INTERNATIONAL BUILDING PERFORMANCE SIMULATION ASSOCIATION ENGLAND

Proceedings of BSO-V 2020
5th Building Simulation and Optimization Virtual Conference
21-22 September 2020
Loughborough, UK

Towards reproducible dwelling retrofit evaluation: Data, tools & opportunity to address uncertainty

Kate Simpson[1,2]
[1]School of Design Engineering, Faculty of Engineering, Imperial College London, SW7 2AZ
[2]The Alan Turing Institute, British Library, 96 Euston Road, London, NW1 2DB, United Kingdom

Abstract

Reproducible research allows scientific work to be trusted. It enables data to be shared and methods developed. Retrofit evaluation relies on mixed method data collection, processing and analysis techniques. This includes building performance simulation and indoor environment evaluation. However, few studies to date have successfully combined and shared these data. Reproducible retrofit evaluation is an emerging topic.

In building simulation, there is a recognised gap between modelled and measured energy use. In previous research, the most sensitive parameters influencing domestic space heating energy demand were found to be temperature set points and heating duration. These are aleatory uncertainties and almost impossible to eliminate, but more data could increase knowledge.

Government, organisations and researchers do share data openly, where ethical and possible, or required. However, not all data in academic research is designed to be open. Where ethical principles are followed, opportunity exists to progress this space.

This paper aims to highlight the opportunity for space heating to be better characterised with more open data, particularly on indoor air temperature. A synthesis of recent work on open and reproducible research in relation to retrofit evaluation is presented. A reproducible guide is highlighted. This is followed by a summary of open datasets that inform retrofit decision-making. A gap is identified for more measured indoor air temperature. A selection of open-source tools and suitability for retrofit evaluation are discussed.

The novel contribution is the identification of opportunities for reproducible retrofit research, using open data to address aleatory uncertainty in terms of heating behaviour, derived from air temperature data. In addition, ideas towards data repositories are made.

Introduction

Dwelling retrofit can assist the UK to develop a more efficient energy system, as set out in the Clean Growth Strategy (HM Gov, 2017). Retrofit can assist alleviate fuel poverty, increase comfort and contribute to climate resilient infrastructure. The practice of retrofit could aid industry recovery from the COVID-19 crisis (Hurst, 2020), if decisions are made at pace.

The declaration of climate emergency across the UK and beyond may require energy researchers to work quicker and more collaboratively (Oreszczyn et al., 2020). An example of collaborative working is The Turing Way Guide to reproducible data science (TTWC, 2019). This is an open book project, co-written from the start.

Reproducible research, or open science approaches can enable trust in results (TTWC, 2019; Kraker et al, 2012). In relation to building performance studies, including retrofit analysis, this could lead to accessible shared datasets, data collection and analysis methods that are open to all researchers (Firth et al. 2018), as a starting point in the cultural shift required.

Retrofit, and energy efficiency research influences public welfare and the likelihood of direct policy recommendations can be higher than other research disciplines (Huebner et al., 2017), highlighting the need for transparency. Contextual factors in energy research will lead to variations in findings (Huebner and Fell., 2020). This is especially the case for individual dwelling evaluation studies.

The data processes in the building energy simulation community could be improved with further consideration to time to synchronise the project with objectives, careful management of people and their skillset, technology and the managing of interfaces between each, and quality including consistent metrics (Higginson et al., 2018). Data protocols to establish the collecting and sharing of data could help avoid a sector replicability crisis (Higginson et al., 2018).

Taking retrofit and building performance as a first example. While computational models serve to estimate, there are often large energy performance gaps between modelled energy estimates and measured energy data in buildings (Zero Carbon Hub, 2014). This can be due to multiple reasons including the physical construction, design factors but also a lack of context specific data for model inputs.

Individual dwellings vary in terms of construction material and thermal performance, quality of build, age of structure, geographic factors such as climate and strata and occupancy behaviour. In space heating modelling, the most sensitive parameters influencing space heating demand have been found to be heating behaviour factors such as heating set point and heating duration (Firth et al. 2009; Hughes et al., 2013). These are aleatory uncertainties and almost impossible to reduce (Booth., 2013) but the collection and sharing of more measured data, with consistent metrics can help increase knowledge and address uncertainty.

Therefore, this article presents a review on literature and data in relation to uncertainty in building simulation, aleatory uncertainty and indoor air temperature data, a selection of existing open data and tools and highlights opportunities for building simulation and energy performance researchers. This starts with uncertainty, then simulation approaches for existing dwellings, temperature studies, open data and related factors before discussing The Turing Way Guide. The article then discusses potential future work and concludes.

Uncertainty in existing building simulation

A number of variables influencing building thermal performance are inherently uncertain, such as occupant behaviour, weather conditions and thermal properties of the building envelope (Tian et al., 2018). Energy-efficiency is therefore not just about technologies but also the human behaviour, with a stochastic nature (Hong et al., 2015). With a focus on occupant behaviour, in its simplest form this relates to adjusting thermostat settings, opening and closing windows, switching lights, adjusting blinds and controlling ventilation settings and movement between spaces (Hong et al., 2016). Space heating energy demand, the largest portion of domestic heating demand, is most sensitive to heating set point and duration (Firth, 2009; Hughes, 2013), both of which are typically driven by occupant behaviour.

Uncertainty analysis can be divided into two categories: forward and inverse uncertainty quantification (Tian et al., 2018). With forward propagation being most widely used to date and focusing on quantifying the uncertainty in system outputs propagated via uncertain input variables using mathematical models (ibid). Conversely, inverse uncertainty analysis, which can also be considered as model calibration, focuses on quantifying unknown input variables through building energy models after collecting data from buildings (ibid).

To further explore uncertainty, a distinction is typically made between stochastic uncertainties, such as occupant behaviour and uncertainty due to a lack of knowledge; named aleatory uncertainty and epistemic uncertainty respectively (Tian et al., 2018). As aleatory uncertainty arises from processes that are inherently random, such as space heating behaviour (Booth, 2013). This can be addressed using a probabilistic framework (Tian et al., 2018), or using more measured data to lump parameters (Booth, 2013), but is almost impossible to eliminate.

Building simulation software are most suitable for predicting energy use of new-buildings at design stage, so that building properties and system parameters can be assumed with engineering design specifications (Heo et al., 2012). However, existing buildings come with variation with how the buildings and their components are actually operated, so the energy savings from simulation models of existing buildings require calibration (ibid). As such, monitored energy consumption becomes integral to the modelling process to generate a baseline model which closely matches monitored energy consumption of a building, so that designed retrofit solutions can be evaluated against a calibrated simulation (ibid). One approach that has been demonstrated previously in learning from samples of monitored data is a Bayesian approach (ibid, Booth, 2013). Example of this have been in identifying cost-benefits of macro level retrofit programmes using a steady-state models (Heo et al., 2012; Booth, 2013). The prior data informing similar Bayesian approaches could potentially be developed if based on more observed, monitored data.

Simulating existing dwellings

The computational modelling analysis tools for dwellings can be broadly divided into top-down, statistical bottom-up and engineering based (which tend to be bottom-up) (Booth, Choudhary and Spiegelhalter, 2012). Top-down models are often used to make decisions about a whole stock of houses (Swan and Ugursal, 2009). Engineering-based approaches are often used for one building allowing much more granular data on individual building elements relevant to that unit to be applied.

Engineering-based, also referred to as physics-based models, are predominantly based on the Building Research Establishment's Domestic Energy Model (BREDEM) in the UK (BRE Group, 2020). The current Standard Assessment Procedure, based on the BREDEM methodology, is the National Calculation Methodology used to generate UK Energy Performance Certificates (EPCs). This is consistent with standard BS EN ISO 13790 and the European Energy Performance Buildings Directive (ibid). The BREDEM methodology was developed in the 1980s and has undergone several revisions. This provides the calculation methodology for many bottom-up building energy models used within academic research, practice and policy. The technical documentation is open access enabling individual parameter calculations to be adjusted according to use.

BREDEM is a monthly-based steady-state tool used to estimate annual energy consumption and carbon emissions. The calculation takes into account the following parameters: building fabric materials, insulation, air leakage, efficiency and control of the heating system(s), solar gains, fuel source, space cooling (if applicable) and potential renewable energy technologies. It is independent of household size and composition, ownership and efficiency of specific electrical appliances and individual heating patterns and temperatures (BRE, 2014). This enables comparison between dwellings.

One of the major weaknesses of engineering-based models, such as BREDEM, is in the generalization of occupant behavioural patterns (Cheng and Steemers, 2011). BREDEM assumes a temperature of 21◦C in the main living area and 18◦C elsewhere, with heating on for

9 hours per day on weekdays and 16 hours at weekends. The lack of empirical data about actual heating patterns has led to models using such standard assumptions (Kavgic et al., 2010). While this enables comparison between model outcomes, it does not lead to accurate representations of energy demand per household.

Hughes et al. (2013) and Firth et al (2009) identified the characteristics of the heating systems and heat loss of the dwellings as the most influential factors of dwelling CO_2 emissions. Both Hughes et al. (2013) and Firth et al. (2009) found the heating demand temperature had the greatest influence with a normalised sensitivity co-efficient of 1.54 and 1.55 respectively. This means for every 1% increase in temperature, a 1.55% increase CO_2 emissions were predicted. Similarly, Cheng and Steemers (2011) found that a 7% increase in average dwelling energy consumption results from an 0.8ºC rise in internal temperature.

A model was developed by Cheng and Steemers (2011) which includes a feature to account for occupancy patterns, based on household employment status. They found that a variation of ±2.5oC in internal temperature introduces an uncertainty of ± 23% in the estimation of dwelling CO_2 emissions. A recommendation was made that a better understanding of internal temperatures and occupant behaviour will help to improve the performance of the model.

Temperature studies

Existing knowledge of preferences related to thermostat set-point and internal temperature is limited (Cheng and Steemers, 2011). Previously, 427 households in England were studied and an average set point of 21.1ºC was found and a standard variation of 2.5ºC (Shipworth, 2011). Patterns of heating have been derived from 249 dwellings in Leicester, via measured hourly temperatures, as opposed to set-points, and face to face socio-technical surveys (Kane et al., 2015). Of the 93% that were centrally heated, 51% heated for two periods a day and 33% just one period a day (ibid), similar results were reported by Huebner et al. (2015) on a similar sample. Furthermore, mean winter indoor temperatures in the rooms studied varied from 9.7ºC to 25.7ºC, all of which varies from the BREDEM assumptions.

Therefore, the BREDEM-based approach to UK home energy use could misrepresent energy use prediction, or the targeting of energy efficiency retrofit measures (Kane et al., 2015). There is now increasing interest in operational energy ratings (Lomas et al., 2019). However, evaluating retrofit impacts on individual properties will still require baseline data within a simulation method, which includes inputs for heating or temperature parameters. One approach to reducing uncertainty in this is through more measured data.

Open data, on indoor air temperatures, could be used to derive heating practices through inverse uncertainty quantification. More open data could enable archetype groups to be better represented. Previous research suggested socio-demographic groups could inform categories of heating behaviour (Kane et al., 2013; Huebner et al., 2015).

Open data

Open data is gaining increasing attention. A move towards this could require behaviour change from researchers, organisations and citizens (Dodds et al. 2020). A manifesto to increase access to data in engineering was set out (ibid) to include nine factors to build capacity, opportunity and motivation including consideration of potential negative impacts such as privacy (ibid). Privacy is concerned with international human rights laws and national and regional data protection laws, the most influential perhaps the General Data Protection Regulation (GDPR) of the European Union (Fjeld et al., 2020). The principles within privacy are considered to be consent, control over data use, ability to restrict processing, the right to rectification, the right to erasure, privacy by design (ibid). Therefore, data concerning home energy requires careful attention in terms of protecting householder privacy. Data should be anonymised, at the very least. Opportunities for householders to be the gate keepers of their data could be further explored.

Openly sharing data can shift power dynamics inherent in the control of information and allows anyone to use the data to innovate (Williams, 2015). This is especially useful where data is used to make decisions, such as around dwelling retrofit. If this went a step beyond researchers, making the data accessible and open to the people whose homes are represented could allow them to critique it and suggest personal insight which could annotate the data. Perhaps via open scholarship.

A recent example of open scholarship is through the use of a mobile application that public users (N=350,000) can input data to, including COVID-19 results from testing and symptoms assessments (N = 100 million) (covid.joinzoe.com, 2020). This data has enabled a recovery model to be built at pace, which estimates 52.2% of people recover within 13 days (ibid). Another output is a geographic heat map that can inform the public of cases in their local area (ibid). Early insight from this data has been published along with the code used to develop the application (Menni et al. 2020).

Data is openly shared by academics, Government and non-government organisations among others. Where Government data lacks, opportunity for researchers, organisations and citizens to share data may be more valuable. Platforms such as Figshare (2020) have been developed to enable data sharing for institutions, such as Loughborough University and publishers, such as Springer Nature, as well as individuals. Known data relating to retrofit analysis which could inform building simulation, in the UK as a start, is presented in Table 1.

Table 1. Existing open data in relation to retrofit evaluation

Overview	**Project and scale**	**Sector**	**Building data**	**Technologies and physical data**	**Social data**	**Opportunities**	**Link**
Mixed data on household energy use and local climate	REFIT: 20 core homes	Academic	Survey data	Gas consumption, electrical load measurements, internal air temperature, local climate data and sensor measurements	One survey (1,054 households), Semi-structured interviews and one structured survey on perceptions of smart homes	More case studies could build on this. (20.05.20 :9342 views, 2096 downloads & 4 citations since 06.2017)	https://www.refitsmarthomes.org/datasets/
Mixed data on household energy and occupant behaviour	LEEDR: 20 households	Academic	Building survey data, Images appliances, sensor location and calibration	Window and door opening, Electrical power measurement of incoming supply and appliances Gas and hot water consumption	Window and door opening, Passive Infra-Red room activity monitoring, Data processing procedures for all data.	More case studies could build on this. (at 20.05.20 had 1511 views, 223 downloads and 4 citations since 07.2018)	https://repository.lboro.ac.uk/articles/LEEDR_project_home_energy_dataset/6176450
National Energy Efficiency Data-Framework	NEED: UK wide	Government	Building attribute data (from Valuation Office Agency and Scottish Assessor)	Matches gas and electricity data (from Energy suppliers and transporters) with energy efficiency measures installed (from HEED and policy measures)	Household characteristics (from Experian) and socio-demographics (Government data)	Government-led overview. Local area data could be more granular. Academic research could fill gaps.	https://www.gov.uk/government/collections/national-energy-efficiency-data-need-framework
Home Energy Efficiency Data	HEED: UK wide	Government	Property characteristic and insulation measures installed	Heating systems and micro-generations installed	--	Universities, Government and non-profit organisations can access, but not householders	https://localhomesportal.est.org.uk/
Gas consumption data	Postcode level across UK	Government	Postcode level gas	Gas consumption	-	Compare EPC and gas use data at postcode level	https://www.gov.uk/government/statistics/postcode-level-gas-statistics-2018-experimental
Energy Performance Certificate data	Individual property level: UK wide	Government	Archetype, fabric construction	Heating system and micro-generations installed, estimates of gas and electricity use derived from BREDEM methodology	No. of occupants	Combined with operational data this could offer greater value (Lomas et al, 2019; Crawley et al. 2020)	https://epc.opendatacommunities.org/
English Housing Survey (50 years of data)	Housing circumstance, condition, energy efficiency	Government	Physical attributes	Energy efficiency	Household circumstance	Data could be used to inform 'property passport' data via the Trustmark data warehouse (below)	https://www.gov.uk/government/publications/50-years-of-the-english-housing-survey
Data stemming from Retrofit for The Future, an Innovate UK project for 100 social homes	Low Energy Building Database: case studies across UK	Non-Government organisation	Photographs (archetype, geometry, materials and age)	Energy use, air-tightness, indoor temperature, for some	Householder perception and experience, for some	Database could be advanced with more measured data	https://www.lowenergybuildings.org.uk/
Data warehouse on retrofit and social network	Trustmark	Government led organisation	Property passport data, measures installed	Data required under PAS 2035	Installer compliance, supply chain overview	Organisations and consumers can become early adopters	https://www.trustmark.org.uk/ourservices/data-warehouse
A list of data on energy consumption, fuel poverty and	Open Data Collaboration Initiative, CSE	Non-Government Organisation	Links to Government shared data on energy consumption, efficiency and census	All focused around energy consumption and efficiency	Fuel poverty data	Data can be accessed shared and discussed with CSE	https://www.cse.org.uk/projects/view/1259

Data already exists in open platforms, which can inform UK domestic retrofit evaluation and policy. However, there is opportunity to expand on this with research data to inform building simulation of existing buildings. A particular gap has been noted in measured indoor air temperature data that can address aleatory uncertainty and potentially form a stronger prior for Bayesian analysis techniques.

Open repositories for research development

Open repositories allow sharing of project information. The open access environment continues to evolve in terms of promoting open access journal publications (OAD, 2020a) and sharing of research data (OAD, 2020b). However, adopting practices during the research process may require a cultural shift. Practitioners at 81 universities were interviewed on their perspective of open research and for the majority of scholars', their understanding or embrace of openness within research dissemination was found to be *"patchy, ill-informed or confused"* (Johnson, 2018).

This movement is understood by the author to be pioneered by computer programmers and developers. For example, through making open source programming and analysis tools and sharing via GitHub (overview on Table 2). Collaboration is encouraged and methods can be continually improved by taking this approach. Following on the GitHub example, this allows effective version control as collaborators can 'clone' an existing project, work on the project and add their own developments without affecting the original version. They can then 'push' their work over to each other and 'pull' others work into their own copy (TTWC, Version control, 2019). This can help understand the thinking that has led to changes made, and eventually the method adopted. A selection of open repositories are highlighted in Table 2 with opportunities and challenges for each.

Table 2. A selection of open repositories and portals

Tool and link	Overview	Opportunities	Challenges
GitHub	Open repository for code development, blogs and project documenting	Global contributions can accelerate progress	Requires skills, intended for development of computer code
Jupyter-notebook	A code editing tool and notebook documents to represent computation outputs with annotation	Enables sharing of working progress and storing the development of code	Requires skills, intended for development of computer code
Hack.md	An open multi-platform collaborative knowledge base	Ideas can be co-developed and multiple data formats shared	This is a blog-like tool and requires pro-active sharing.
Pre-review https://content.prereview.org	A portal to share pre-prints of papers for community peer-review, prior to journal submission	Allows early feedback and sharing of ideas and research paper publications	Multi-disciplinary so likely requires a network to flag a paper to

This is just a snapshot of a variety of platforms, but as yet a suitable repository for developing socio-technical tools *and* sharing data has not been identified by the author. This would entail challenges to overcome. For example, Firth et al. (2018) identified that existing data formats are not suited to all studies and may need to be adapted to fit. The data structure being key to enable both ease-of-interpretation and the development of suitable code for analysis. Python and Jupyter notebooks were found to meet Open Methodology requirements by Firth et al. (2018). However, these are particulalrly useful for quantitative data and would likely require skill development by some transdisciplinary researchers to be accessible.

Retrofit evaluation factors and data types

This section provides a concise overview of factors and multiple data types that can be considered as part of retrofit evaluation, as shown in Table 3. The collection of data, or finding, combining and arranging data in the right format for retrofit building performance studies take time. Energy research is multidisciplinary and uses multiple methods including monitoring, modelling, interviews and case studies (Huebner and Fell., 2020). While time familiarising with research methods is valuable, perhaps an open suite of data and methods to analyse retrofit performance would save time and allow more results to be shared.

Table 3. Overview of retrofit data themes, insights & examples

Data theme	What these data can provide insight on	Data examples (not exhaustive)
Geographic location	Climate and strata conditions, plus potential flood risk	Maps, geographic survey data
Building characteristics	Geometry, age, archetype and physical construction of the building	Building surveys, photographs and satellite images
Retrofit products	Materials installed, their attributes and performance criteria	Product information from manufacturers, research evidence of risks, compliance data
Building performance	Impact of retrofit, in terms of energy demand, air-tightness, material performance and moisture flows	Measured energy data (pre/post), normalised, modelled data, air leakage tests, u-value measurements, hygrothermal surveys
Thermal comfort	Impacts of retrofit in terms of occupant comfort	Measured air temperature, velocity and occupant comfort
Indoor environment quality	Insight on the quality of internal environment pre and post retrofit	Measured indoor pollutant parameters and householder perception, ventilation rates and infiltration rates
Control	How heat, power and cooling are controlled and if effective	Usability questions to householder and evaluation of actual energy consumption
Householder behaviour	Heating, appliance, window and movement behaviour and energy impact	Sensors, diaries, surveys and interviews. For heating behaviour: indoor air temperature and building data.
Aesthetics	Consumer satisfaction, beyond environment and energy data	Qualitative data from occupants, photographs
Financial cost	Economic impacts	Householder survey, material and labour costs
Time	Time spent on the retrofit process	Householder or installer interviews, project schedules
Process	Physical installation practice & challenges encountered	Householder or installer interviews: qualitative

The data collected and shared for retrofit evaluation is transdisciplinary (Lomas, 2009). However, the number of studies with complete monitoring, equally capturing building data, technologies and people is limited (Higginson et al., 2018). As a consequence, this limits the impact of this research on policy, as reported by the Buildings Performance Institute Europe (BPIE) (Economiduo et al., 2011). Retrofit evaluation data could also be used to inform supply chain actors (Killip et al., 2020). A socio-technical retrofit evaluation tool and data repository could assist with these challenges. The building simulation community likely have experience, tools, data and ideas which could aid in shaping this.

As a move towards open, reproducible data requires behaviour change (Dodds et al., 2020), ethical practices (Fjeld et al., 2020), knowledge and skills of researchers (Johnson, 2018) and a framework (Kostkova et al. 2016), the next section presents The Turing Way Guide which could be adopted or adapted for use within the building simulation community.

The Turing Way Guide

The Turing Way Guide (The Guide) to reproducible data science aims to provide all the information that researchers need at the start of their project to ensure that they are easy to produce at the end (TTWC, 2019). It is an open project written in Jupyter book (2020).

The Guide is based upon the premise that reproducible research can ensure scientific work can be trusted by enabling code and results to be verified and built on in further work. The guide chapters include definitions in reproducible data science, open research, version control, licensing, collaborating on GitHub or Gitlab, community communications, data management, reproducible environments, code quality, testing and reviewing, continuous integration and more. Researchers who are new to open science and reproducible research may face barriers including insufficient documentation, disconnection from online communities or a lack of awareness of resources (Cereceda and Quinn, 2020).

The assessible format of the guide enables a quick check of reproducible queries. It is evolving and the collaborative approach has enabled a diverse range of inputs. The Guide also links to wider information and tools, such as FAIR (Findability, Accessibility, Interoperability and Reusability) (GOFAIR, 2020). The intent of FAIR is not only for the ‘data’ but also the methods, tools and workflows that led to the data (Wilkinson et al., 2016). Good data management and working processes are a pre-condition supporting knowledge discovery and innovation.

While building simulation for retrofit research is context dependent, fully reproducible results may not be feasible for every case, but it can serve as a goal to work towards and open principles can be used. The Guide’s chapter on open research is particulalrly valuable as a starting point. It covers open data, software, hardware, access and notebooks. In addition to the concept of open scholarship including open educational resource, equity, diversity and inclusion and citizen science. The Guide could inform open retrofit evaluation data and tools, with a focus on indoor air temperature initially.
The Guide targets academic researchers but could be of use to industry actors who are working in building performance evaluation. Figure 1 illustrates some of the guiding principles to the collaborative community approach and the journey to reproducible research.

Figure 1. Illustrations to highlight principles of The Turing Way Guide and journey to reproducible research (Scriberia)

Discussion and conclusion

Retrofit evaluation methods include building performance simulation of existing buildings to predict impacts of energy-efficiency technologies. Space heating energy demand of dwellings could be better predicted with more open data on indoor air temperature. This can potentially reduce uncertainty in building simulation as heating set point and duration were previously found to be the factors space heating energy demand is most sensitive to (Firth et al, 2009; Hughes et al., 2013). Researchers could assist in filling this gap, perhaps by taking a reproducible open approach, where ethical to do so. This could lead to more accurate predictions of space heating energy demand and potential savings following retrofit through the application of data relevant to existing buildings. Existing buildings are challenging to simulate (Heo et al., 2012) due to the variation between physical components and building occupancy and control. Further data on heating practices, anonymised to protect householder privacy, could inform calibration data.

Wider socio-technical factors evaluated in retrofit, are captured by a variety of data types. There is a need for a repository that can handle multiple datasets and formats. It should be accessible to researchers of multiple disciplines to provide input, develop the tools and share data. This can enable progress towards consistent metrics.

This paper has presented The Turing Way Guide as a tool for researchers and practitioners. This can inform reproducible research practice, and potentially be adopted and adapted by the community of building performance and retrofit evaluation researchers. There appears to be emerging interest to develop this space.

The UK is progressing in terms of open research and data. Particularly in data shared by the Government and movements towards open data in academic institutions. Retrofit and building performance research has not always been carried out in an open way and towards reproducibility. This could be partly due to building evaluation being context dependent and partly due to culture and practices in addition to ethics concerned with the personal nature of data within the home environment.

At a time of declared climate emergency, enabling more informed retrofit decision-making is necessary. Future opportunities exist to collaboratively advance open source tools, enabling wider input for development in addition to data sharing by researchers and practitioners. This could address some uncertainty in existing building simulation and ultimately accelerate progress in carbon reduction.

Acknowledgement

This work was carried out under the Grand Challenge area Data-Driven Design Under Uncertainty within The Alan Turing Institute's Data Centric Engineering programme under the Lloyd's Register Foundation grant G0095. The author is grateful to the BSO-V 2020 reviewers who provided constructive comments which strengthened this paper.

References

Booth, A., Choudhary, R. and Spiegelhalter, D.J. (2012) Handling uncertainty in housing stock models, Building and Environment, Vol. 48, pp 35-37

Booth, A. (2013) Handling uncertainty in the retrofit analysis of the UK housing stock, PhD Thesis, Robinson College, University of Cambridge

BRE Group (2020) Standard Assessment Procedure (SAP 2012). Available from: https://www.bregroup.com/sap/standard-assessment-procedure-sap-2012/ [Accessed 16/04/20]Cereceda, O. and Quinn, D.E.A. (2020) A graduate student perspective on overcoming barriers to interacting with open-source software, Facets, DOI: https://doi.org/10.1139/facets-2019-0020

Cheng, V. and Steemers, K. (2011) Modelling domestic energy consumption at district scale: A tool to support national and local energy policies, Environmental Modelling & Software, Vol. 26 (10), pp. 1186-1198

Crawley, J., McKenna, E., Gori, V., & Oreszczyn, T. (2020). Creating Domestic Building Thermal Performance Ratings Using Smart Meter Data. Buildings and Cities, 1(1), 1–13. DOI: http://doi.org/10.5334/bc.7

Covid.joinzoe.com (2020) Data update: New COVID prevalence figures for the UK, July 8th, 2020. Available via: https://covid.joinzoe.com/post/data-update-prevalence-covid [accessed 09/07/20]

Dodds, L., L'Henaff, P.,Maddison, J. and Yates, D. (2020) A manifesto for increasing access to data in engineering, Data-Centric Engineering, Vol. 1: e5, doi:10.1017/dce.2020.3

Economidou, M., Atanasiu, B., Despret, C., Maio, J., Nolte, I., & Rapf, O. (2011). Europe's Buildings Under the Microscope: A Country-by-country Review of the Energy Performance of Buildings. Brussels: Buildings Performance Institute Europe (BPIE).

Figshare (2020) Figshare: Store, share discover research. Available from: https://figshare.com/ [Accessed 22/-5/20]

Firth, S.K., Cole ,G., Kane, T., Fouchal, F. and Hassan, T.M. (2018) An Open Science Approach for Building Performance Data Schemas and Analysis Algorithms. Proceedings of BSO 2018, 4th Building Simulation and Optimization Conference, Cambridge, UK: 11-12 September 2018

Firth, S., Lomas, K. and Wright, A.J. (2009) Targeting household energy-efficiency measures using sensitivity analysis, Building Research & Information, Vol. 38 (1), pp. 25-41. DOI: 10.1080/09613210903236706

Fjeld, J, Achten, N, Hilliglos, H, Nagy, A, Srikumar, M (2020) Principled Artificial Intelligence: Mapping Consensus in Ethical and Rights-Based Approaches to Principles for AI. Available at https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3518482 [accessed 17/07/20]

GOFAIR (2020) FAIR Principles, available from: https://www.go-fair.org/fair-principles/ [accessed 20.05.20]

Heo, Y., Choudhary, R., Augenbroe, G. (2012) Calibration of building energy models for retrofit analysis under uncertainty, Energy and Buildings, Vol. 47, pp. 550-560, doi.org/10.1016/j.enbuild.2011.12.029

Higginson, S., Topouzi, M., Andrade-Cabrera, C., O-Dwyer, C., Darby, S. and Finn, D. (2018) Achieving data synergy: The socio-technical process of handling data, pp. 63-81. In: Foulds C., Robison R. (eds) Advancing Energy Policy. Palgrave Pivot, Cham

HM Gov (2017) Clean Growth Strategy: Leading the way to a low carbon future

Hong, T., D'Oca, S., Turner, W.J.N. and Taylor-Lange, S.C. (2015) An ontology to represent energy-related occupant behavior in buildings. Part I: Introduction to the DNAs framework, Building and Environment, Vol. 92, pp. 764-777, doi.org/10.1016/j.buildenv.2015.02.019

Hong, T., Taylor-Lange, S. C., D'Oca, S., Yan, D. and Corganti., S.P. (2016) Advances in research and

applications of energy-related occupant behaviour in buildings, Energy and Buildings, Vol. 116, pp. 694-702, doi.org/10.1016/j.enbuild.2015.11.052
Huebner, G., & Fell, M. J. (2020). TReQ Tools: How to Improve Transparency, Reproducibility and Quality in Energy Research. doi:10.6084/m9.figshare.11663466
Huebner, G.M., Nicolson, M.L., Fell, M.J., Kennard, H., Elam, S., Hanmer, C., Johnson, C. and Shipworth, D. (2017) Are we heading towards a replicability crisis in energy efficiency research? A toolkit for improving the quality, transparency and replicability of energy efficiency impact evaluations, proceedings of ECEEE Summer Study Proceedings
Huebner, G.M., McMichael, M., Shipworth, D., Durand-Daubin, M. and Summerfield,A.J. (2015) The shape of warmth: temperature profiles in living rooms, Building Research & Information, Vol. 43 (2), pp. 186-196
Hughes, M., Palmer, J., Cheng, V. and Shipworth, D. (2013) Sensitivity and uncertainty analysis of England's housing energy model, Building Research & Information, Vol. 41 (2), pp. 156-167, DOI:10.1080/09613218.2013.769146
Hurst, W. (2020) Calls for retrofit drive to spark post-Covid 'green economic recovery', Architects Journal. Available from: https://www.architectsjournal.co.uk/news/calls-for-retrofit-drive-to-spark-post-covid-green-economic-recovery/10047044.article?blocktitle=More-architecture-news&contentID=16149 [Accessed 13.05.20]
Johnson, G. J. (2018) Cultural, ideological and practical barriers to open access adoption within the UK Academy: an ethnographically framed examination, Insights 31: 22. DOI: http://doi.org/10.1629/uksg.400
Jupyter book (2020) Overview and installation. Available from: https://jupyterbook.org/start/overview.html [Accessed 22/05.20]
Kane, T., Firth, S., Allinson, D., Irvine, K. and Lomas, K. (2013) Understanding occupant heating practices in UK dwellings, N: Moshfegh, B. (ed.). World Renewable Energy Congress - Sweden, 8-13 May, 2011, Linköping, Sweden, vol.3 (057), pp.1126 -1133.
Kane, T., Firth, S.K. and Lomas, K.J. (2015) How are UK homes heated? A city-wide, socio-technical survey and implications for energy modelling, Energy and Buildings, Vol. 86, pp. 817-832, doi.org/10.1016/j.enbuild.2014.10.011
Kavgic., M., Mavrogianni, A., Mumovic, D., Summerfield, A., Stevanovi, M., Djurovic-Petrovic, M. (2010) A review of bottom-up building stock models for energy consumption in the residential sector, Building and Environment, Vol. 45, pp. 1683-1697
Killip, G., Owen, A. and Topouzi, M. (2020) Exploring the practices and roles of UK construction manufacturers and merchants in relation to housing energy retrofit, Journal of Cleaner Production, Vol. 251, 119205
Kostkova, P., Brewer, H., de Lusignan, S., Fottrell, E., Goldacre, B., Hart, G., Koczan, P., Knight, P., Marsolier, C., McKendry, R.A., Ross, E., Sasse, A., Sullivan, R., Chaytor, S., Stevenson, O., Velho, R. and Tooke, J. (2016) Who Owns the Data? Open Data for Healthcare. *Front. Public Health* 4:7. doi: 10.3389/fpubh.2016.00007
Kraker, P., Leony, D., Reinhardt, W. and Beham, G. (2012) The case for open science in technology enhanced learning, International Journal of Technology Enhanced Learning, Vol. 3 (6). DOI: https://doi.org/10.1504/IJTEL.2011.045454#
Lomas, K. (2009) Carbon reduction in existing buildings: a transdisciplinary approach, Building Research & information, Vol. 38 (1), pp. 1-11
Lomas, K. J., Beizaee, A., Allinson, D., Haines, V.J., Beckhelling, J., Loveday, D.L., Porritt, S.M., Mallaband, B., Morton, A. (2019) A domestic operational rating for UK homes: Concept, formulation and application, Energy and Buildings, 201, pp.90-117, ISSN: 0378-7788. DOI: 10.1016/j.enbuild.2019.07.021.
Menni, C., Valdes, A.M., Freidin, M.B. et al. (2020) Real-time tracking of self-reported symptoms to predict potential COVID-19. Nat Med. https://doi.org/10.1038/s41591-020-0916-2
OAD (Open Access Directory) (2020a) Journals that converted from TA to OA. Available from: http://oad.simmons.edu/oadwiki/Journals_that_converted_from_TA_to_OA. [accessed 20.05.20]
OAD (2020b) Data Respositories. Available from: http://oad.simmons.edu/oadwiki/Data_repositories [Accessed 20.05.20]
Oreszczyn, T., Huebner, G., Shipworth, D. (2020) How should energy researchers respond to a climate emergency? UCL Energy Institute Working Paper. Available from: https://figshare.com/articles/How_should_energy_researchers_respond_to_the_climate_emergency_/12074226 [Accessed 13.05.2020]
Scriberia illustrations for The Turing Way community and used under a CC-BY licence: Available from here: https://doi.org/10.5281/zenodo.3332807
Shipworth, M. (2011) Thermostat Settings In English Houses: No Evidence Of Change Between 1984 And 2007, Building and Environment, 46(3), 635-642
Swan, L. G. and Ugursal, I. (2009) Modeling of end-use energy consumption in the residential sector: A review of modelling techniques, Renewable and Sustainable Energy Reviews, Vol.13 (8), pp. 1819-1835
Tian, W., Heo, Y., de Wilde, P., Li, Z., Yan, D., Soo Park, C., Feng, X. and Augenbroe, G. (2018) A review of uncertainty analysis in building energy

assessment, Renewable and Sustainable Energy Reviews, Vol. 93, pp. 285-301

TTWC (The Turing Way Community), Becky Arnold, Louise Bowler, Sarah Gibson, Patricia Herterich, Rosie Higman, Anna Krystalli, Alexander Morley, Martin O'Reilly and Kirstie Whitaker. (2019, March 25). The Turing Way: A Handbook for Reproducible Data Science (Version v0.0.4). Zenodo. 10.5281/zenodo.3233853.

Wilkinson, M., Dumontier, M., Aalbersberg, I. et al. (2016) The FAIR Guiding Principles for scientific data management and stewardship. Sci Data 3, 160018. https://doi.org/10.1038/sdata.2016.18

Williams, S. (2015) More than data: Working with big data for civics. A Journal of Law and Policy, Vol. 11 (1), pp. 181-199

Zero Carbon Hub (2014) Closing the gap between design & as-built performance: End of term report. Available from: http://www.zerocarbonhub.org/sites/default/files/resources/reports/Design_vs_As_Built_Performance_Gap_End_of_Term_Report_0.pdf [Accessed 20.05.20