the thematic analysis results are presented. In Section 5, a discussion of the study is presented, including the design, research findings, and limitations, followed by concluding remarks in Section 6.

2. Data and Modelling in the Retrofit Process

2.1. The Retrofit Process

A characterisation of the retrofit process, its main stages, and stakeholders is presented below to anchor the literature on data and modelling in the retrofit process.

The LETI Climate Emergency Retrofit Guide framework [1] is a holistic framework that breaks the retrofit process down into five stages, covering the whole retrofit project timeline. The LETI framework maps its substages to the RIBA Plan of Work, an industry-recognised project management tool used in UK built environment projects [17], and to different stages in the PAS2035:2019 compliance map. PAS2035 is a standards specification for the whole-building energy-efficient retrofit of domestic dwellings, compliance to which is required for publicly funded retrofit projects. The PAS2035 offers risk assessment pathways, categorised as A, B, or C, indicating projects with increasing levels of risk and the number of properties impacted [10]. The LETI framework maps to risk pathway B, which is used for retrofit programmes with multiple measures that are to be installed in multiple properties [1]. The LETI framework is therefore considered the standard retrofit process for the remainder of the paper, used as a guide to frame the existing literature on different steps of the retrofit process and to frame the research presented. Likewise, the findings here are relevant to PAS2035:2019 Pathway B compliant retrofit projects.

Various actors and stakeholders are involved at different stages throughout the retrofit process. Beyond user stakeholders, such as leaseholders, occupants, and landlords, there are many actors involved in the planning, design, and implementation of retrofit programmes, both on individual dwellings and larger housing stock. The PAS2035 specification defines specific roles which must be in place for compliance: these include a retrofit coordinator, a certified professional who oversees the planning and project management; retrofit assessors, who undertake assessments on the building; and retrofit designers, typically architects who design the retrofit interventions [10]. Additional PAS2035 roles include installers and

evaluators responsible for installation and post-installation checks, respectively. There is no requirement that different individuals hold each of these roles. In practice, an individual may hold multiple roles within the project. The LETI framework also describes different roles and stakeholders that engage with the different stages. In Figure 1, the retrofit process as laid out by LETI is shown, mapping the RIBA Plan of Work stages and the corresponding stakeholders involved throughout the process. Loose categories of stakeholders have also been provided to better signpost responsibilities: this is particularly useful when describing the different roles of stakeholders with overlapping interests, for example, the local government who can act as housing providers, funding bodies, and data and information holders. Figure 1 highlights the scope of the topics explored in this paper, focusing on the first three stages of the LETI framework.
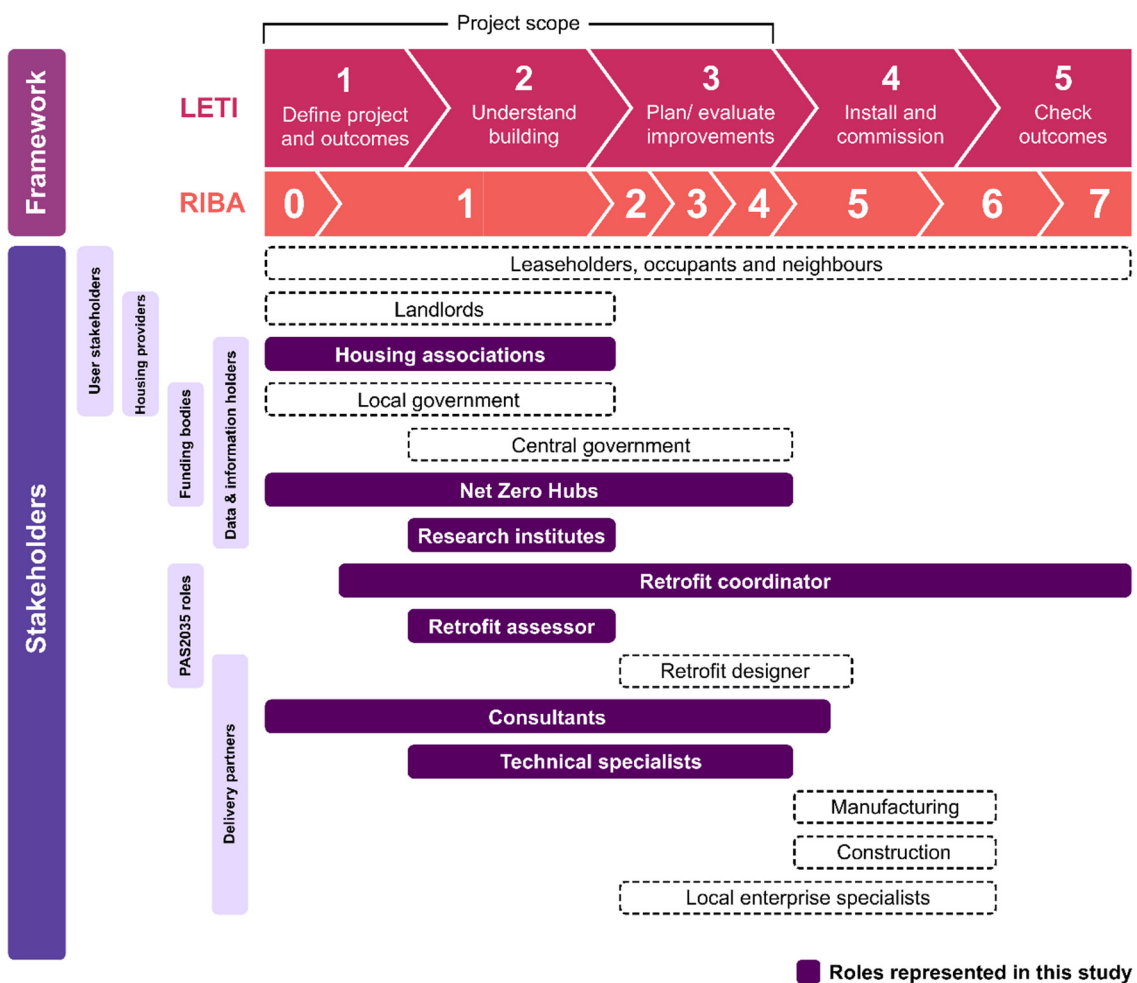
Figure 1. Outline of the different stages of the retrofit process, as defined in the LETI Climate Emergency Retrofit Guide, with corresponding RIBA Plan of Work Stages. Different actors and stakeholders and their engagement with the process are presented. The stages and roles represented in this study are also highlighted.

2.2. Data and Modelling

Throughout the retrofit process, many types of data and models are both used and generated, including building surveys, stock models, and EPCs, as well as direct measurements made by surveyors or retrofit assessors during design stages. In this study, the terms' data and models are distinguished to provide structure in the interviews. Data are defined as true and measured information from a building, examples of which might include metered energy use or building dimensions. On the other hand, models are considered to produce a representation or simulation either before or after retrofit: these include

building energy models such as SAP, as well as the outputs of such models. For example, the estimated energy consumption in an EPC is considered a model, but the measured information entered into SAP to produce that EPC represents the data.

Despite the prevalence of data and modelling throughout the retrofit development process, issues with the quality of, and confidence in, the metrics and measures used to make decisions are long outstanding [2]. For example, Hardy and Glew (2019) catalogue several errors in the EPC register, ranging from incorrect energy estimates to missing data [13]. Contemporary research into these problems has sought to address these issues by developing new models and associated data collection processes, such as D'Angelo et al. (2022) who developed a workflow based on building information modelling (BIM) [14], and Ward et al. [15] who proposed a framework for estimating energy efficiency over a large scale using mobile sensing. Large scale stock models have also been developed to present a data-rich model of residential homes [18]; this paper found that the availability of validated data on housing stock, in particular, is lacking.

There is broad consensus that paying greater attention to developing high-quality data and models is necessary to implement effective retrofit at a large scale [12,19]. Different impacts and effects of data and models for characterising and designing aspects of retrofit have been widely studied. For example, Estrella Guillén et al. [20] compared different metrics for benchmarking energy and comfort and showed that they are not interchangeable, suggesting that different metrics can drastically affect the perception of a building's performance. Likewise, Fawcett and Topouzi [21] investigated the role of metrics used in domestic retrofit and whether they are fit-for-purpose, finding that there is a particular importance in the choice between measurements of carbon and of energy, which is corroborated by [22]. Simpson et al. (2020) performed a bibliometric analysis of data-centric research in retrofit across Northwestern Europe, looking at computational models, monitoring frameworks, and statistical analyses [23]. They identified recent trends in data-driven approaches to energy performance, heat and power, the indoor environment, and retrofit practice, and they highlighted the potential improvements more data can provide, especially on the impacts of retrofit.

Understanding the needs of practitioners in retrofit is an area of research previously explored by [24]: they conducted a series of interviews focussed on understanding the capabilities of actors in the area. One of the themes identified in that study was the accessibility of information, although this was primarily focused on developing the knowledge of the practitioners themselves. Fylan and Glew (2022) used focus groups to study barriers in the installation of retrofit interventions [25]. The primary focus of the research in this paper works in parallel to these studies, investigating the needs of stakeholders from data and modelling in domestic retrofit planning. Given the importance of retrofit decision-making supported by data and modelling, the study focuses on stakeholders who work in the planning stages of retrofit interventions, where data and modelling results have a critical effect on the work conducted. The roles and responsibilities of these stakeholders were limited to the first three stages of the LETI retrofit process, pertaining to the project definition, understanding, and planning stages. Key stakeholders within the domestic retrofit sector who work or have worked across multiple properties, as opposed to individual homeowners with limited experience, were identified based on a direct search and through recommendations. The project scope and categories of stakeholders involved in the study are highlighted in Figure 1, which is further expanded on in the next section.

3. Materials and Methods

This study was designed to investigate the following research question: what are the data- and modelling-related needs in the preparation phases of retrofit projects for residential buildings?

In framing the research question, data and models were distinguished to identify themes that related to both the acquisition and use of measured data, and in the use of models and their respective outputs. The concepts of the two are entangled, and the

outputs of models are typically referred to as data; similarly, stock models, for example, may be solely formed of true measured information. Clear definitions of retrofit data and models, as outlined in Table 1, were used with all participants in the study to ensure mutual understanding between the research team and participants.

Table 1. Grounding definitions given to the participants at the start of the interview.

	Definition	Examples
Retrofit Data	True and measured information regarding the houses being retrofitted.	Such as the results of metered energy use and cost inflows/outflows.
Retrofit Model	A model that produces a representation or simulation of energy consumption before/after retrofit; material use; cost of retrofit; and other building performance metrics.	Such as building energy models (like SAP, EnergyPlus) or financial modelling.

Stakeholders working in the early stages of the LETI retrofit process in the domestic retrofit sector were contacted with a request to fill out a survey regarding their use of data and modelling in retrofit, and a subset of respondents that met the inclusion criteria then participated in a qualitative interview. The research presented is based on an analysis of these interviews. Thematic analysis allowed for depth in the understanding of a participant's experiences with data use in the retrofit sector. Within the field of building and construction research, qualitative methodology has been similarly employed, particularly in the context of data and modelling adoption and implementation [24,26].

The sampling criteria for participant recruitment for qualitative interviews were to have worked or be working in the retrofit sector for domestic retrofit with experience using data as part of that role. Refined sampling criteria required stakeholders to have primary experience in the planning stages of domestic retrofit projects. The refined sampling criteria allowed the study team to focus on a subset of the retrofit journey pipeline and focus on professionals in the retrofit industry with multiple experiences in various projects, excluding homeowners and occupants. Potential individual participants were contacted from the authors' network. Relevant organisations identified from online searches, reports, and academic literature were also contacted using a two-stage recruitment approach where suitable participants were then identified by the organisation. Individuals and organisations were contacted via email. Additionally, snowball sampling was used where contacted parties were invited to suggest potential participants for the study.

A total of 87 individuals and organisations were contacted and invited to be part of the study with the following breakdown: 39 organisations in the private sector, 26 governmental organisations, 18 voluntary organisations, and 4 academic laboratories. Thirty-two participants responded to the sampling survey which served as a filtering point for eligible interview participants, of which eight were recruited for in-depth interviews. Demographic information about the interview participants is detailed in Table 2.

This study was approved by the University of Sheffield's Ethics Review Board, as administered by The Department of Civil and Structural Engineering. Consent was obtained from each participant before the interview.

A semi-structured interview protocol was developed based on the literature review and the research question. Each interview aimed to obtain detailed stories and experiences of where data and modelling are used in the domestic retrofit sector, as well as the industry's needs from data and modelling for the future implementation of retrofit. The interview protocol was developed iteratively, following four pilot interviews conducted to improve the protocol.

Table 2. Interview participants, detailing their job title; stakeholder category; and with what stages of the LETI retrofit process their work is involved.

ID	Job Title	Stakeholder Category	LETI Stage (s)
P1	Retrofit consultant and net zero technical analyst	Consultancy	1, 2, 3
P2	Director of research group	Research Institute	2, 3
P3	Programme manager—retrofit	Housing Partnership	1, 2, 3, 4, 5
P4	Research assistant	Research Institute	1, 2
P5	Data and information manager	Public Sector Body	1, 2
P6	MCS manager and retrofit coordinator	Community Interest Company	1, 2, 3, 4, 5
P7	Retrofit assessor	Social Enterprise	2
P8	Joint head of sustainability	Architecture firm	2, 3

The interview was semi-structured: a set of topics was created based on the research question [27], and follow-up questions were tailored to participant responses. At the start of each interview, the researcher gave the participant the definition of "data and modelling within retrofit"; these definitions are provided in Table 1. The structure of the interview was as follows: introduction; background questions; what you do with data and modelling, e.g., types of data and models used, mapping exercise, deep dive into an example; what is needed from data and modelling, e.g., motivations, access, quality and confidence, stakeholders involved; and wrap-up questions. The interviews lasted between 45 and 75 min, with an average time of 60 min, and were conducted both in person and online.

After automated transcripts were manually checked and de-identified, an inductive thematic analysis method was adopted to analyse the interviews, which allowed themes within the data to emerge [28,29]. In adopting a needs assessment approach to the thematic analysis where barriers and enablers of data and modelling of retrofit were investigated (following the research question), emergent themes included lived experiences of how data and modelling were being used and complaints about data and modelling, or experiences where the limitations of data and modelling hindered a participant's progress. The themes were collated into a codebook that was iterated through several rounds of analysis. Two members of the research team coded the same three interviews, selected randomly, to establish inter-rater reliability [30]. The resulting transcript-level inter-rater reliability was found to be 100%, as both researchers identified the same set of themes in each transcript.

The interviews were also coded with which types of data and modelling respondents indicated that they used, e.g., where the participant collected or used data, or created or used a model or its results.

4. Results

4.1. Characterisation of Study Participants' Professional Experiences with Retrofit Data and Modelling

To further contextualise the qualitative analysis, the types of data and models used by interview participants in their professional roles, which formed the basis of the experiences discussed, are summarised in Figure 2. In the interviews, participants were asked to organise data and models that were commonly used in their profession into categories, which led to the categorisation presented in Figure 2. Those types of data and models that featured most frequently in both survey responses and during interviews are outlined in each category. Definitions used for each category are provided in Table 3.
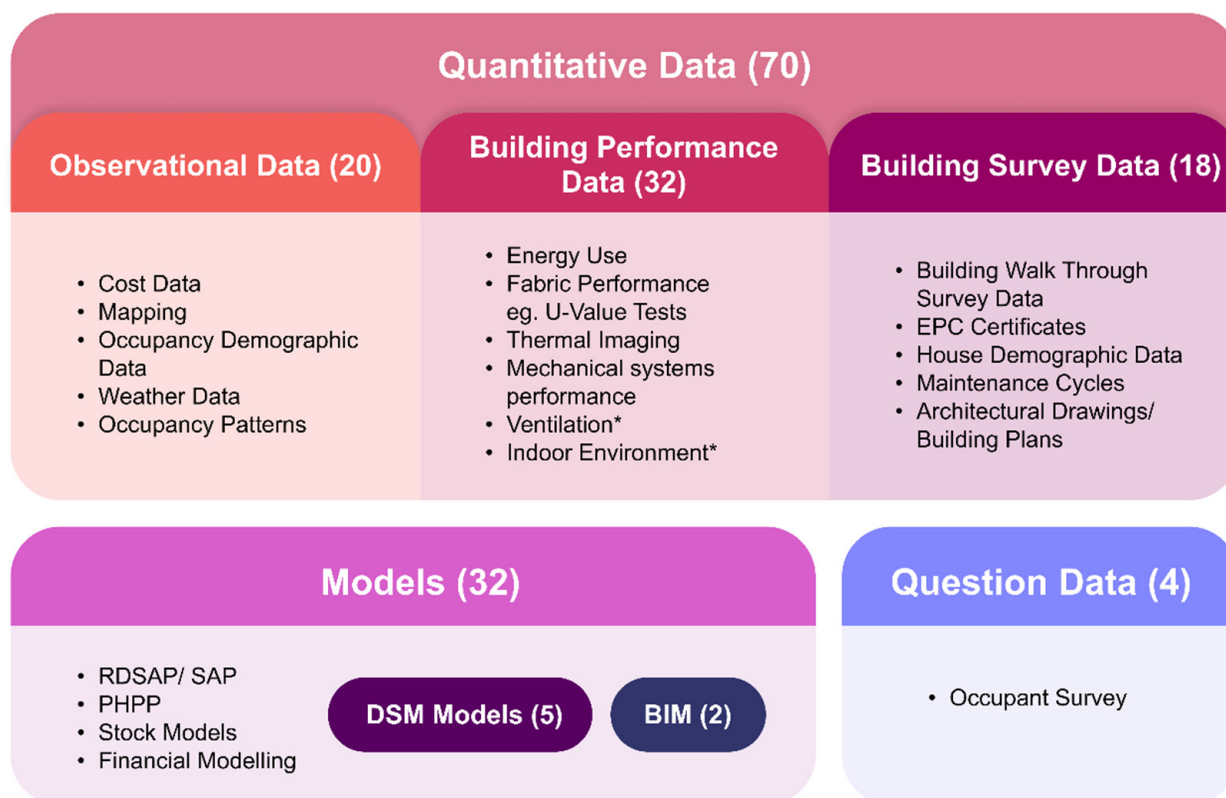
Figure 2. Map of data and models commonly used by participants in their profession. Counts represent the number of mentions across all different data and modelling tools and across all participants in both the survey responses and subsequent interviews, for each category. The five most mentioned data and modelling tools for each category are listed and (*) indicates equal counts.

Table 3. Definition of data and model map categories.

Name	Definition
Quantitative Data	Measurements and direct observations concerning the building and its surrounding.
Observational Data	External factors that influence the building.
Building Performance Data	Data related to how the building is functioning.
Building Survey Data	The physical features of the building.
Models	Simulation software and planning tools used in the retrofit process.
BIM	Building information modelling (BIM) uses various tools and technologies to store a lot of information within a 3D model.
DSM	Dynamic Simulation Models (DSM) model dynamic processes. For example, Energy Plus simulates energy consumption over time, in the context of energy efficiency.
Question Data	Findings from interviews and questionnaires with occupants and stakeholders.

4.2. Thematic Analysis

Nine themes captured what stakeholders do with and need from data and modelling in retrofit. Participants described both their current practice using data and modelling in retrofit and what they identified as key issues and opportunities for improvement. The themes, their definition, examples from the interviews, and excerpts from interview transcripts are presented as a thematic codebook in Table 4.

Table 4. Inductive codebook. The counts are reported both per interview (I) and per excerpt (E). (I) counts how many interviews were coded to a specific theme. (E) counts how many distinct interview passages (sometimes from the same interview) were coded to a specific theme.

Theme	Definition	Examples	Quote	I	E
A. Data and modelling are [not] trustworthy if…	Participants find the quality of secondary data and modelling results used to inform retrofit actions and input into models to be poor. Participants only have confidence in data and modelling under certain conditions.	Untrustworthy data and models included inaccurate EPC ratings; outdated costing data limited to certain typologies. Trustworthy data and models included data collected under controlled conditions; regulated and standardised data collection equipment; and data collected by a trusted surveyor; where only a subset of retrofit measures are modelled.	"There's a pre-EPC calculation and the difference might be one or two EPC bands commonly". P5	8	10
B. Data increases confidence in retrofit work	Participants collect, analyse, and communicate data before and after retrofit to increase the confidence of clients, occupants, and themselves in the retrofit work carried out.	E.g., thermographic images facilitate communication between residents, financial stakeholders, and contractors before retrofit; energy use data ensure that discrepancies post retrofit can be addressed, such as poor occupancy behaviour and MEP performance.	"I've used [thermal imaging] to show that cavity wall insulation has degraded or slumped because you can see a change in temperature in a wall". P7	6	11
C. Better data collection methods are needed to help inform retrofit actions	Participants find current data collection methods are invasive, unreliable, and unacceptable to occupants, limiting opportunities for participants to collect usable data to inform retrofit actions.	E.g., drop-out data due to malfunctioning technology; invasive methods of measurement such as U-Value tests leaving marks on walls; using a drill on the wall to better understand the building fabric.	"There often is dropout data [with data monitoring systems] or a bit of data gets lost or there is a big spike or something in the data. And I find I always have to do a bit of checking of the data and deleting those kinds of outliers. But […] where's the limit of is this an outlier bit of data or is this actual data?" P1	5	13
D. More opportunities to input specific data into models are needed to increase model accuracy	Participants believe there are not enough opportunities to input highly specific data into retrofit models leading to the use of outdated inputs in models.	E.g., rdSAP/SAP carbon predictions are inaccurate due to outdated carbon factors; rdSAP/SAP provide assumptions such as the performance of built elements which impacts the accuracy of results.	"One of the things that's been wrong with the SAP data is the carbon intensity estimates of electricity. They haven't really changed for quite a long time, yet, we've decarbonised our electricity supply quite a lot. So, things like heat pumps will come out with worse performing SAP points because SAP is still using estimates for electricity being generated from gas, oil, coal". P3	5	11

Table 4. *Cont.*

Theme	Definition	Examples	Quote	I	E
E. Data and models are used to target retrofit measures within the financial constraints	Participants discussed using data and models to target which homes to retrofit and what retrofit measures to implement to best use a limited budget.	E.g., EPC certificates are used to target particular houses which are eligible for funding; rdSAP model is used to target particularly effective measures.	"Because the cost of installing measures has gone up and the funding from the treasury hasn't, we have to also look for perhaps areas which won't all be very large attached houses because that's very expensive to retrofit". P5	5	9
F. Occupant-related data impacts retrofit work	Participants change the retrofit design based on occupancy-related information.	E.g., occupants have different priorities, such as the aesthetics of wall insulation, which informs the retrofit measures; whether the property is owner-occupied or social housing impacts the retrofit measures implemented; when undertaking a PHPP model changing the occupancy patterns greatly impacts the results.	"We as a design team were being asked to guarantee the performance of the finished building. And it became apparent in our PHPP model that if you change the occupancy, you radically change particularly the risk of overheating, which is actually the bigger challenge. So, it became apparent that it was important that we actually found out how many people live there, not just assumed, based on the floor area or the bedrooms". P8	4	5
G. Data and modelling could help inform and educate occupants (speculative)	Participants discuss how occupant behaviour can undermine the impacts of retrofit measures. Participants identify a lack of education of occupants on retrofitting measures and show that data collection and monitoring can be used to assist occupant education.	E.g., occupants not knowing how to use the new equipment (thermostat) in their homes, resulting in higher energy usage than predicted; occupants not running the washer when PV panels are operating; data monitoring to inform occupants of the cause of mould within homes.	"If [social housing landlords] have that data to hand, they can just say—Look you put your fan on once in the past month and this room hasn't been heated at all and that's why you've got all the black mould". P6	4	5
H. Models are too expensive to purchase (speculative)	Participants discuss the high cost of purchasing models, preventing better models being used more widely.	E.g., when trying to purchase models for large amounts of data, BIM, particularly Revit.	"[Models] are pretty robust but they are quite expensive, so how do you get the sweet spot between something being robust and people being able to take decisions on the basis of them, at cost". P2	3	3
I. Occupants are excluded in data monitoring	Participants do not monitor or retrofit buildings with occupants in them due to an increase in complexity.	E.g., co-heating tests involve heating a home to 25 degrees which is assumed as unacceptable for occupants to live through; large sample sizes needed to account for vast variation in occupant behaviours and add difficulties in monitoring occupied homes.	"Because there are so many occupant angles that affect energy use, it becomes impossible [so] we are all focusing purely on building fabric". P4	2	3

4.3. Trade-Offs between Precision, Confidence, and Burden of Collection in Data and Modelling

In theme A "Data and modelling are [not] trustworthy if…", participants described specific conditions under which they trust data collection, where results from models relating to a dwelling are untrustworthy because they are often outdated or inaccurate, as illustrated in the quote by P5 in Table 4. Participants also described their confidence in standardised and controlled data collection methods, as described by P2, and in model outputs, as described by P8 (excerpts provided below).

> *"I'm reasonably confident things like smart meter data have to be right on the whole, they're not 100%, but 99.99% you get the right data from there because it's so standardised and regulated".* P2

> *"I do [have confidence] if it's the right model and I've been sufficiently in control of what's gone into that model".* P8

Notably, P6 and P8 both described having confidence in model outputs if they had confidence in the data that informed the model, either by inputting it themselves or by trusting the person who did.

In theme D "More opportunities to input specific data into models are needed to increase model accuracy", participants called for more opportunities to enter data and modify data in modelling. In the excerpt below, P1 described how some models propose default values, in this case, for U-values, which does not encourage retrofit stakeholders to question the actual building elements.

> *"I think it's just easier in SAP or rdSAP to put the default value in or use an assumed value because there's a lot of guidance on that as well… If you've got something that's constructed in 1950 and it's made of cavity wall, here's the U Value you can use. You go oh great I'll just use that. You know you're not quite thinking, is it a really thick cavity or is it a really thin cavity?"* P1

Other participants described how micro-climate data and heat pump data cannot be entered into the models they use (P7 and P6, respectively).

In theme C "Better data collection methods are needed to help inform retrofit actions", while acknowledging some advances in data collection methods, participants described the current retrofit measurements as too invasive for occupants and discussed potential ways of performing measurements in less invasive ways while trading off on precision. For example, P1 discussed that drop out data and loss of data can occur during collection (see quote from Table 4). Outliers from data collection need to be manually verified.

Some participants discussed their assumptions around the acceptability of certain data collection methods. For example, P4 discussed how it is assumed that it would not be acceptable for occupants to live in the building while it is being retrofitted and that even testing is difficult with occupants present:

> *"The heat flux plates are the most reliable ones we've found. But the problem is then: Is a wall uniform all the way along it for the U value? And then actually it probably isn't. So how many heat flux plates do you need? […] And then you just end up in an absolute data nightmare […] That's the person who has the passive house. And we can put all these things on your walls and do U values. And she's like, "does it damage the walls?" [Answer] "Well it might leave a sticky mark" You might need to redecorate". You think, if that's just from one test of your house over a weekend. When we test for the co heating you heat someone's house up to 25 degrees, you don't want that. You can't live like that and it's too expensive to move them out". […]* P4

In theme F "Occupant-related data impacts retrofit work", participants described the difficulty of designing retrofits for 'unknown' or 'average' occupants, which brings additional uncertainty into their predicated outcomes, as described by P8 in Table 4, where changes in occupancy numbers radically change the results of model outputs. As further described by P6 below:

"Where you've got social landlords, it's alright focusing on the needs of the current occupant, but there's nothing to say that they're going to be living there in two years. [. . .] You're wanting to [design] for the conditions of your average person who's going to be renting that dwelling, not just one person who might have like idiosyncrasies in the way that they use that property". P6

Lastly, in theme I "Occupants are excluded in data monitoring", participants described how monitoring practices are changing towards ignoring occupancy altogether. Rather than monitoring over 18 months or 2 years, P2 described changing practices towards pre- and post-retrofit assessments that are not based on continuous monitoring once occupants are in the building.

"We tend to do more building performance [tests]: 'Does the building work?' rather than 'What's the occupant doing?' But we used to do that, we would monitor buildings for 18 months and then two years. We are doing less that and doing more physical building performance type stuff. So pre and post [retrofit], because [. . .] they are very hard to manage". P2

4.4. Who Accesses the Data and Modelling Results in Retrofit?

Participants described using data and modelling results to choose what retrofit measures to undertake and to check that the results of retrofit measures were as expected, thereby increasing their confidence in the retrofit work, as demonstrated by theme B "Data increases confidence in retrofit work". For example, P4 found that using thermal images with contractors helped them to see where the gaps were and represented a good visual tool. They claimed that contractors found it useful to see the impact of the retrofit measures that were being installed.

"And I think when we've been doing the building work, what's been most useful to the contractors is our thermal images. So, they want to know that, you know, cause what, so we've been doing like air tightness retrofits on some properties, which basically just involves the contractor, trying to just bung up as many holes as possible in the leaky house we were doing. Um, but you know, if they could see where the gaps are, where the leaks are, it helps". P4

Participants also described several instances of using data with multiple additional stakeholders in theme B "Data increases confidence in retrofit work". These included occupants, financial stakeholders, and contractors. For example, P1 described holding manufacturers accountable and purchasing a better heat pump, as the one installed was not working as manufacturers said it would.

"We found that, you know, bits weren't working quite like they were expected. [. . .] So, we're in the process of resolving that. So, making sure that the heat pump gets switched out for one that is as efficient as the design said it was going to be. It's their responsibility because there's the measured data to be able to say that that wasn't as good as the design said". P1

In another instance, P7 discussed that being able to better *show* the impact of retrofit measures through, e.g., 3D models, rather than just verbally describing them would help to inform occupants and get them on board to 'better target' retrofit measures.

"It [would] be nice to be able to generate that sort of [results of model] in 3D or even 2D without having to learn an entirely new skill [. . .] It would just be to be able to more accurately show a client where the weaknesses are in a dwelling. And thus, be able to better target interventions. [. . .] So, some of the local authorities are going around trying to do thermal imaging as an engagement process with their residents. So, they're planning to go down the street in the winter and go, "Oh, looks like you've got a few leaky bits on your house, would you like to sign up to our retrofit scheme?" P7

In theme E "Data and models are used to target retrofit measures within the financial constraints", participants use data and modelling to carry out the most impactful retrofit

measures, demonstrating that data and modelling, in the hands of financial decision makers, can help to orient the 'right' retrofit measures. Theme H "Models are too expensive to purchase" also refers to the cost of models, which could be prohibitive.

Speculatively (since only two participants had experience engaging with occupants), participants described how there are missed opportunities for data and models to be leveraged with occupants to explain how to best live with the new retrofit installations, describing a high potential for data and modelling as educational tools (theme G "Data and modelling could help inform and educate occupants"). For example, P3 discussed how more education of occupants on retrofit technologies can lead to higher impacts, using the example of solar PV usage optimisation.

> *"What are we doing about teaching people about how they can utilise their solar PV better are we just putting it in? And they just get home from work and at seven o'clock, they're sticking their washing machine on because that's what they've done. And they're not utilising the free energy. You can't change people's behaviour unless they know why they should be doing it. And if you can take them through that process, you can help them make an impact".* P3

5. Discussion

A prominent theme in the analysis and subsequent codebook was the participants' confidence in the data and models used for retrofit planning. Greater control over data collection and its use in models bolstered this confidence. Direct data collection, for instance, was perceived to enhance reliability (theme A "Data and modelling are [not] trustworthy if..."), and more flexible data input points in models instilled confidence in their output accuracy (theme D "More opportunities to input specific data into models are needed to increase model accuracy"). Indeed, there was concern that aiming for user-friendly models, characterised by pre-set variables, might compromise robustness. In line with this concern, Parker et al., 2021, noted that RdSAP's data inputs are often oversimplified and hence imprecise, advocating for more adjustable inputs like DSMs [31].

Conversely, collecting retrofit data is a major limitation of housing stock modelling [32]. Participants voiced apprehensions about the intricacy and intrusiveness of data collection methods, making them difficult to execute (theme C "Better data collection methods are needed to help inform retrofit actions"). For instance, while P1 advocated for integrating accurate thermal transmittance (u-values) in SAP models, P4 underscored the challenges in measuring them. Additionally, measuring retrofit data directly created occupant acceptability hurdles for participants (theme C "Better data collection methods are needed to help inform retrofit actions"). Many data collection processes disrupt occupants [11,31], with private homeowners finding retrofit interventions notably more intrusive than social landlords (Theme F "Occupant-related data impacts retrofit work"). Nevertheless, some literature has shown that residents' tolerances vary [6], which suggests an opportunity for the strategic targeting of both dwellings and occupants. Simultaneously, more efficient collection and modelling methods are in demand [11,31].

Striking a balance between precise data collection and the complexity of its collection is pivotal for effective retrofit actions. Similarly, models must allow customisation without becoming overwhelmingly intricate. Parker et al., 2021, suggest examining RdSAP assumptions for potential improvements in precision, identifying specific values—like u-values—that significantly affect accuracy [31]. Refining default values could pave the way for intermediate data collection methods, enhancing model accuracy without the hefty costs and time of field measurements. Parker et al., 2021, hint at expanding default values, including material properties, to strike this balance [31].

In exploring variations in data and model use across stakeholder groups, the possibility for expanding the use of underutilised data and models arises. Residents are often not provided with access to comprehensive information and guidance on how to effectively adapt to retrofit installations due to cost barriers (theme H "Models are too expensive to purchase") or broader exclusions (theme I "Occupants are excluded in data monitoring").

The exclusion of residents creates potential barriers to their ability to be more involved and aware of the retrofit process and its outcomes, thereby impacting their agency and the overall impact of the retrofit undertaken.

Occupant behaviour significantly influences retrofit outcomes [33], and newer models increasingly account for this behaviour [34]. However, in properties with high occupant turnover, design often caters to the 'average' resident, diminishing retrofit efficacy (theme F "Occupant-related data impacts retrofit work"). A proposed solution is to use data and modelling insights to educate occupants about optimising their living environment (theme G "Data and modelling could help inform and educate occupants"). Participants, in theme B "Data increases confidence in retrofit work", even mentioned leveraging these insights to promote retrofit benefits, especially to owner-occupants, suggesting that occupants could be a focal point for data-driven retrofit education. Such initiatives, coupled with the broader climate crisis and the UK's cost-of-living crisis, could shift occupant perceptions about the acceptability of interventions.

Later in the retrofit process, data and model use could be extended further when engaging contractors (Theme B "Data increases confidence in retrofit work" and theme G "Data and modelling could help inform and educate occupants") to emphasise the importance of meticulous workmanship and foster dialogue about buildability between designers and builders, which can bridge the gap in retrofit implementation quality, as noted by the Zero Carbon Hub [35]. Therefore, making retrofit predictions and results transparent to stakeholders, including builders and residents, can foster a deeper understanding of necessary behavioural adjustments to maximise retrofit advantages. The power of robust data and modelling can be truly realised when it is made accessible to a broad audience.

Limitations

The number of participants was limited by the timeframe of the study and participant availability in this period; by the breadth of existing connections; and by the visibility of potential participants beyond the sampling strategies described in the methods, e.g., staff profiles or LinkedIn pages. Efforts to expand the interview pool were made, knowledge and contacts of the wider research community were leveraged, alongside the inclusion of an option for survey respondents to recommend other professionals working in the space. Small sample sizes are, however, common when utilising qualitative research methods and focus on facilitating in-depth explorations of each stakeholder's perspective within a specified segment of the retrofit pipeline working with housing providers. Such an approach enabled the creation of a codebook that resonated with and built upon the existing literature on data and modelling in retrofit. Despite this limitation, the findings in this paper contribute to a broader understanding of the needs of actors in domestic retrofit planning.

6. Conclusions

Through a qualitative analysis of semi-structured interviews of stakeholders working in the planning stages of domestic retrofit in the UK, this paper sought to characterise what the data-related needs in the preparation phases of retrofit projects are. A consensus among participants indicated that the reliability of data and modelling results hinged on specific conditions. Notably, confidence surged with standardised data collection overseen by credible entities and when there was trust in the underlying assumptions either through meticulous data entry or faith in the stakeholders overseeing these procedures. Furthermore, adapting retrofit measures in alignment with budget constraints was frequently encountered by the respondents.

The key findings in the paper highlight the common themes when critically appraising the use of data and models in the retrofit design process, including the perceived trade-offs that affect decision making. There are significant improvements required in the collection, curation, and usage of data, and the subsequent modelling used to design, optimise, and target retrofit interventions. Participants called for both improved occupant-friendly data

collection techniques, and opportunities to refine assumptions and data entry into models. When affecting any changes, however, there needs to be significant consideration made about the trade-offs between the precision, trust, and ease-of-use of any data collection.

Several opportunities for a more holistic understanding and study of data and modelling applications in retrofit emerged. For example, participants described using data to increase confidence in their work among diverse stakeholders, including occupants and clients. They postulated the potential of data and modelling as tools to inform and educate occupants about retrofit interventions in their dwellings, thereby amplifying the efficacy of such measures. Broadening the beneficiaries of data and modelling in retrofit may not only amplify the advantages of prevailing data and models but may also pave the way for innovative data collection approaches and model refinement. The data and models that participants described using, summarised in Figure 2, define the scope of the thematic analysis and can serve to highlight key areas where themes and recommendations can be targeted for future work. Future studies in this area could expand the number of stakeholders in the study to incorporate those that exist at later stages in the LETI guide (Figure 1), or explore the impact of the findings in this paper on, for example, residents.

Author Contributions: Conceptualisation, M.J.C. and W.O.C.W.; methodology, M.J.C. and D.A.; formal analysis, M.J.C., D.A. and C.T.S.; writing—original draft preparation, M.J.C., W.O.C.W. and D.A.; writing—review and editing, M.J.C. and W.O.C.W.; visualisation, C.T.S.; project administration, W.O.C.W.; funding acquisition, W.O.C.W. and M.J.C. All authors have read and agreed to the published version of the manuscript.

Data Availability Statement: The anonymised interview transcripts and survey responses are available upon request to the corresponding author.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. LETI. *LETI Climate Emergency Retrofit Guide*; Low Energy Transformation Initiative C.I.C.: London, UK, 2021.
2. UKGBC; Wheeler, J.; Alker, J.; Box, P. *The Retrofit Playbook*; UKGBC: London, UK, 2021.
3. Holmes, G.; Hay, R.; Davies, E.; Hill, J.; Barrett, J.; Style, D.; Vause, E.; Brown, K.; Gault, A.; Stark, C. *UK Housing: Fit for the Future?* Committee on Climate Change: London, UK, 2019.
4. Hamilton, I.; Milner, J.; Chalabi, Z.; Das, P.; Jones, B.; Shrubsole, C.; Davies, M.; Wilkinson, P. Health Effects of Home Energy Efficiency Interventions in England: A Modelling Study. *BMJ Open* **2015**, *5*, e007298. [CrossRef]
5. Heo, Y.; Augenbroe, G.; Graziano, D.; Muehleisen, R.T.; Guzowski, L. Scalable Methodology for Large Scale Building Energy Improvement: Relevance of Calibration in Model-Based Retrofit Analysis. *Build. Environ.* **2015**, *87*, 342–350. [CrossRef]
6. Haines, V.; Mitchell, V. A Persona-Based Approach to Domestic Energy Retrofit. *Build. Res. Inf.* **2014**, *42*, 462–476. [CrossRef]
7. Collins, M.; Curtis, J.A. *Willingness-to-Pay and Free-Riding in a National Energy Efficiency Retrofit Grant Scheme: A Revealed Preference Approach*; ESRI Working Paper; The Economic and Social Research Institute (ESRI): Dublin, Ireland, 2016.
8. Jafari, A.; Valentin, V. An Optimization Framework for Building Energy Retrofits Decision-Making. *Build. Environ.* **2017**, *115*, 118–129. [CrossRef]
9. Kim, J.; Nam, S. Do Household Time, Risk, and Social Preferences Affect Home Energy Retrofit Decisions in Korea? *Sustainability* **2021**, *13*, 4152. [CrossRef]
10. *British Standards Institution PAS 2035:2019*; Retrofitting Dwellings for Improved Energy Efficiency—Specification and Guidance. British Standards Institute: London, UK, 2022.
11. Patterson, J.; Cui, Z.; Coma-Bassas, E. *Developing a Practical Retrofit Early-Stage Survey Tool to Inform the Decision-Making Process for Existing Homes*; Project Report; Welsh Government: Wales, UK, 2023; ISBN 978-1-899895-28-1.
12. Tan, L.M.; Arbabi, H.; Ward, W.; Li, X.; Densley Tingley, D.; Khan, A.; Mayfield, M. Towards an Automated Workflow for Large-Scale Housing Retrofit. *Environ. Res. Lett.* **2023**, *18*, 6. [CrossRef]
13. Hardy, A.; Glew, D. An Analysis of Errors in the Energy Performance Certificate Database. *Energy Policy* **2019**, *129*, 1168–1178. [CrossRef]

14. D'Angelo, L.; Hajdukiewicz, M.; Seri, F.; Keane, M.M. A Novel BIM-Based Process Workflow for Building Retrofit. *J. Build. Eng.* **2022**, *50*, 104163. [CrossRef]

15. Ward, W.O.C.; Li, X.; Sun, Y.; Dai, M.; Arbabi, H.; Tingley, D.D.; Mayfield, M. Estimating Energy Consumption of Residential Buildings at Scale with Drive-by Image Capture. *Build. Environ.* **2023**, *234*, 110188. [CrossRef]

16. Cristino, T.M.; Neto, A.F.; Wurtz, F.; Delinchant, B. The Evolution of Knowledge and Trends within the Building Energy Efficiency Field of Knowledge. *Energies* **2022**, *15*, 691. [CrossRef]

17. RIBA. *RIBA Plan of Work*; Royal Institute of British Architects: London, UK, 2020.

18. Steadman, P.; Evans, S.; Liddiard, R.; Godoy-Shimizu, D.; Ruyssevelt, P.; Humphrey, D. Building Stock Energy Modelling in the UK: The 3DStock Method and the London Building Stock Model. *Build. Cities* **2020**, *1*, 100–119. [CrossRef]

19. Wade, F.; Visscher, H. Retrofit at Scale: Accelerating Capabilities for Domestic Building Stocks. *Build. Cities* **2021**, *2*, 800–811. [CrossRef]

20. Estrella Guillén, E.; Samuelson, H.W.; Cedeño Laurent, J.G. Comparing Energy and Comfort Metrics for Building Benchmarking. *Energy Build.* **2019**, *205*, 109539. [CrossRef]

21. Fawcett, T.; Topouzi, M. Residential Retrofit in the Climate Emergency: The Role of Metrics. *Build. Cities* **2020**, *1*, 475–490. [CrossRef]

22. Parkin, A.; Herrera, M.; Coley, D.A. Net-Zero Buildings: When Carbon and Energy Metrics Diverge. *Build. Cities* **2020**, *1*, 86–99. [CrossRef]

23. Simpson, K.; Whyte, J.; Childs, P. Data-Centric Innovation in Retrofit: A Bibliometric Review of Dwelling Retrofit across North Western Europe. *Energy Build.* **2020**, *229*, 110474. [CrossRef]

24. Simpson, K.; Murtagh, N.; Owen, A. Domestic Retrofit: Understanding Capabilities of Micro-Enterprise Building Practitioners. *Build. Cities* **2021**, *2*, 449–466. [CrossRef]

25. Fylan, F.; Glew, D. Barriers to Domestic Retrofit Quality: Are Failures in Retrofit Standards a Failure of Retrofit Standards? *Indoor Built Environ.* **2022**, *31*, 710–718. [CrossRef]

26. Zahrizan, Z.; Mohamed Ali, N.; Haron, A.; Marshall-Ponting, A.; Hamid, Z.A. Exploring the Adoption of Building Information Modelling (BIM) in the Malaysian Construction Industry: A Qualitative Approach. *Int. J. Res. Eng. Technol.* **2013**, *2*, 384–395. [CrossRef]

27. Weiss, R.S. *Learning from Strangers: The Art and Method of Qualitative Interview Studies*; Weiss, R.S., Ed.; Free Press: New York, NY, USA; Maxwell Macmillan International: Toronto, ON, Canada, 1994.

28. Braun, V.; Clarke, V. Using Thematic Analysis in Psychology. *Qual. Res. Psychol.* **2006**, *3*, 77–101. [CrossRef]

29. Patton, M.Q. *Qualitative Research & Evaluation Methods: Integrating Theory and Practice*; SAGE Publications: Thousand Oaks, CA, USA, 2014; ISBN 978-1-4833-1481-5.

30. O'Connor, C.; Joffe, H. Intercoder Reliability in Qualitative Research: Debates and Practical Guidelines. *Int. J. Qual. Methods* **2020**, *19*, 1609406919899220. [CrossRef]

31. Parker, J.; Fletcher, M.; Thomas, F.; Miles-Shenton, D.; Brooke-Peat, M.; Johnston, D.; Gorse, C.; Glew, D. *Demonstration of Energy Efficiency Potential; Literature Review of Benefits and Risks in Domestic Retrofit Practice and Modelling*; Department for Business, Energy & Industrial Strategy: London, UK, 2021.

32. Booth, A.T.; Choudhary, R.; Spiegelhalter, D.J. Handling Uncertainty in Housing Stock Models. *Build. Environ.* **2012**, *48*, 35–47. [CrossRef]

33. Wei, S.; Hassan, T.M.; Firth, S.K.; Fouchal, F. Impact of Occupant Behaviour on the Energy-Saving Potential of Retrofit Measures for a Public Building in the UK. *Intell. Build. Int.* **2017**, *9*, 97–106. [CrossRef]

34. Ben, H.; Steemers, K. Modelling Energy Retrofit Using Household Archetypes. *Energy Build.* **2020**, *224*, 110224. [CrossRef]

35. Zero Carbon Hub. Closing the Gap between Design and As-Built Performance. 2014. Available online: https://www.designingbuildings.co.uk (accessed on 30 August 2023).