

# Defining and identifying complex-to-decarbonise homes and retrofit solutions

Annex B – framework report



© Crown copyright 2023

This publication is licensed under the terms of the Open Government Licence v3.0 except where otherwise stated. To view this licence, visit [nationalarchives.gov.uk/doc/open-government-licence/version/3](https://nationalarchives.gov.uk/doc/open-government-licence/version/3) or write to the Information Policy Team, The National Archives, Kew, London TW9 4DU, or email: [psi@nationalarchives.gov.uk](mailto:psi@nationalarchives.gov.uk).

Where we have identified any third-party copyright information you will need to obtain permission from the copyright holders concerned.

---

# Contents

Executive summary	4
Introduction	6
Rationale for the Identification Framework	6
Framework aims and scope	7
Report structure	7
Glossary	8
Methodology: approach to developing the framework	9
Understanding existing approaches and datasets	9
Development of use cases	10
Definition of framework development principles	11
Attribute identification, mapping, and weighting	12
Overview of framework structure	12
Illustrative applications of the framework	14
Recommendations for further development	19
Appendices	20
Appendix 1: Dataset mapping	20
Appendix 2: Attribute identification, mapping, and weighting methodology	25
Appendix 3: Detailed framework components	33
Appendix 4: Framework workflows	34
Appendix 5: Application of the framework: methodology for data harmonisation	36

# Executive summary

## Annex introduction and purpose

This annex accompanies the main report for ‘Defining and identifying complex-to-decarbonise homes and retrofit solutions’. As part of the wider research undertaken to develop knowledge and evidence of approaches to upgrading underperforming and technically difficult to treat housing, this technical annex sets out our rationale and approach to developing a framework for identifying homes considered to be complex-to-decarbonise (CTD) within the UK stock.

The annex uses the terminology complex-to-decarbonise (CTD) homes, where relevant, to describe homes which have been identified as those with either one, or a combination of, certain physical, locational, occupant demographic, or behavioural attributes that prevent the effective decarbonisation of that home until they are addressed. These attributes might constrain the design and delivery of measures to improve energy efficiency, decarbonise heating, or realise occupant benefits (e.g., increased comfort and affordability of domestic heat and energy). These effects may be amplified by one or a combination of numerous system-level factors including financial (e.g., feasibility and affordability of measures), economic (e.g., supply chain and materials availability), and/or organisational capacity and capability (e.g., workforce skills).

This CTD terminology has been developed in this research and is informed by evidence which recommends several improvements on existing terminology such as ‘hard-to-treat’ and ‘hard-to-decarbonise’.

## Overall aim, approach, and methods

By enabling the identification of CTD homes, this initial framework intends to enable a wide range of analysis use cases defined by stakeholders. Examples of this include further analysis that can:

- Support the design and rollout of policies and energy efficiency programmes CTD homes (e.g., Capital Scheme Design, Social Housing Decarbonisation Fund)
- Inform the development of nuanced/ flexible approaches to energy efficiency measures for specific typologies of CTD homes (e.g., heritage/listed homes and those with traditional construction).

### **The specific objectives were to:**

- Develop a stakeholder-informed list of physical, technical and occupant attributes that render a home to be CTD.
- Address current data gaps in identifying CTD homes, by integrating a wide range of datasets that include variables relating to attributes that describe e.g., exposure.

- Create a framework with broad analysis coverage by considering data at various regional and national scales (e.g., national housing stock and devolved administration EPC data).
- Assess approaches to improve usability of EPC databases to inform stock analysis.

The framework was developed using an iterative process, informed by a Rapid Evidence Review (REA) and qualitative interview insights as well as feedback from Department for Energy Security and Net Zero (DESNZ) stakeholders. These informed the development of framework use cases and principles describing its focus (i.e., its aim) and intended functionality (i.e., how it may be used) and to generate a list of physical, technical and occupant attributes that may render a home to be CTD and weight them according to their relative importance.

## Framework functionality and outcomes

The resulting framework was designed to be data/dataset-agnostic (i.e., non-dataset specific) and modular (comprised of modules or 'analysis units' coded in Python). This versatile analysis environment automates data processing, allowing the framework to support a wide range of dataset types and to process large amounts of data efficiently. Such an approach also means that the framework can be customised or soft-linked to further datasets to suit the array of analysis needs highlighted by stakeholders.

The functionality of the framework is illustrated through two use cases. These demonstrate its ability to employ a spectrum of technical, physical and occupant attributes to identify homes that are considered to be CTD from a whole house retrofit perspective at both:

- The individual dwelling level, highlighting the relative possibility of each home being CTD.
- At stock level, where a series of datasets were 'layered' to determine the incidence of CTD homes at both regional and LSOA scales. This highlighted the distribution of these homes across England and Wales and specific areas where CTD homes were concentrated.

## Future pathways for framework development

To support more robust analysis of the impact of future policies on these homes, further development of the framework should be underpinned by in-depth engagement with other (non-policy) stakeholders. In addition, it is recommended that measures to improve the usability and quality of data that may be used to inform further analysis be introduced (e.g., systematically checking and correcting EPC data at source, provision of guidelines to improve consistency between datasets that are collected). Finally, the use of Unique Property Reference Numbers (UPRNs) has proved to be useful for linking datasets used in the framework their adoption as a standard identifier (where appropriate and in line with data protection requirements) in relevant future data collection exercises related to the UK housing stock should be considered.

# Introduction

As part of the wider research undertaken to develop knowledge and evidence of approaches to upgrading underperforming and technically difficult to treat housing, this report sets out the rationale and approach to developing a framework for identifying homes considered to be CTD within the UK stock. This annex accompanies the main report for 'Defining and identifying complex-to-decarbonise homes and retrofit solutions.'

## Rationale for the Identification Framework

Across the domestic sector, 'complex-to-decarbonise' (CTD) homes have been regarded as being more challenging to decarbonise as this will involve higher costs, barriers that are more difficult to overcome, or solutions that are more complex to implement. Recent research has highlighted that these homes which are estimated to be up to 28% of the residential stock are responsible for over 25% of all direct residential sector emissions in the UK<sup>1</sup>, these homes are therefore an emerging priority in the built environment's response to the climate crisis.

An in-depth Rapid Evidence Assessment (REA) was undertaken in Work Package 3 (WP3) of the overall project to collect and critically analyse the current knowledge on CTD homes. This highlighted that overall, there is little clarity of how to define, measure and identify these homes in the UK. As such, a key recommendation arising from the REA was to develop a framework for identifying CTD homes to enable more holistic analysis of these homes to support the formulation of relevant policy initiatives (e.g., Capital Delivery Schemes).

To gain real-world experience and insights on experiences, challenges, and barriers of retrofitting CTD homes as series of qualitative interviews with stakeholders was undertaken as part of Work Package 4 (WP4). Emergent insights from these interviews also highlighted a clear need for a flexible identification framework to support array of uses high-level applications for CTD homes such as:

- Signposting and providing steer on the guidance, standards and regulations that can be followed.
- Enabling the identification of potential risks or key considerations given property and context details – and through the home's project lifecycle
- Supporting consistency across suppliers, especially in able-to-pay markets and help address work quality issues.
- Assisting in the provision of consumer advice and guidance, as well as supporting resident and community agents to navigate challenges and manage contractors.

---

<sup>1</sup> Foster, S., Tahir, F., Orchard, K., Walker, I., Schwartz, Y., Raslan, R., 2019. Analysis on abating direct emissions from 'hard-to-decarbonise' homes. Committee on Climate Change.

## Framework aims and scope

This proposed framework aims to employ a systematic and data-driven approach that uses a spectrum of technical, physical and occupant attributes to identify homes considered to be CTD from a whole house retrofit perspective. To ensure that the framework is comprehensive and representative, our specific objectives were to:

- Develop an expanded list of physical, technical and occupant attributes based on feedback derived from the REA, qualitative interviews and through consultation with DESNZ stakeholders.
- Create a framework that offers UK-wide analysis coverage at various resolutions/scales by considering data such as Housing/Housing Condition Surveys and EPC data for devolved administrations, national and local authority level heritage and fuel poverty data.
- Explore how a framework may be developed to allow a wider range of datasets that include attributes that describe such aspects as exposure categorisation may be used to address current data gaps in identifying CTD homes.
- Assess approaches to improve dwelling level analysis such as using EPC dataset records and national statistics identifiers.

The defined scope of functionality of this framework is to support the identification of CTD homes within the UK housing stock. This scope defines the focus of an initial framework that provides a strong foundation to develop further with industry through engagement and testing.

## Report structure

The report is structured as follows:

- Executive summary
- Introduction: an overview of the rationale to developing the framework and defines its aims and scope
- Methodology: summarises the approach to developing the framework
  - Use cases
  - Framework development principles
  - Attribute identification, mapping, and weighting
  - Overview of framework structure
- Illustrative examples: an example of the application of the framework.
- Conclusions

- Appendices
  - Dataset mapping
  - Attribute identification, mapping, and weighting methodology
  - Detailed framework components
  - Framework workflows
  - Application of the framework: methodology for data harmonisation

## Glossary

**Table 1. Glossary of report terms**

Term	Definition
<b>Data agnostic (system)</b>	A system in which the format of data transmission is irrelevant to its function. This means that the system can receive data in multiple formats or from multiple sources, and still process that data effectively.
<b>Lower Layer Super Output Areas (LSOA)</b>	A geographic hierarchy designed to improve the reporting of small area statistics in England and Wales.
<b>Module/Modular</b>	A system that is constructed using similar units (modules) for flexibility and variety in use
<b>Use case</b>	Use cases define the function(s) the framework through the applications it may be used for and, where relevant, the interactions of intended users with it.
<b>Unique Property Reference Number (UPRN)</b>	A unique identifier for every addressable location in the UK.
<b>CTD Index</b>	A numeric indicator that denotes the relative possibility of a home being CTD.



# Methodology: approach to developing the framework

To guide the development of a flexible, comprehensive framework, our approach for developing the framework was based on an iterative process that incorporated findings from the REA (WP3) and emergent insights from the qualitative interviews (WP4) alongside feedback from DESNZ stakeholders. The steps involved are summarised as follows, and detailed in the following sections:

1. Understanding existing approaches and datasets: Findings from the REA relating to existing models and datasets were used as the basis of further research to identify, map and document data availability and limitations, and assess how earlier methodologies may have considered and categorised CTD attributes.
2. Development of use cases: Emergent insights from the qualitative interviews were used to develop high-level (initial) use cases for the framework. DESNZ stakeholder feedback was then used to refine these use cases.
3. Definition of framework principles: Use case feedback from the qualitative interviews and DESNZ stakeholders was used to define a set of framework principles describing its focus and intended functionality.
4. Attribute identification, mapping and weighting: Insights from both the REA and the qualitative interviews were integrated to inform the development of a list of physical, technical and occupant attributes and weight<sup>2</sup> them according to their relative importance.
5. Development and implementation of the framework structure: Building on an initial structure, feedback from stakeholders was integrated into the development of a final framework. The final framework was developed using a modular<sup>3</sup> approach which automated framework processes using customised Python<sup>4</sup> scripts.

## Understanding existing approaches and datasets

As part of the REA, examples of datasets and models that underpinned research into CTD homes were identified. REA findings pointed to the general absence of reporting on CTD homes and many of their associated attributes within the datasets used for building stock level analysis. This indicated that a variety of datasets would need to be analysed and combined to enable the identification of CTD homes for this framework.

---

<sup>2</sup> Weighting assigns a coefficient to a variable through which its effect on the calculation process to reflect its importance.

<sup>3</sup> A system that is constructed using similar units (modules) for flexibility and variety in use.

<sup>4</sup> An object-oriented, high-level programming language, which encourages program modularity and code reuse.

Therefore, as part of the development process, a review of existing datasets that may be used for identifying CTD homes was undertaken to analyse the format (e.g., spreadsheet, spatial), coverage (e.g., UK, Devolved Administration, City) and resolution of each (LSOA, individual dwellings. etc) and map the relevant CTD attributes included within them. The outcome of this analysis (see Appendix 1) provided an indication of the array of location/climate, housing stock, EPC, socio-economic and indexing datasets that the framework may potentially interact with and should therefore be designed to consider.

## Development of use cases

Use cases define the applications that a framework may be used for and, where relevant, the interactions of intended users with it. Three high level use-cases were identified in the initial feedback gathered through the analysis of qualitative interview insights:

1. Industry: Diagnostic – Diagnostic use case to help industry to identify CTD elements, be steered to guidance and select practical solutions.
2. Policymaker and asset owner: Mapping – Macro-assessment of the scope of the challenge.
3. Civil society and representative organisations: Knowledge and information – Nature of the problem and how it affects their stakeholders.

In line with the current project scope, the mapping-focused high level use case was prioritised for development. A consultation process with DESNZ stakeholders (see Box 1) provided detailed feedback on the array of applications in which the identification framework may be used. These encompassed functions such as supporting guidance for capital scheme design and informing cost-benefit analysis.

### **Box 1: Stakeholder use case consultation questions**

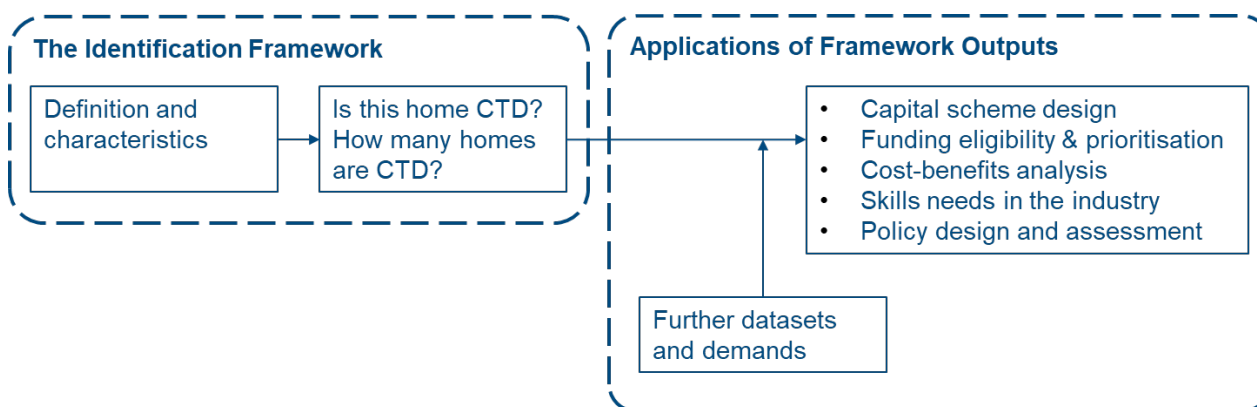
1. What use cases you can see the identification framework being used in your policy areas?
2. What data/intelligence gaps do you currently see which a framework could help with?
3. Which audiences/stakeholders do you imagine could make use of a definition framework – and what are their requirements?

## Definition of framework development principles

The dataset analysis and use case feedback were combined to determine the following set of principles that underpinned the development of the framework:

- **Function:** The primary function of the framework is to aid in the identification of CTD homes. The rationale for this is that through enabling the identification of CTD homes, the framework may potentially be used (subject to the availability of suitable data) for a number of the use cases as described through the aforementioned stakeholder feedback exercise.
- **Analysis level(s):** The framework structure should allow the identification of CTD homes at both a) individual dwelling level (to determine if a particular dwelling is CTD) and b) housing stock analysis (to determine the incidence of CTD homes in a housing stock of a given size).
- **Data interaction approach:** The framework should allow interaction with an array of datasets, therefore be designed to be data/dataset-agnostic<sup>5</sup>. This approach ensures that the framework is versatile, not dependant on specific databases and can be customised<sup>6</sup> to suit the array of needs highlighted by stakeholders.
- **Structure and Environment:** The framework will be designed around a modular structure (formed of modules or 'analysis units') and coded in Python to support the automation of data processing. This will allow for flexibility, variety in use the implementation of multiple use cases.
- **Outputs and soft-linking:** The framework will not seek to directly output final 'results' regarding such aspects as cost of retrofit. It will however be designed to allow soft-linking to additional datasets by stakeholders to facilitate such analysis (Figure 1).

**Figure 1. Application of framework outputs**



<sup>5</sup> A system in which the format of data transmission is irrelevant to its function. This means that the system can receive data in multiple formats or from multiple sources, and still process that data effectively, whereby it can work with information received from heterogeneous databases, i.e., databases with dissimilar data formats.

<sup>6</sup> Biljecki, F., Chow, Y.S., 2022. Global Building Morphology Indicators. Computers, Environment and Urban Systems 95, 101809. <https://doi.org/10.1016/j.compenvurbsys.2022.101809>

## Attribute identification, mapping, and weighting

To identify and map, weight and script the attributes that underly the analysis processes used by the framework into the modular implementation structure, the following process was used. For full details on each of these stages and attribute tables see Appendix 2.

**Stage A: Categorisation of principal and proxy attributes**<sup>7</sup>: Findings from the REA were synthesized with insights from interviews to identify an initial pool of attributes that may render a home to be CTD. These were then filtered and categorised into:

- Principal attributes that directly determine/contribute to a dwelling being considered CTD, such as hard-to-treat insulation and space constraints.
- Proxy attributes that indicate the existence of one or multiple principal attributes, such as age band and dwelling type.

The mapping from proxy to principal attributes is primarily used by the framework to interface with housing stock databases, it can also inform assessment of individual dwellings as defaults when certain principal attributes are hard to determine.

**Stage B: Assignment of attribute weights**<sup>8</sup>: A simple evidence-based ranking scheme which aimed to reflect the relative importance of attributes was used to assign weights to the principal and proxy attributes identified in Stage A.

**Stage C: Scripting**<sup>9</sup>: The processes and data used to implement the framework was then scripted in Python to automate processes.

## Overview of framework structure

Figure 2 represents the schematic structure of the identification framework, which consists of four functional components (Interface 1, 2 and Module 1, 2) and three data components (Primary CTD Attributes, Proxy CTD Attributes, Individual CTD Result). The specific roles and functions of each is detailed in Appendix 3.

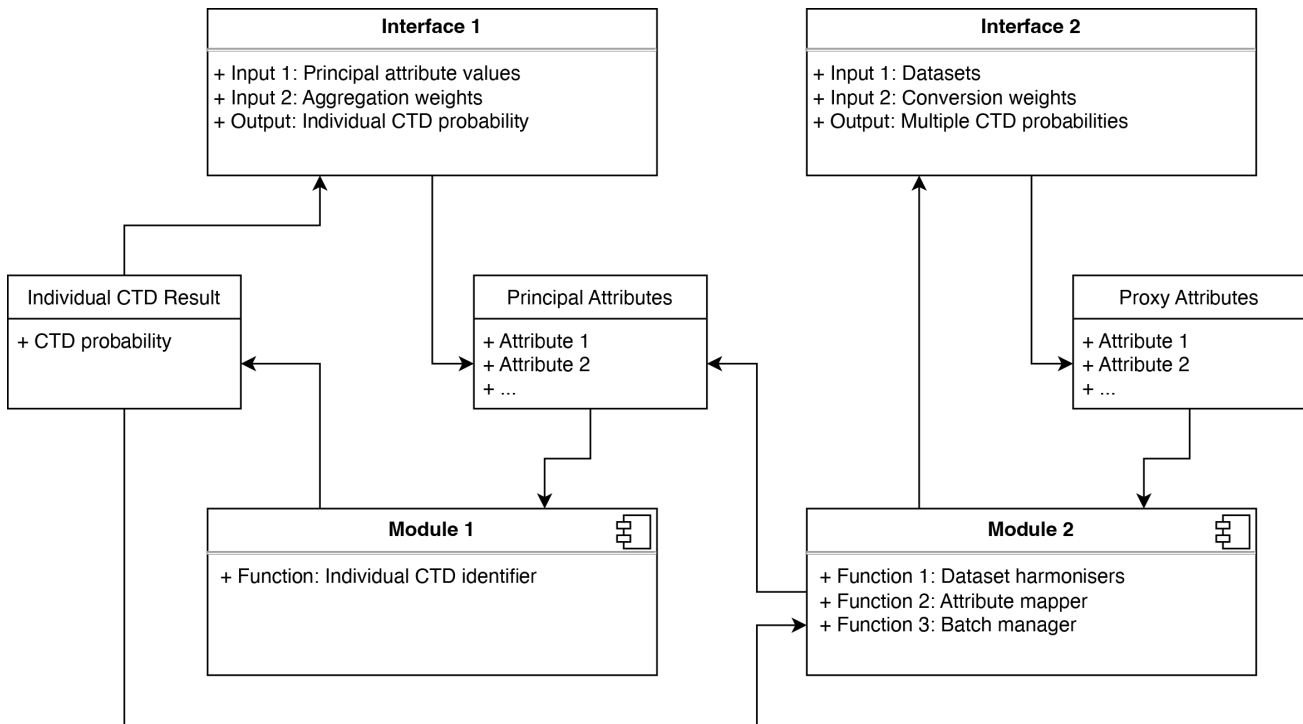
---

<sup>7</sup> This process is categorised as “low uncertainty / low customisation needed”, since evidence to support it was fairly abundant in the REA and interviews, therefore no or limited customisation will be needed to apply the framework to the specific use cases highlighted in the DESNZ stakeholder feedback

<sup>8</sup> This process is categorised as “high uncertainty / customisation needed”, since evidence to support from the REA and interviews is limited and/or highly context-dependent. Thus, a degree of customisation will be needed to apply the framework to the specific use cases highlighted in the DESNZ stakeholder feedback

<sup>9</sup> This process is categorised as “medium uncertainty / customisation needed”, since the script components regarding attribute conversion and aggregation is dataset-agnostic, yet those regarding interfacing specific datasets are subject to adaptation, especially when involving internal datasets. Thus, a degree of customisation will be needed to apply the framework to the specific use cases highlighted in the DESNZ stakeholder feedback

**Figure 2. CTD Identification framework structure**



To use the framework, the following implementation steps should be followed:

**A - For analysis at the individual dwelling level**, where a home may be assessed to determine if it is CTD

1. Input the principal attributes in a csv file.
2. Adjust the principal attribute weights if necessary.
3. Run the principal aggregator script.
4. Analyse/Utilise the resultant CTD Index.

**B - For analysis at the stock level**, where various datasets were 'layered' to determine the incidence of CTD homes at defined scale.

1. Gather all datasets to be used to derive proxy attributes.
2. Filter/correct the datasets if necessary.
3. Prepare the dataset harmoniser scripts for new datasets.
4. Adjust the batch manager script when new datasets are used.
5. Adjust the principle and/or proxy attribute weights if necessary.
6. Run the batch manager script.
7. Analyse/Utilise the resultant CTD Index dataset.

## Illustrative applications of the framework

This section illustrates the functionality of the framework through two use cases. While these use cases are not intended to be exhaustive, they demonstrate its ability to employ a spectrum of technical, physical and occupant attributes to identify homes that are considered to be CTD from a whole house retrofit perspective at individual dwelling level and stock level.

The framework workflows involved in implementing each of these applications are detailed in Appendix 4. For these illustrative cases, an analysis scope of England and Wales was defined. The datasets used to focus on this analysis scope, and the processes used to compile and filter them are described in Appendix 5.

Figures 3 to 6 illustrate key outputs summarising the outcomes of the two illustrative levels of analysis as follows:

- **Individual Assessment**<sup>10</sup>: Each individual dwelling identified within the analysis area (England and Wales) was individually assessed and an individual record of the assessment produced (Figure 3). The colour scale on the diagrams denotes the relative possibility of a home being CTD (CTD Index). For readability, a subsample of the analysis of individual dwellings in London is illustrated in Figure 4.
- **LSOA/Regional Distribution**: At the stock level, the identification results of assessed dwellings were averaged across LSOAs (Figure 5) and Regions (Figure 6) to create a map of the distribution of CTD homes across areas of different spatial resolutions.

These outputs can directly:

- Determine if a single dwelling is CTD and the degree of relative difficulty associated with its retrofit (based on a higher CTD Index).
- Highlight areas/ regions where there is a higher concentration of CTD homes and may therefore be prioritised.

Illustrative potential follow-on applications of these outputs that may be achieved through further soft linking by stakeholders are listed below in Table 2.

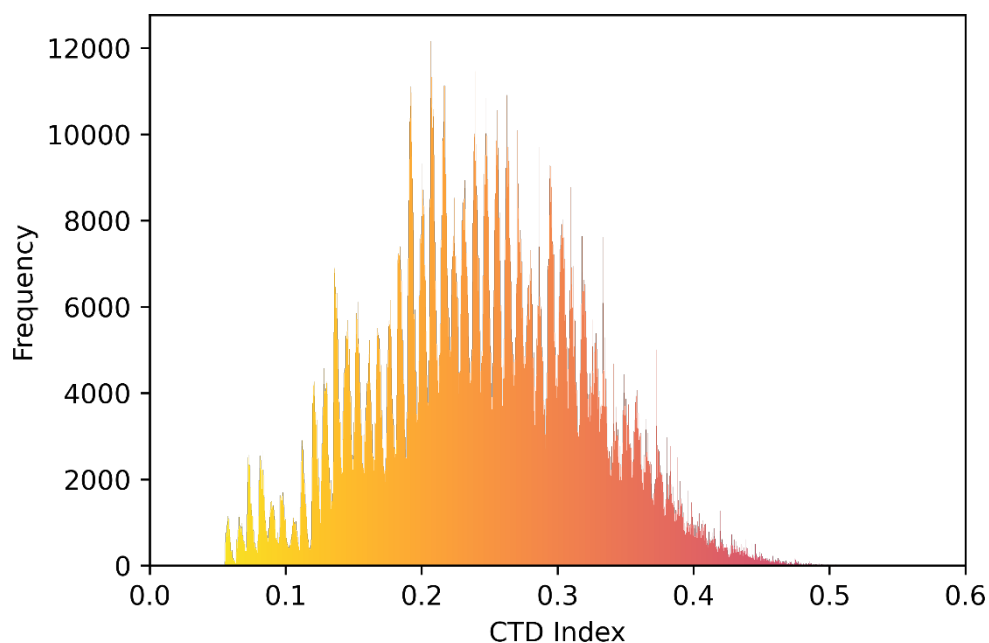
While these are not intended to be exhaustive of its potential uses, they provide examples that demonstrate how outputs may inform some of further analysis needs highlighted in DESNZ stakeholder feedback.

---

<sup>10</sup> It should be noted, based on the assumptions, thresholds and data used, this illustrative analysis found that 20-25% of the dwellings assessed were CTD

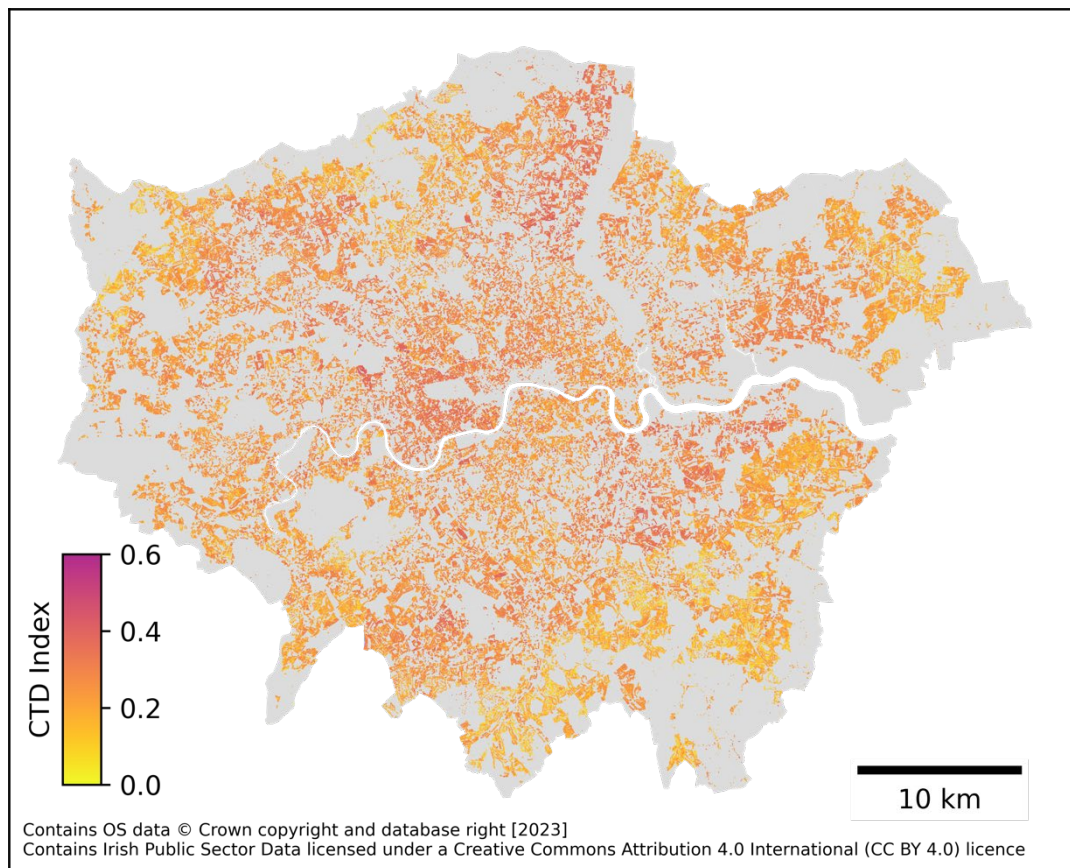
**Table 2. Illustrative further applications of framework outputs**

Illustrative application	Framework output	Types of further datasets required to be soft-linked by stakeholders	Potential further outputs
Cost of retrofit	Individual home assessment	Evidence-based cost databases (e.g., from SHDF projects)	Estimate of costs of retrofitting these homes based on degree of difficulty.
Cost of retrofit	Regional distribution of CTD homes	Regional cost indexes	Regional cost variability of retrofitting CTD homes
Skills needs	Regional distribution of CTD homes	Data on local construction skills/employment statistics	Local/ regional skills shortages
Eligibility	Individual home assessment/LSOA distribution	Data on deprivation/ LILEE	Eligibility of households based on such aspects as low income

**Figure 3. Individual dwelling assessment: Frequency and CTD Index**


Note: The scale shown denotes the CTD Index of a dwelling. This scale can be customised in line with specific analysis requirements.

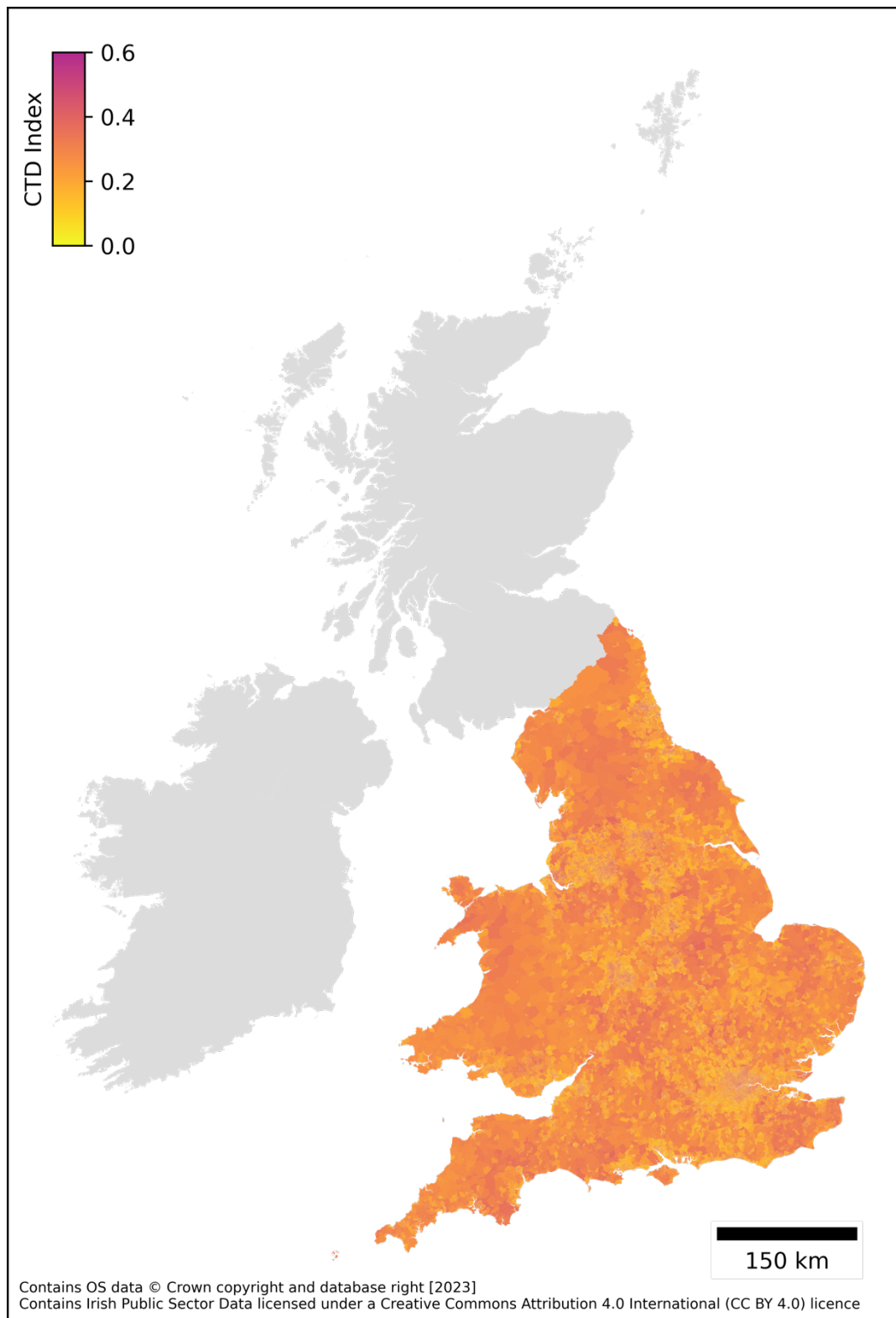
**Figure 4. Distribution of CTD homes in London: UPRN estimates**



Note: The scale shown denotes the CTD Index of a dwelling. This scale can be customised in line with specific analysis requirements.

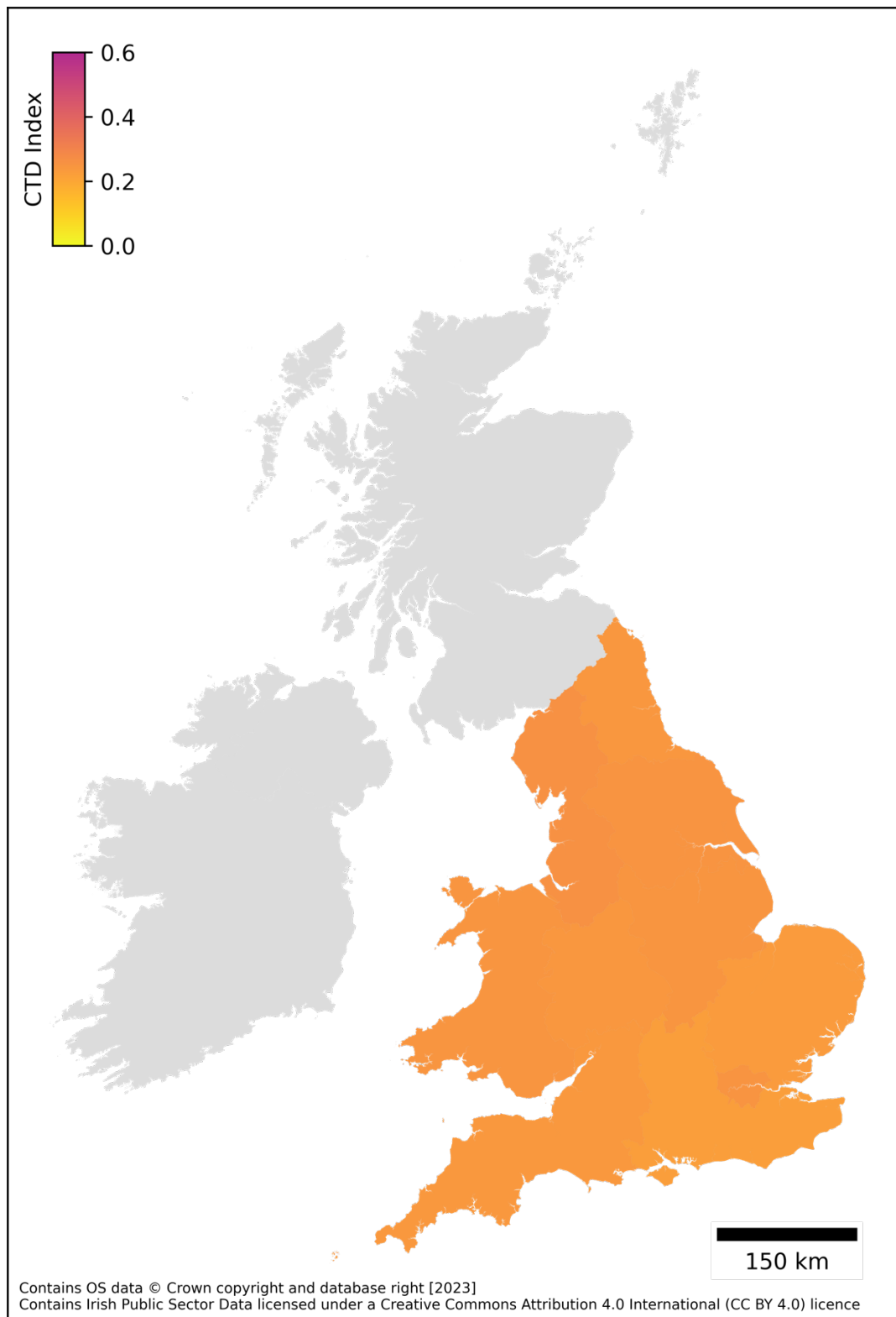


**Figure 5. Distribution of CTD homes in England and Wales: LSOA estimates**



Note: The scale shown denotes the CTD Index of a dwelling. This scale can be customised in line with specific analysis requirements.

**Figure 6. Distribution of CTD homes in England and Wales: regional estimates**



Note: The scale shown denotes the CTD Index of dwelling. This scale can be customised in line with specific analysis requirements.

## Recommendations for further development

The functionality of the framework, illustrated through the two use cases, has demonstrated its ability to:

- Employ a spectrum of technical, physical and occupant attributes to identify homes that are considered to be CTD from a whole house retrofit perspective.
- Undertake analysis at both the individual dwelling level and stock level and use individual dwelling national statistics identifiers (UPRNs) to improve analysis at higher resolution.
- Interact with diverse datasets that include data on a wide range of CTD attributes at various resolutions/scales and link them for the purposes of identification.

To support more robust analysis of the impact of future policies on these homes, recommendations are proposed. These address requirements for the further development of this framework and the datasets that may underpin future analyses targeting CTD homes (and the housing stock in general):

- The success of this framework has been underpinned through engagement with policy stakeholders to inform the iterative development processes and envision the use cases it may be used for. Further development of the framework targeting its use for industry and civic society stakeholders should be underpinned by similar engagement in-depth processes.
- EPC datasets, where available, were used to allow analysis at individual housing level. This level of analysis is invaluable for such applications as assessing individual dwelling/household need, suitability and eligibility for energy efficiency retrofit schemes. However, to improve the usability of these vast datasets, issues with data quality (addressed in this report through cleaning and filtering processes applied) should be systematically checked for and corrected at source.
- UPRNs have provided a useful method by which to link some, but not all, datasets that were used for the analysis. It is therefore recommended that UPRNs should be systematically used (where appropriate and in line with data protection requirements) in relevant future data collection exercises related to the UK housing stock in general and energy efficiency retrofit in particular.
- Guidelines for future data collection exercises should aim to improve consistency between datasets and minimise uncertainty. For example, the provision of a clear methodology for mapping data from previous LSOA 2011 definitions.

# Appendices

## Appendix 1: Dataset mapping

**Table 3. Mapping of relevant datasets, with dataset information and attribute mapping (y=included, n=not included, ?=unclear)**

Name	Description	Coverage	Resolution	Age	Typology	Heritage	Height	Location	Materials	Floor	Height/ age	Region	Tenure	Notes
Climate Risk	Spatial maps of climate exposure & vulnerability across Greater London. Flood & exposure risk data	Greater London	LSOA	n	n	n	n	y	n	n	?	n	n	None.
National Coastal Erosion Risk Map	Spatial data splitting the coastal baseline to 'frontages' (lengths of coast)	UK	TBC	n	n	n	n	y	n	n	?	n	n	None.
Solid wall & off gas network properties	Number & percentage of properties with solid walls/not connected to gas mains.	England	LSOA	n	n	n	n	n	n	n	?	y	n	None.

Name	Description	Coverage	Resolution	Age	Typology	Heritage	Height	Location	Materials	Floor	Height/ age	Region	Tenure	Notes
HadUK-Grid	Climate data for spatial grids	UK	1x1m grid	n	n	n	n	y	n	n	?	n	n	None.
Radar Rainfall Map Layer	Spatial data of recent rain rates in mm/hour over the UK	UK	TBC	n	n	n	n	y	n	n	?	n	n	None.
2011 rural/urban classification	A rural/urban view of datasets	UK	Output area (OA), super output area (SOA) & ward level.	n	n	n	n	y	n	n	?	y	n	None.
Continuous Household Survey	Household level characteristics & attitude data	Northern Ireland	Dwelling/ household level	n	y	n	n	n	n	n	?	n	y	None.

Name	Description	Coverage	Resolution	Age	Typology	Heritage	Height	Location	Materials	Floor	Height/ age	Region	Tenure	Notes
Scottish Household Survey	Household level characteristics & attitude data	Scotland	Dwelling/ household level	n	y	n	n	n	n	n	?	y	y	None.
EPC data England & Wales	Data from certificates lodged on the Energy Performance of Buildings Registers since 2008	England & Wales	Dwelling/ household level	y	y	n	n	y	y	y	?	y	y	Identificat ion of high rise through FLAT_ST OREY_C OUNT
EPC data Scotland	Data from certificates lodged on the Energy Performance of Buildings Registers since 2013	Scotland	Dwelling/ household level	y	y	n	n	y	y	y	?	y	y	Identificat ion of high rise through FLAT_ST OREY_C OUNT
Scottish House Condition Survey	Stock level data of homes in Scotland: dwelling & occupant socio-economic data	Scotland	Dwelling/ household level	y	y	n	n	n	y	y	?	y	y	CTD y/n identifier, FP status
English Housing Survey	Stock level data of homes & occupant socio-economic data in England	England	Dwelling/ household level	y	y	n	n	n	y	y	?	y	y	Nature of area/ Rurality indicator included

Defining and identifying complex-to-decarbonise homes and retrofit solutions: Annex B

Name	Description	Coverage	Resolution	Age	Typology	Heritage	Height	Location	Materials	Floor	Height/ age	Region	Tenure	Notes
Northern Irish Housing Condition Survey	Stock level data of homes in NI: dwelling & occupant socio-economic data	Northern Ireland	Dwelling/ household level	y	y	n	n	n	y	?	?	y	y	FP status, urban/ rural/ fringe only
Welsh Housing Condition Survey	Stock level data focusing on dwelling characteristics/condition	Wales	Dwelling/ household level	y	y	n	n	n	y	y	?	n	y	Urban/ rural/ fringe only
National Heritage List for England	Spatial Data on Heritage Buildings	England	TBC	?	?	y	?	y	?	?	?	y	?	None.
Heritage Local Authority Profiles	Datasets & information on the historic environment in England	England	LA	n	n	y	n	n	n	n	?	y	n	None.
Heritage Buildings	Heritage Buildings in the City	London	LA	n	n	?	?	y	?	?	?	y	?	None.
Discovery, identification & understanding	Scope & distribution of Heritage buildings & Conservation areas	England	LSOA/LA	n	n	y	n	n	n	n	n	y	n	None.
Welsh Index of Multiple Deprivation	Areas of deprivation in Wales	Wales	LSOA	n	n	n	n	n	n	n	n	y	n	None.

Defining and identifying complex-to-decarbonise homes and retrofit solutions: Annex B

Name	Description	Coverage	Resolution	Age	Typology	Heritage	Height	Location	Materials	Floor	Height/ age	Region	Tenure	Notes
Fuel poverty detailed tables	2022 data using the Low Income Low Energy Efficiency (LILEE) indicator.	England	High level statistical release	y	y	n	y	n	n	y	n	y	y	None.
Fuel poverty by Lower Super Output Area	2023 data using the Low Income Low Energy Efficiency (LILEE) indicator.	England	LSOA	n	n	n	n	n	n	n	n	y	n	None.
LSOA estimates of properties not connected to gas network	Estimates for the number of properties without mains gas.	UK	LSOA	n	n	n	n	n	n	n	n	y	n	None.
Experimental statistics on heat networks	Heat network data collected following Heat Metering & Billing Regulations	UK	LA	n	n	n	n	y	n	n	n	y	n	None.
OS Open UPRN	All UPNRs found in AddressBase	Great Britain	Dwelling level	n	n	n	n	y	n	n	n	y	n	None.
National Statistics UPRN Lookup	Allocates GB addresses to Output Area (OA) using UPNRs	Great Britain	Dwelling/ household level	n	n	n	n	y	n	n	n	y	n	None.
Valuation Office Data	Data underpinning the banding of properties for Council Tax	UK	Dwelling/ household level	y	y	n	y	?	?	y	?	y	?	Data restricted



## Appendix 2: Attribute identification, mapping, and weighting methodology

The following three stage process identified, mapped and weighted the attributes that would underly processes used by the framework to identify CTD homes.

Stage A: Categorisation of principal and proxy attributes<sup>11</sup>

1. **Keyword identification/synthesis:** Keywords describing CTD attributes were identified from REA (WP3) and interview (WP4) transcripts to form an initial attribute pool. This included keywords mentioned in the context of both CTD definition and challenges.
2. **Initial filtering:** The initial attribute pool was filtered as follows:
  - Only attributes related to the building itself and people directly associated with it (occupants/ owners/ neighbours) were included.
  - From these, attributes for which there was limited evidence from both REA and interviews were further excluded.
3. **Categorisation of attributes:** The resultant attribute pool was split into two categories:
  - **Principal attributes (Figure 4):** These directly determine/contribute to a dwelling designated as CTD, such as hard-to-treat insulation and space constraints. They are used either to assess an individual dwelling, or as part of the “backend” calculations to support stock level analysis.
  - **Proxy attributes (Table 5):** These indicate the existence of one or multiple principal attributes, such as age band and dwelling type. The mapping from proxy to principal attributes is primarily used by the framework to interface with housing stock databases, it can also inform assessment of individual dwellings as defaults when certain principal attributes are hard to determine.

**Mapping principal and proxy attributes:** The mapping for these attributes was derived via their co-occurrence in both REA and interviews (**Table 6**).

---

<sup>11</sup> This process is categorised as “low uncertainty / low customisation needed”, since evidence to support it was fairly abundant in the REA and interviews, therefore no or limited customisation will be needed to apply the framework to the specific use cases highlighted in the DESNZ stakeholder feedback

**Table 4. Principal attributes**

<b>Attribute</b>	<b>CTD Principal category</b>	<b>CTD principal subcategory</b>
Technical attributes (the building itself)	insulation	non-fillable cavity wall infeasible loft insulation
Technical attributes (the building itself)	space	internal wall insulation heating/energy storage low-carbon heating/energy
Technical attributes (the building itself)	heating source	off gas grid off heat network
Technical attributes (the building itself)	construction	external wall insulation poor maintenance history non-standard construction other additional work
Social attributes (people directly associated with it)	willingness	self and fellow occupants' resistance self and the landlord resistance self-resistance
Social attributes (people directly associated with it)	affordability	low affordability
Social attributes (people directly associated with it)	disruption	disruption to neighbours disruption to self

**Table 5. Proxy attributes**

Attribute	CTD proxy category	CTD proxy variations
Category based	age	pre-1919 1920 -- 1944 1945 -- 1964 1965 -- 1989 post-1990
Category based	typology	detached semi-detached terraced bungalow flat converted
Category based	heritage	non-heritage conservation area grade i grade ii* grade ii
Category based	floor area	floor area < 16 floor area >= 16
Category based	height	height < 6 height >= 6
Category based	tenure	owner-occupied private rented social rented
Binary/Numerical	mixed tenure	Y/N
Binary/Numerical	rainfall	0~1
Binary/Numerical	remote	Y/N
Binary/Numerical	gas grid	0~1
Binary/Numerical	heat network	0~1
Binary/Numerical	deprivation	Y/N
Binary/Numerical	fuel poverty	0~1

**Table 6. Mapping from proxy to principal attributes**

Proxy attribute	Principal attribute
age	non-fillable cavity wall external wall insulation poor maintenance history non-standard construction other additional work
typology	infeasible loft insulation internal wall insulation heating/energy storage low-carbon heating/energy external wall insulation non-standard construction self and fellow occupants' resistance self and the landlord resistance self-resistance low affordability disruption to neighbours
heritage	non-fillable cavity wall infeasible loft insulation low-carbon heating/energy external wall insulation disruption to self
floor area	internal wall insulation heating/energy storage
height	non-fillable cavity wall off gas grid external wall insulation
tenure	poor maintenance history self and fellow occupants' resistance self and the landlord resistance self-resistance low affordability
mixed tenure	heating/energy storage self and fellow occupants' resistance self and the landlord resistance self-resistance low affordability
rainfall	external wall insulation

Proxy attribute	Principal attribute
remote	non-fillable cavity wall non-standard construction self-resistance
gas grid	off gas grid
heat network	off heat network
deprivation	low affordability
fuel poverty	low affordability

## Stage B: Assignment of attribute weights<sup>12</sup>

Weight assignment followed a simple evidence-based ranking scheme. The following process was implemented to then assign weights based on “importance” to attribute identified in Stage A. Two types of weights were used:

- Weights for aggregation of principal attributes:** These were derived from a combination of evidence from interview and analysis team judgement. Interview evidence consisted of “number of mentions” by interviewees, however as this is subject to such aspects as interviewee project experience and personal interest, consistency/correlation<sup>13</sup> between number of mentions and relative importance of that attribute cannot be strongly established. Hence, equal weights were used in the following analysis, and they should be customised as per preference of specific stakeholders in future practice. Any positive number can be used for this set of weights to reflect relative importance between principal attributes, they are normalised automatically in the framework.
- Weights for conversion from proxy to principal attributes:** These were derived from a combination of evidence from REA and the qualitative interviews combined with analysis team judgement. Evidence from REA is presented as the degree of difficulty in retrofit that a proxy attribute leads to a principal one, which is consistent. Evidence from interviews entails previously described limitations and was considered with expert judgement employed where necessary. Using this evidence, a point-based system was adopted, where values from 0 to 3 were used to indicate the correlation between proxy and principal attributes (0 = no correlation, 1 = low, 2 = medium, and 3= high degree of correlation).

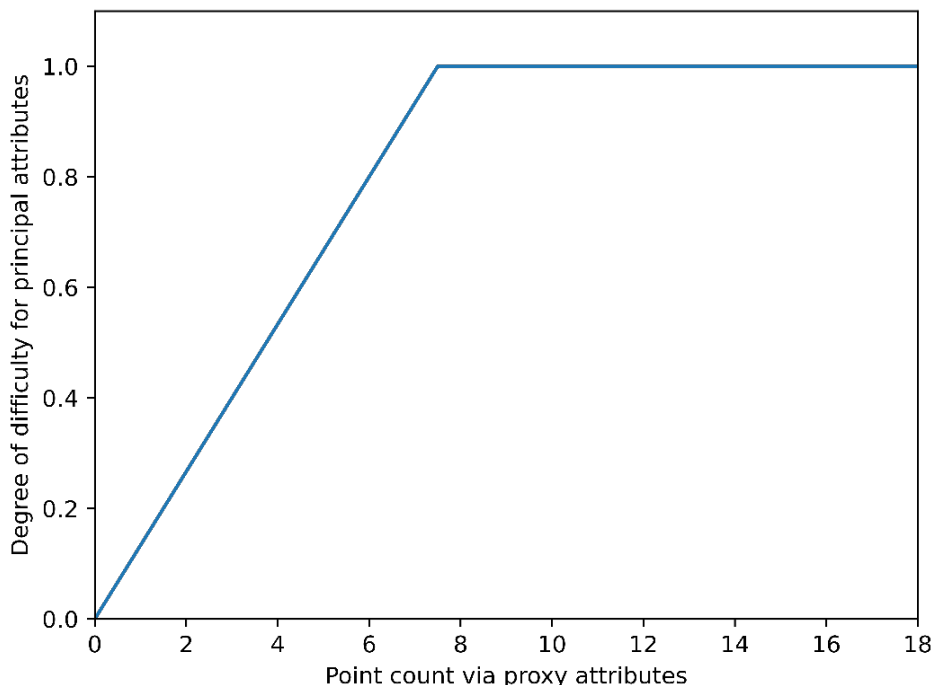
The degree of the retrofit difficulty via proxy is the normalised summation of the points achieved for all the attributes. To accommodate the varied number of related proxy attributes for each principal one, the summed points are normalised linearly and capped at 7.5 (Figure

<sup>12</sup> This process is categorised as “high uncertainty / customisation needed” since evidence to support from the REA and interviews is limited and/or highly context-dependent. Thus, a degree of customisation will apply to the framework to the specific use cases highlighted in the DESNZ stakeholder feedback

<sup>13</sup> Correlation here denotes a connection/relationship or connection between attributes.

7). That is, when a given principal attribute is linked to two and half proxy attributes in the strongest correlation (3), it is considered to have the highest degree of difficulty.

**Figure 7. The cap system for proxy conversion**



**Table 7. Rationale for score assignment**

Attribute	Rationale
Age	Dwelling age was considered to be a proxy that refers to a number of CTD attributes such as the absence of fillable wall cavities, using non-standard construction, infeasibility of external wall insulation (EWI), maintenance history, and the need for additional work to be undertaken to support retrofit <sup>14, 15, 16</sup> High correlation scores were assigned to older buildings.
Archetype	Archetypes that typically are associated with space constraints such as terraced properties, followed by flats, tenements, and semi-detached houses were correlated (with different scores) to challenges in e.g., accommodating internal wall insulation, space-demanding measures such as heating storage and low carbon heating systems. Infeasible external wall insulation was correlated to flats and tenements archetypes due to their heights/scaffolding considerations. Scores for neighbours' resistance to the retrofit uptake and for disruption to neighbours were correlated to the archetype attachment type (e.g., flats, tenements and terraced units were considered more disruptive). The scores assigned to the different archetypes/retrofit measures was found to be in line with the actual retrofit measure costs aggregated in the Home Upgrade Grant reports.

<sup>14</sup> Piddington, J., Nicol, S., Garrett, H., & Custard, M. (2020). The Housing Stock of The United Kingdom. Building Research Establishment. [https://files.bregroup.com/bretrust/The-Housing-Stock-of-the-United-Kingdom\\_Report\\_BRE-Trust.pdf](https://files.bregroup.com/bretrust/The-Housing-Stock-of-the-United-Kingdom_Report_BRE-Trust.pdf)

<sup>15</sup> Raslan, R., Schwartz, Y., Palmer, J., & Terry, N. (2017). Understanding Best Practice in Deploying External Solid-Wall Insulation in the UK. BEIS.

<sup>16</sup> Thompson, R. (2012). Review of the Number of Cavity Walls in Great Britain- Methodology.

Attribute	Rationale
Heritage value/ significance/ designation	In general, the degree of historic significance was correlated to some CTD attributes such as infeasible cavity insulation, infeasible loft insulation, infeasible low carbon heating systems, infeasible EWI, additional work requirements, and disruption to self and neighbours. A distinction between conservation areas and the different degrees of heritage value/significance was considered given that the degree of retrofit infeasibility is expected to be the highest for Grade I listed properties, Grade II*, Grade II, and then conservation areas.
Floor area	Dwelling floor areas below a certain threshold constrain the installation of space-demanding retrofit measures. Hence, infeasible internal wall insulation (IWI) and infeasible heating/energy storage were considered to be directly correlated.
Height	Tall buildings were considered to have medium correlation with infeasible EWI due to scaffolding requirements. In the UK, most tall buildings have some degree of correlation to the absence fillable wall cavities and loft spaces that can be insulated. Moreover, most tall buildings might already be off the gas grid as they would typically be using another medium for heating such as a communal or district heating system.
Occupancy type	Different correlation scores have been assigned to the resistance to the retrofit uptake based on occupancy type. Rented properties were considered more prohibitive than mixed-use and owner occupied. Lack of affordability, maintenance history attributes had similar correlations to the occupancy type. Disruption scores were slightly different to reflect how mixed-use buildings might be more disruptive to neighbours.
Rainfall	Based on interview feedback, rainfall was highlighted as challenge to EWI installation. Hence, rainfall frequency per year above a certain threshold can be a proxy of medium correlation to the infeasibility of EWI.
Remote areas	Remote areas can be a proxy for non-standard construction, an increasing degree of retrofit resistance due to the difficulty of getting material and labour. A low degree of correlation to non-standard construction and the need for additional work has been assigned to remote areas as well.
Proximity to gas grid	A threshold distance is being studied, above which the properties are to be considered off the gas grid.
Proximity to heat network	A threshold distance is being studied, above which the properties are to be considered off the heat network.
Fuel poverty	Informed by DESNZ annual statistical reports, deprivation and fuel poverty can be proxies for retrofit affordability and the resistance to the retrofit uptake.

## Stage C: Scripting<sup>17</sup>

The processes and data used to implement the framework was then scripted in Python in a file containing the commands required to execute the analysis. Four types of scripts were developed, corresponding to the four functions in Module 1 and Module 2 as depicted in Figure 2:

- Individual CTD identifier: [one script] this can be used standalone to assess individual dwellings.
- Attribute mapper: [one script] this maps proxy to principal attributes and communicates with Individual CTD identifier.
- Dataset harmonisers: [multiple scripts, one for each dataset] these translate attributes in specific datasets to proxy attributes in Table 6. Further information on these is provided in Box 2 below.
- Batch manager: [one script] this coordinates the input of different datasets via Dataset harmonisers and feeds the translated attributes of multiple records into Individual CTD identifier via Attribute mapper for stock level identification.

**Table 8. Creation of data harmonisers**

Stage	Description
Creation of dataset harmonisers	A dataset harmoniser is essentially a python script that contains a python function or a series of python functions to convert attributes used in a specific dataset to proxy attributes defined in the framework. The core framework modules utilise pandas <sup>18</sup> DataFrame for data communication, as such, the output of such harmoniser function is expected to be pandas DataFrame objects.
Input	Two major input arguments are expected: An object of the dataset read into python. A pandas DataFrame object for the collection of proxy attributes, which contains coordinates and postcodes for each UPRN that facilitates the mapping of datasets. Any other input arguments that are necessary to harmonise specific datasets can be included.
Output	The output argument is a pandas Series that contains the harmonised dataset to be concatenated to the main proxy DataFrame.

<sup>17</sup> This process is categorised as “medium uncertainty / customisation needed”, since the script components regarding attribute conversion and aggregation is dataset-agnostic, yet those regarding interfacing specific datasets are subject to adaptation, especially when involving internal datasets. Thus, a degree of customisation will be needed to apply the framework to the specific use cases highlighted in the DESNZ stakeholder feedback

<sup>18</sup> An established third-party python library that handles tabular data and facilitates their analysis (<https://pandas.pydata.org>). A pandas DataFrame is a matrix of data, and a Series is a vector.



## Appendix 3: Detailed framework components

**Table 9. Framework components**

Component	Description
Principal CTD Attributes	These attributes - identified from literature review and interviews - directly determine/contribute to home being considered CTD. These should not be changed by framework end-users but may be customised by developers/maintainers should more evidence emerge in the future. Each of them needs to be assigned with a value between 0 and 1.
Proxy CTD Attributes	These attributes indicate the principal ones. These should not be changed by framework end-users but may be customised by developers should more evidence emerge in the future. There are potentially three value types: Boolean (yes/no), numeric (0-1), categorical (string).
Individual CTD Result	This is the final identification result on the individual dwelling level, which determines if it is CTD via a number ranging from 0 to 1.
Interface 1	This is the interface for individual dwelling assessment. Framework end-users are required to fill out degree of difficulty values for each principal attributes. Framework maintainers need to define aggregation weights for principal attributes.
Interface 2	This is the interface for stock level analysis, where framework end-users are required to input the datasets of interest, whilst framework maintainers need to define conversion weights for proxy attributes. Both end-users and maintainers can contribute to constructing datasets attribute maps.
Module 1	This hosts the core function that determines whether an individual dwelling is CTD. The algorithm essentially performs a weighted average of the degree of difficulty values for each principal attributes.
Module 2	This hosts three functions that facilitates the CTD identification at stock level: the dataset harmonisers, the attribute mapper and the batch manager. The dataset harmonisers are pre- or user-defined; the attribute mapper translates the harmonised proxy attributes into principal ones in the framework; the batch manager communicates with Module 1 to distribute the inputs and collect the outputs in batch.

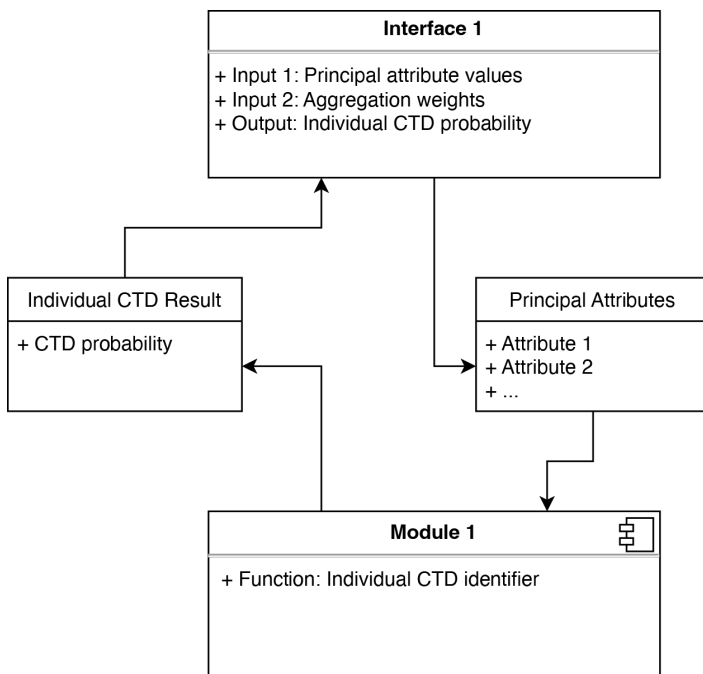
## Appendix 4: Framework workflows

The following appendix describes the automated workflows and interaction between modules for the illustrative use cases for the framework.

### Individual dwelling assessment

For an individual dwelling level assessment, the active components of the framework are Interface 1, Module 1, Principal CTD Attributes and Individual CTD Result (Figure 4). The user manually inputs data entries gathered via measurement, surveying or other viable means that correspond to the principal attributes in Interface 1. Once entered, the data is fed into the identification algorithm in Module 1, which outputs the identification outcome through a series of automated processes and returned to Interface 1.

**Figure 8. Workflow for individual dwelling assessment**



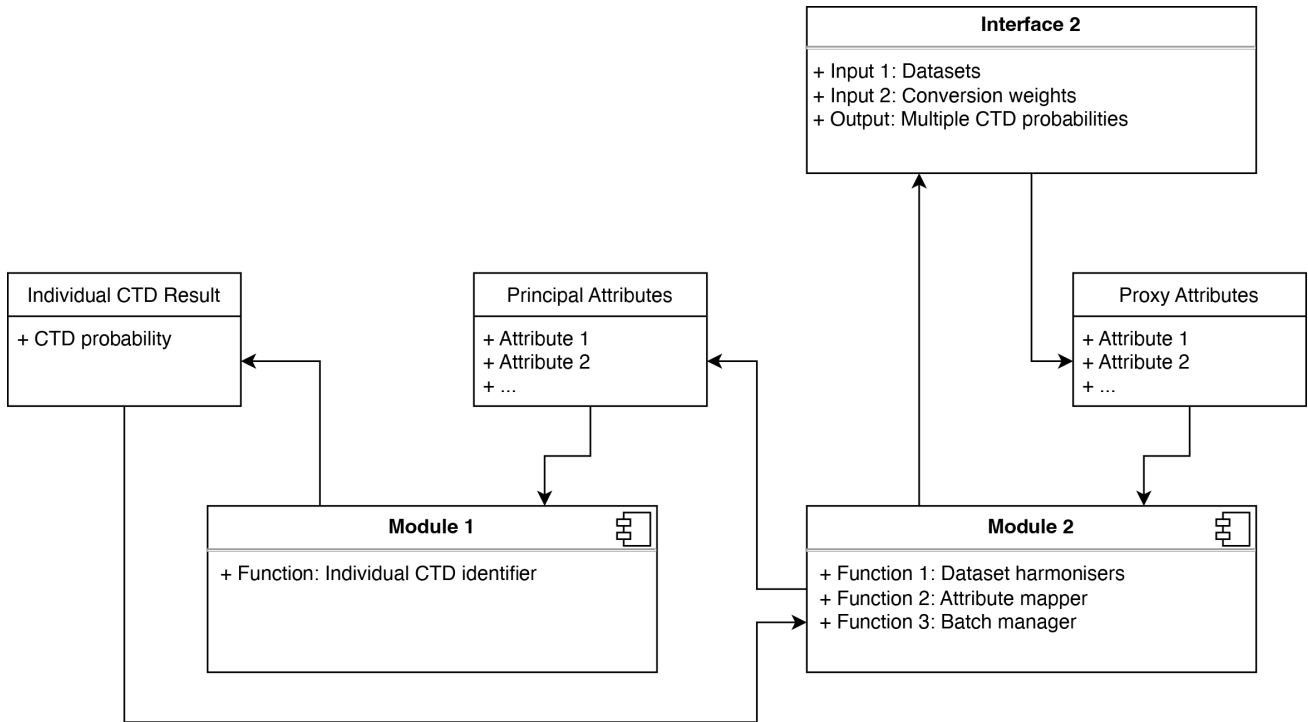
### Housing stock analysis

The stock level analysis can extend the analysis to the UK as a whole or to subset of the UK stock (e.g., devolved administration level, city level, local authority level, housing provider level). For this analysis level, the active components of the framework are Interface 2, Module 1, Module 2, Primary CTD Attributes, Secondary CTD Attributes and Individual CTD Result (Figure 5).

For this application, the user specifies the housing stock level datasets to be used according to the analysis aims. The user can use pre-defined dataset attribute maps, or define their own, to translate dataset attributes to proxy ones, which will be later mapped to principal attributes required by the algorithm in Module 1.

After receiving these inputs, Module 2 will loop over all the dwelling entries in the dataset, map the attributes and send them to Module 1 for individual identification. The collated analysis outcomes are then returned to Interface 2 for further analysis.

**Figure 9. Workflow for housing stock analysis**



## Appendix 5: Application of the framework: methodology for data harmonisation

The following details the approach for applying the framework for a stock level analysis. The approach is based on the ‘layering’ of multiple datasets, each of which includes key data regarding CTD attributes. These datasets are linked through the Unique Property Reference (UPRN), which is used as the unique identifier across datasets<sup>19</sup>.

Similar to the individual dwelling assessment, the core resolution of the stock analysis is on the UPRN level. In the instance that a given dataset has a lower geographic or non-geographic resolution, it is mapped to individual households using the UPRN Lookup table, as a shared characteristics in a certain area.

### Datasets used

The following table lists the dataset ‘layered’ and linked for the analysis detailing the usage of each and/or the attributes that are derived/mapped from it. These datasets represent a subset from those listed in Appendix 1, selected in line with the analysis scope required for the illustrative cases (England and Wales) and the proxies they cover.

**Table 10. Datasets used for illustrative use cases**

Dataset	Link	Usage
National Statistics UPRN Lookup	<a href="https://geoportal.statistics.gov.uk/datasets/national-statistics-uprn-lookup-april-2023">https://geoportal.statistics.gov.uk/datasets/national-statistics-uprn-lookup-april-2023</a>	Base for UPRN records with associated geographic information
OS Open UPRN	<a href="https://osdatahub.os.uk/downloads/open/OpenUPRN">https://osdatahub.os.uk/downloads/open/OpenUPRN</a>	Cross-checking for UPRN validity
Energy Performance of Buildings Data: England & Wales	<a href="https://epc.opendatacommunities.org">https://epc.opendatacommunities.org</a>	Base for stock analysis, providing proxies: age, typology, floor area, height, tenure, mixed tenure
National Heritage List for England (NHLE)	<a href="https://opendatahistoricalengland.hub.arcgis.com/datasets/historical-england::national-heritage-list-for-england-nhle">https://opendatahistoricalengland.hub.arcgis.com/datasets/historical-england::national-heritage-list-for-england-nhle</a>	Providing the proxy: heritage
Conservation Areas	<a href="https://opendatahistoricalengland.hub.arcgis.com/datasets/historical-england::conservation-areas">https://opendatahistoricalengland.hub.arcgis.com/datasets/historical-england::conservation-areas</a>	Providing the proxy: heritage
HadUK-Grid Gridded Climate Observations	<a href="https://catalogue.ceda.ac.uk/uuid/4dc8450d889a491ebb20e724debe2dfb">https://catalogue.ceda.ac.uk/uuid/4dc8450d889a491ebb20e724debe2dfb</a>	Providing the proxy: rainfall

<sup>19</sup> In this section, the term UPRN is used interchangeably with the individual property and its record.

Dataset	Link	Usage
Rural Urban Classification (2011) of Lower Layer Super Output Areas in England & Wales	<a href="https://geoportal.statistics.gov.uk/datasets/ons::rural-urban-classification-2011-of-lower-layer-super-output-areas-in-england-&amp;-wales-1">https://geoportal.statistics.gov.uk/datasets/ons::rural-urban-classification-2011-of-lower-layer-super-output-areas-in-england-&amp;-wales-1</a>	Providing the proxy: remote
LSOA estimates of properties not connected to the gas network	<a href="https://www.gov.uk/government/statistics/lsqa-estimates-of-households-not-connected-to-the-gas-network">https://www.gov.uk/government/statistics/lsqa-estimates-of-households-not-connected-to-the-gas-network</a>	Providing the proxy: gas grid
Experimental statistics on heat networks	<a href="https://www.gov.uk/government/publications/energy-trends-march-2018-special-feature-article-experimental-statistics-on-heat-networks">https://www.gov.uk/government/publications/energy-trends-march-2018-special-feature-article-experimental-statistics-on-heat-networks</a>	Providing the proxy: heat network
Index of Multiple Deprivation (Dec 2019) Lookup in England	<a href="https://geoportal.statistics.gov.uk/datasets/ons::index-of-multiple-deprivation-dec-2019-lookup-in-england">https://geoportal.statistics.gov.uk/datasets/ons::index-of-multiple-deprivation-dec-2019-lookup-in-england</a>	Providing the proxy: deprivation
Index of Multiple Deprivation (Dec 2019) Lookup in Wales	<a href="https://geoportal.statistics.gov.uk/datasets/ons::index-of-multiple-deprivation-dec-2019-lookup-in-wales">https://geoportal.statistics.gov.uk/datasets/ons::index-of-multiple-deprivation-dec-2019-lookup-in-wales</a>	Providing the proxy: deprivation
Sub-regional fuel poverty data 2023 (2021 data) - England	<a href="https://www.gov.uk/government/statistics/sub-regional-fuel-poverty-data-2023-2021-data">https://www.gov.uk/government/statistics/sub-regional-fuel-poverty-data-2023-2021-data</a>	Providing the proxy: fuel poverty
Estimated Fuel Poverty (by LSOA) - Wales	<a href="https://datamap.gov.wales/layers/inspirewg:fuel_poverty_by_lsoa_geom_vw">https://datamap.gov.wales/layers/inspirewg:fuel_poverty_by_lsoa_geom_vw</a>	Providing the proxy: fuel poverty
LSOA (2011) to LSOA (2021) to Local Authority District (2022) Lookup for England & Wales (V2)*	<a href="https://geoportal.statistics.gov.uk/datasets/ons::lsqa-2011-to-lsqa-2021-to-local-authority-district-2022-lookup-for-england-&amp;-wales-version-2">https://geoportal.statistics.gov.uk/datasets/ons::lsqa-2011-to-lsqa-2021-to-local-authority-district-2022-lookup-for-england-&amp;-wales-version-2</a>	Cross-mapping between LSOA (2011) to LSOA (2021) for datasets that are yet to be updated to LSOA (2021)
Local Authority District (2011) to Local Authority District (2021) Lookup for England & Wales*	<a href="https://geoportal.statistics.gov.uk/datasets/ons::local-authority-district-2011-to-local-authority-district-2021-lookup-for-england-&amp;-wales">https://geoportal.statistics.gov.uk/datasets/ons::local-authority-district-2011-to-local-authority-district-2021-lookup-for-england-&amp;-wales</a>	Cross-mapping between LAD (2011) to LAD (2021) for datasets that are yet to be updated to LAD (2021)

\*Denotes lookup datasets used to cross map datasets available at different levels

## Data compilation

Since datasets obtained from various sources formats, a compilation exercise is needed for each dataset to unify their format for later use. In general, each type of dataset was compiled into a single file, with non-spatial data in \*.csv format and spatial data in \*.geojson format. During compilation, data columns/attributes which were found to be irrelevant to proxies that were identified for CTD homes were removed.

## Data filtration

Following compilation, a subset of compiled datasets were filtered to remove invalid records (e.g. duplicate records or data for homes that did not exist), and to correct errors in the remaining records to ensure that they were framework-friendly (i.e. easily read and processed by the framework). The filtration and corrections made to the data records are detailed in **Table 11**.

**Table 11. Data filtration and correction processes for datasets used in illustrative use cases**

Dataset	Compilation	Filtration/Correction
National Statistics UPRN Lookup	Accessible via a set of csv files per region, they were combined into a single csv file.	Remove UPRNs not in OS Open UPRN Remove UPRNs with a difference in coordinates (> 0.5) from OS Open UPRN
OS Open UPRN	Accessible via a single csv file.	N/A
Energy Performance of Buildings Data: England & Wales	Accessible via a set of csv files per local authority, they were combined into a single csv file.	Remove UPRNs not address matched Remove duplicate UPRNs without a latest INSPECTION_DATE merge Maisonette into Flat, as they were treated similarly in SAP Change Flat with a BUILT_FORM to Converted Remove House with no BUILT_FORM Remove Bungalow & Park home that are neither Detached nor Semi-Detached Remove UPRNs with missing TENURE CONSTRUCTION_AGE_BAND NUMBER_HABITABLE_ROOMS TOTAL_FLOOR_AREA Unify values for TENURE CONSTRUCTION_AGE_BAND Remove UPRNs on the same POSTCODE with conflicting PROPERTY_TYPE

Dataset	Compilation	Filtration/Correction
		Remove Flat single on a given POSTCODE & with no FLAT_STOREY_COUNT Use the highest FLAT_STOREY_COUNT for multiple Flat UPRNs on the same POSTCODE Remove duplicate UPRNs with conflicting TENURE CONSTRUCTION_AGE_BAND TOTAL_FLOOR_AREA / NUMBER_HABITABLE_ROOMS BUILT_FORM PROPERTY_TYPE Merge remaining duplicate UPRNs via mode
National Heritage List for England	Accessible via a single csv file.	Separate records with multiple coordinates Remove duplicates
Conservation Areas	Accessible via a single csv file.	Remove duplicates
HadUK-Grid Gridded Climate Observations	12km daily rainfall in 2021 used, accessible via a set of nc files/ month, frequency of rain heavier than drizzle (>0.5) calculated across the year & geographic points converted to polygons (12x12) before writing into single geojson file.	N/A
Rural Urban Classification (2011) of Lower Layer Super Output Areas in England & Wales	Accessible via a single csv file, it was mapped LSOA (2021).	N/A

**Table 11. Continued data filtration and correction processes for datasets used in illustrative use cases**

Dataset	Compilation	Filtration/Correction
LSOA estimates of properties not connected to the gas network	Accessible via a single xlsx file.	N/A
Experimental statistics on heat networks	Accessible via a single xlsx file.	N/A
Index of Multiple Deprivation (Dec 2019) Lookup in England	Accessible via a single csv file, their native attribute names were unified. A threshold of 10% top deprivation area was applied, before combined into a single csv file	N/A
Index of Multiple Deprivation (Dec 2019) Lookup in Wales	Accessible via a single csv file, their native attribute names were unified. A threshold of 10% top deprivation area was applied, before combined into a single csv file	N/A
Sub-regional fuel poverty data 2023 (2021 data) - England	Accessible via a single csv file (2021).	N/A
Estimated Fuel Poverty (by LSOA) - Wales	Accessible via a single csv file, mapped at LSOA level (2021).	N/A
LSOA (2011) to LSOA (2021) to Local Authority District (2022) Lookup for England & Wales (Version 2)	Accessible via a single csv file.	N/A



Dataset	Compilation	Filtration/Correction
Local Authority District (2011) to Local Authority District (2021) Lookup for England & Wales	Accessible via a single csv file.	N/A

---

This publication is available from:

[www.gov.uk/government/publications/defining-and-identifying-complex-to-decarbonise-homes](http://www.gov.uk/government/publications/defining-and-identifying-complex-to-decarbonise-homes)

If you need a version of this document in a more accessible format, please email [alt.formats@energysecurity.gov.uk](mailto:alt.formats@energysecurity.gov.uk). Please tell us what format you need. It will help us if you say what assistive technology you use.