A Digital Curation Model for Post-Occupancy Evaluation Data

Abstract

Post-Occupancy Evaluation (POE) is an important component of a cyclical design process that allows cost savings and promotes accountability of designers, owners, and users, as well as allowing identification of facilities issues and informed decision-making. It is reasonable to expect that the combined effect of new data-gathering technologies and the increasing urgency of sustainable building design will be that POE studies will increase and, eventually, be applied at a mass scale. Digital Curation (DC) allows for the long-term preservation and constant re-use of research data. Yet there has been no attempt so far to develop a DC strategy for POE data. This paper introduces the concept of Digital Curation and relates it to objective, quantitative POE data. A combination of a literature review and a questionnaire sent to expert respondents is utilised in order to identify data types and major Digital Curation issues relevant to POE. A Digital Curation Model (DCM) for POE data is presented, drawing from the existing models, examples, and best practice available, intended to act as a framework that enables the storage, archiving, management, recovery and utilisation of such data on a mass scale. Finally, the importance, benefits, obstacles and implications of developing a suitable DC strategy for Post-Occupancy Evaluation are discussed.

keywords: Post-Occupancy Evaluation; Digital Curation; Data Management; Building Monitoring

INTRODUCTION

Background to Post-Occupancy Evaluation

Post-Occupancy Evaluation (POE), defined by Zimring and Reizenstein (1980) as “the examination of the effectiveness for human users of occupied design environments”, has been conducted in a systematic fashion for more than 40 years (W. F. E. Preiser, 2005). The main aim of POE is to provide feedback of the building performance to designers, owners, users and other stakeholders. This can facilitate a cyclical design process but it can also inform the building operation, management and use. The benefits of an effective POE are therefore manifold, from short-term benefits, such as allowing quick identification of facilities issues and supporting informed decision making, to longer-term benefits such as cost savings throughout the building lifecycle and greater accountability of designers, owners and users for the building performance (W. F. Preiser, White, & Rabinowitz, 2015). More generally, the application of POE to a mass scale will contribute to general design knowledge about the performance of buildings and therefore enable continuous improvement (Zimmerman and Martin, 2001).

The broad definition of POE given above includes a very wide range of techniques that cover a great variety of building aspects. Bordass and Leaman (2005) presented a range of techniques, that allow stakeholders to obtain qualitative feedback from users. The acceptance by the design community of the phenomenon of the performance gap (De Wilde, 2014) has also increased interest in quantitative forms of POE, where objective parameters, such as energy consumption or indoor environmental properties, are measured and analysed in order to identify the actual performance of a building (Menezes, Cripps, Bouchlaghem, & Buswell, 2012). A recent review by Li *et al.* (2018) found that the published research work on POE showed a substantial increase after 2010.

It is reasonable to expect that this increase will continue. The interest in understanding better the performance of buildings is becoming greater, particularly with regard to their environmental impact in the context of the climate emergency. Post-occupancy evaluation is an important mechanism to achieve this enhanced understanding. Simultaneously, the cost of collecting data is dropping constantly, and more properties can be measured more quickly and thoroughly than ever before, while more methods of analysis are available via Machine and Deep Learning techniques. The introduction of the Internet of Things (IoT) will increase exponentially the data that can and will be collected. Individual elements, from building management systems to everyday personal objects, such as household appliances or mobile phones, can come equipped with their own in-built sensors, data logging software, and capacity to upload to the cloud on a constant and instant basis. This will mean a multitude of formats and datasets, with different owners and standards. This not only raises a host of issues, such as privacy, ownership, storage, and duplication, but also presents a missed opportunity, by not unifying this data for a common purpose. Digital Curation (DC) provides a strategy for the long-term storage, archiving, management, recovery and utilisation of data. However, no attention has been paid until now to the need for, and importance of, a DC strategy for POE data.

Such a strategy can deliver substantial benefits. The nominal lifecycles of buildings are 50-120 years, and often substantially more. These periods are extremely long compared to the timescales of data storage and management tools and formats (often changing every 10-20 years), as well as the stakeholders and managers themselves (who can change as often as every year in some cases). If the lifecycle performance is to be assessed consistently and effectively, appropriate digital curation techniques need to be found. Applied at a mass scale, a standardised approach to the storage and management of such data, promises substantial benefits: it can allow the comparison between different buildings of similar typologies; it can facilitate the identification of community performance at various levels, from community to international; it can create the necessary common basis to allow the implementation of Big Data/Machine Learning techniques. These, however, requires appropriate Digital Curation strategies.

Background of Digital Curation

With digital technologies and the switch from sequential processing to random access storage, the preservation, sharing, and usage of digital data have developed at a rapid pace during the past 15 years. Data management has become a field of research in itself, with different groups setting guidelines and standards for data retention and sharing, particularly on the World Wide Web (WWW), whose development has promoted a rush on the part of institutions to create web repositories for data that act as data management infrastructures.

A number of definitions have been suggested to describe these practices, such as digital data management, data curation, knowledge management (Girard, 2015), and different groups have been using different terminologies. Digital Curation was first used by JISC in 2003, defined as "[t]he activity of managing and promoting the use of data from its point of creation, to ensure it is fit for contemporary purpose, and available for discovery and re-use" (Lord, 2003).

For the purposes of this work, the more recent definition by the Digital Curation Centre (DCC) is used, namely that of "maintaining, preserving and adding value to digital research data throughout its lifecycle" (Digital Curation Centre). Digital Curation allows for the long-term preservation and constant re-use of research data. What adds value to the original data is the set of actions performed on the dataset that allows for the long-term preservation, accessibility, openness and re-usability.

Various models have been developed in order to provide a general framework for Digital Curation, with the objective of allowing researchers to benefit from the sharing possibilities and analytical tools offered by digital technologies, promoting "data infrastructure literacy" which allows participation in the collection and analysis of data (Gray, Gerlitz, & Bounegru, 2018).

Discipline-specific projects have also been set up, with the purpose of building online repositories for research data. These projects can be examples of Tim Berners Lee's "Social Machine": digital research infrastructures that allow humans and machines to cooperate in order to create knowledge (Berners-Lee, 2008). Unfortunately, as has been recently indicated (Huebner and Mahdavi, 2019), the existing digital ecosystem surrounding building infrastructure dataset seems to be completely scattered, with little effective discussion amongst practitioners on how to benefit from data sharing.

There is currently no work specific to POE, neither open repositories, nor guidelines, though some frameworks make general provisions . For example, the Industry Foundation Classes (IFC) used in Building Information Modelling (BIM) contain generic extensible support for Post-Occupancy Evaluation measurements via the IfcSensor class (buildingSMART). Semantic data frameworks such as Project Haystack and BRICK contain relevant tags for environmental properties and sensors. However, these examples focus on asset representation and not POE data, thus not engaging with the long-term storage, archival, analysis, and recovery of large environmental datasets. Moreover, they are intended to be holistic, not specialised. This has an impact on two levels: firstly, they typically require a substantial level of information to be workable and often, particularly in the case of BIM, via highly complex, proprietary software applications. Secondly, they are geared towards capturing building elements that don't change often. This means that large-scale building monitoring over prolonged periods becomes difficult or impossible. For example, a BIM or BRICK file containing POE data for 10,000 buildings over 10 years would be highly uneven in its data distribution and ultimately unwieldy.

It can be concluded that existing frameworks have been developed with the aim of capturing the holistic, current picture, and without an emphasis on long-term and mass-scale data capture and retention. Disciplines such as biology (NIH, 2020) and economics (ESRC; Lab) have adopted principles of digital curation and set up specialised, centralized repositories for the collection, sharing, and preservation of the datasets produced by their scientific communities. By contrast, DC does not appear as part of POE / Building Monitoring. Thus, there is a clear research gap in this area.

Aim and Objectives

The aim of this work is to introduce the concept of DC for objective, quantitative POE data, via a prototype model that facilitates the storage, archiving, management, recovery, and utilisation of such data on a mass scale, while highlighting the importance, benefits, and obstacles of developing a suitable DC strategy for POE.

POE places importance on both qualitative/subjective data and quantitative/objective data. However, the focus of this paper is on quantitative/objective data only, with an emphasis on environmental performance and other factors that are likely to lead to energy consumption. The reasons for this are twofold:

- Environmental data are quantifiable data, measured precisely in physical units. Therefore, they are more homogenous, lend themselves better to engineering analysis, and can be collected automatically without human involvement. By contrast, qualitative data are widely varied, by nature subjective, and require human input for their collection. Thus they tend to be case-specific and without the capacity to generate substantial volumes of data (often referred to as "Big Data") under a common framework.
- The contribution of buildings to climate change, now given greater urgency through recognition as the climate emergency, pushes the environmental performance and energy consumption of buildings more urgently as a top priority for designers, researchers, policy makers, and other users/stakeholders (Construction, Agency, & Programme, 2019).

The objectives of this paper are as follows.

O1. Introduce Digital Curation concepts and practices relevant to objective POE data.

O2. Investigate the current state-of-the-art of Digital Curation practice amongst researchers and practitioners who collect such data, including considerations, methods, and obstacles.

O3. Develop a prototype Digital Curation model for objective environmental POE data.

O4. Discuss the benefits and costs of developing a Digital Curation strategy for POE data at a mass scale, and identify obstacles to this being achieved.

METHODOLOGY

Fundamental DC Concepts and Practices

A general review of the current literature and models proposed was undertaken. Digital Curation models are created by archivists and records managers working in institutions (Public Administration and historical archives), as well as companies, particularly those dealing with large amount of data. As digitalisation processes became more common, general data curation models have been developed to provide organisations with a common basis upon which to build customised data management systems for the entire data lifecycle, including data that need long-term preservation in the long term. A 2016 review identified, nine main lifecycle models (Poole, 2016).

1) The Digital Curation Centre (DCC) Lifecycle Model;
2) The I2S2 Idealized Scientific Research Activity Lifecycle Model by the The United Kingdom Office for Library and Information Networking (UKOLN)
3) The DDI (Data Documentation Initiative Alliance) Digital Combined Life Cycle Model
4) The ANDS (Australian National Data Servces) Data Sharing Verbs
5) The Data ONE Data Lifecycle by the Data Observation Network for Earth
6) The Research 360 Institutional Research Lifecycle by the University of Bath
7) The Capability Maturity Model for Scientific Data Management by the Syracuse University School of Information Studies
8) The UK Data Archive Data Lifecycle from the University of Essex
9) The OAIS (Open Archival Information System) model by the Consultative Committee for Space Data Systems (CCSDS)

For the purposes of this work, three of those frameworks are utilised:

- the OAIS Reference Model, which was accepted as an international standard in 2003 (ISO 14721), and is currently the only ISO standard for Digital Preservation (Laughton and Du Plessis, 2013)
- the Lifecycle model developed by the Digital Curation Centre (Higgins, 2008), which is particularly flexible, yet extensive, and has seen wide adoption from a range of organisations
- the FAIR principles, which are specifically designed for research data and reflect a range of university-research based models (Wilkinson et al., 2016)

These are presented in the “Core Concepts in Digital Curation” section.

POE Literature Review

In order to develop a suitable DC model it is necessary to have an overview of the various types of data currently collected by POE researchers and professionals, as well as any existing DC strategies amongst individual research groups. Scopus was utilised with the following inclusion criteria:

- Research papers published in peer-reviewed journals in the past five years from the time of search (i.e. since 2014), as the focus of this research is the current state-of-the-art in POE data collection. This reflects the adoption of real-time collection of data (typically via Wi-Fi networks), compared to previous systems that relied on periodic collections of data via downloading. Older work is supplemented via literature reviews.
- The article title, abstract, or keywords needed to include at least one of the terms in List A and at least one of the terms in List B in Table 1 below.

Table 1. Inclusion criteria

List A	List B
○ post-occupancy evaluation ○ building monitoring ○ indoor quality ○ indoor assessment These are the terms typically associated with the collection of objective POE data. The terms “indoor quality” and “indoor assessment” are used as supersets that can capture terms such as “indoor environmental quality”, “indoor	○ energy ○ fuel ○ environmental ○ climate ○ thermal ○ light ○ water

environmental assessment", "indoor environmental quality assessment" etc.	These are the terms typically associated with the environmental building performance and energy consumption.

The search returned 429 documents. The abstracts of those papers were read in full, in order to identify research papers that describe projects relevant to the work presented here.

The following types of papers were considered relevant, i.e. providing insights to data that can be measured:

- papers that dealt explicitly with the measuring of objective, environmental data measured automatically, i.e. via data loggers, and presented those results;
- reviews of POE work conducted by third parties;
- conceptual or ontological frameworks that reference POE work by others.

In order to focus on the state-of-the-art, papers that were not in a Q1 or Q2 journal and had no citations (excluding self-citations), were filtered out.

This filtering returned 40 papers, of which 37 were research papers documenting specific projects, 2 were literature reviews, and 1 dealt with building monitoring ontology. A further 12 papers had aspects that dealt with categorisations of objective POE data and have been considered as supplementary material. The 37 papers were studied in order to identify the types of data collected by researchers, with the reviews used to verify the information found in the research papers. The results are presented in the "POE Literature Review Results" section.

POE Questionnaire

The published POE works surveyed did not cover issues relating to DC, especially with regard to archiving and long-term storage. As such, it was deemed necessary to contact the researchers in order to get more information with regard to their DC practices, via an online questionnaire.

The questionnaire sought to identify issues around data collection and retention. On data collection, the questions aimed to gauge the user- and equipment-related issues. On data retention, the questions aimed to gauge the frequency and rationale of data retention, as well as storage and access issues, including data harmonisation aspects. Finally, there were two dedicated questions focusing on the cost proposition of POE. The participants were the researchers identified in the literature review that were either still active in a research role or working in the construction industry in a role with POE involvement. In addition, a small number of prominent practitioners conducting POE research were identified and added to the participant list.

A total of 30 researchers (population size N = 30) were contacted, covering all of the 37 papers (as some authors had multiple appearances in the list of identified papers). The response rate was 70%, as a total of 21 responses were received (sample size n = 21), corresponding to a 12.5% margin of error (e=0.125) at a 95% confidence level (z = 1.96).

While the sample size provides a reasonably small margin of error, there is a question if the chosen population size is adequate to reflect contemporary practice. This population reflects the lead researchers that have published work in the leading journals in this area, and the respondents are from leading institutions covering Europe, Asia, North and South America, and Oceania. This work focuses on the state-of-the-art, as it is likely that the issues the leading researchers identify will be eventually encountered by others that do work in this area using similar techniques. It possible, however, that more insights could be gained by exploring the practices adopted, and issues faced, by researchers and practitioners who work with more established techniques, outside those covered recently in the leading journals. This would be a suitable next step for a further elaboration of this framework.

The results are presented in the “Questionnaire Results” section.

CORE CONCEPTS IN DIGITAL CURATION

The OAIS Reference Model

The OAIS Reference Model describes the environment of an optimal digital repository as composed of four entities (Figure 1) (Lavoie, 2004):

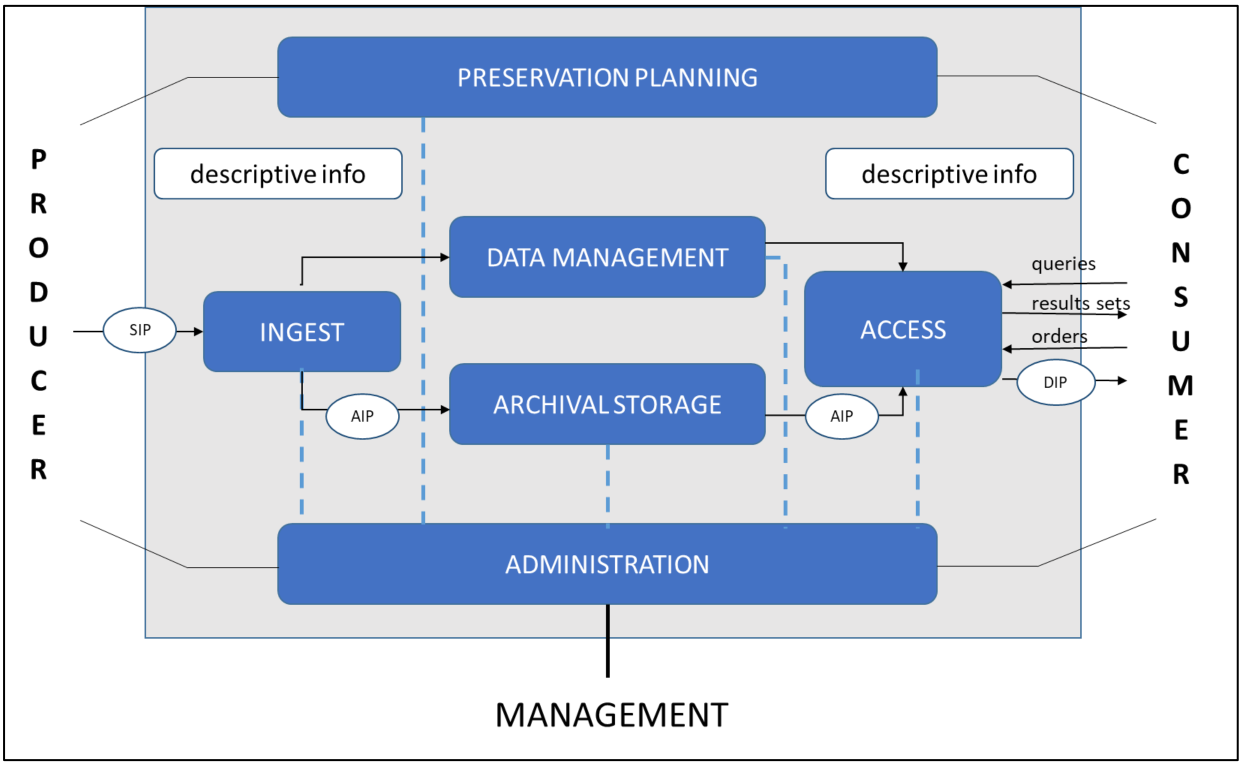

Figure 1. OAIS Reference Model, adapted from (Lavoie, 2004)

- **Producers**, who produce the data and the dataset.
- **Consumers**, anyone with an interest, current or future, in the data; human and/or machines.
- **Management**, the entity responsible for establishing the policy objectives of the archive.
- **Archive**, the repository in itself, indicated as the sum of all datasets, infrastructures where it is held, and the community maintaining it.

The information is circulated via different types of **Information Packages** (defined as AIP, SIP, and DIP in Figure 1).

The Digital Curation Lifecycle Model

Similar to the OAIS Reference Model, but with more emphasis on performance rather than on information collection, the DDC Digital Curation Lifecycle suggests the series of actions that should be performed on a digital object to ensure its preservation and usability (Figure 2) (Higgins, 2008).

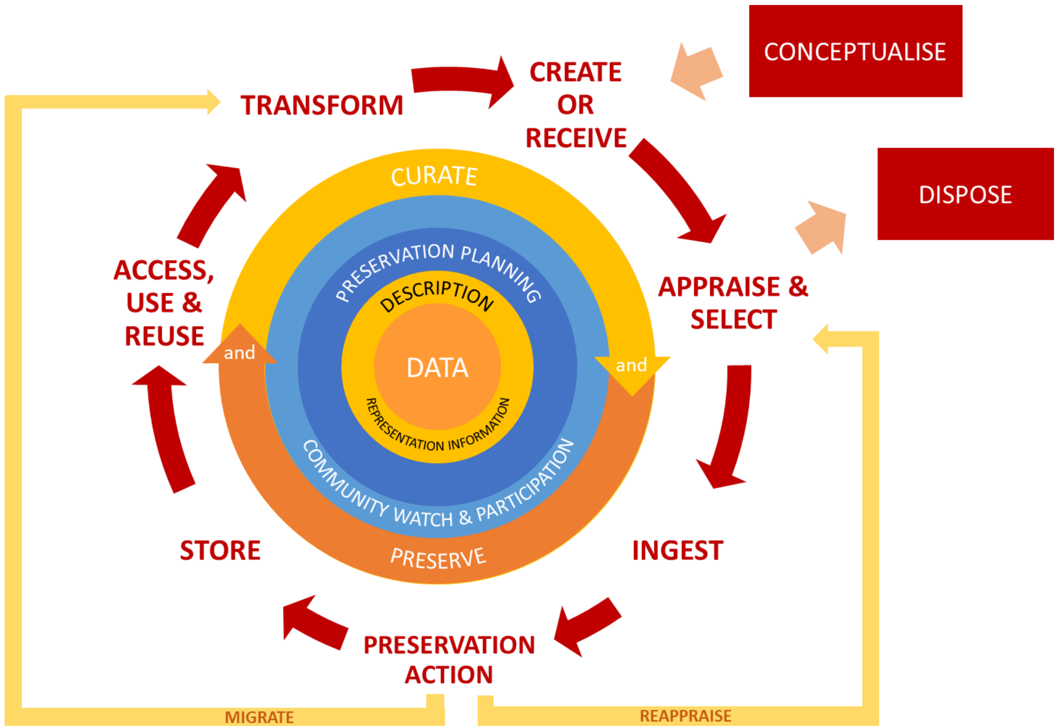

Figure 2. The DDC Digital Curation Lifecycle, adapted from (Higgins, 2008)

The Lifecycle utilises the following concepts:

- **Conceptualise / Create / Receive** – creation / reception of the data and their metadata.
- **Appraise / Select** – evaluation of the digital objects / research data to select those requiring long-term curation and preservation; to be carried out according to a clear set of pre-defined guidelines.
- **Dispose** – disposal of redundant, useless data.

- **Ingest** – ingestion of the data in the system/repository that would allow for access and use.
- **Preservation** – carry out the series of actions to ensure the long-term preservation and retention of the digital object; this may require migration, transformation, etc.
- **Store** – keep the data in a secure manner as outlined by relevant standards: this includes the physicality of the servers, drives, etc, where the digital objects are preserved.
- **Access / Use / Reuse** – ensure that the digital objects can be accessed and used by the designated users, whether the general audience or a restricted group of people.
- **Reappraise** – periodical re-validation of the importance of the digital object. Disposal when necessary.

The FAIR Principles

These Principles were created by the FAIR Initiative, in order to establish the guidelines for data curation amongst researchers from academia and industry (Wilkinson, et al., 2016). FAIR is the acronym for the four principles indicated by the consortium:

- **Findability –** both data and metadata must be easy to find for both humans and computers. Machine-readable metadata are essential for automatic discovery of datasets and services.
- **Accessibility** – authentication and authorisation may be required; however, standardised communications protocols that are open, free, and universally accessible should be implemented.
- **Interoperability –** data and metadata need to be integrated with other data and with workflows for analysis, storage, and processing; they should use a broadly applicable language for knowledge representation, controlled vocabularies and references.
- **Reusability** – data and metadata and data should be well-described so that they can be replicated and/or combined in different settings, with clear data usage licenses, accurate description and clear provenance (Initiative).

FAIR places considerable emphasis on machine actionability, i.e. the capacity of computational systems to find, access, interoperate and reuse data with minimal or no human intervention. Though it is not explicitly mentioned, there are strong references to the concept of the WWW as a "social machine," (Hendler and Berners-Lee, 2010). Several prominent web repositories, such as Dataverse, Mendeley Data, DataHub, and DANS follow the FAIR principles and act as general-purpose digital curation projects for metadata.

Implications for the Development of a DC model for POE

While a central OAIS-compliant repository for POE data would be an ideal outcome, the establishment and maintenance of such repositories is challenging, requiring networks of stakeholders to manage the repository and set the guidelines, as well as funding and time to guarantee long-term sustainability. The work presented here covers the initial stages of data curation, by suggesting a subject-specific metadata model that can be adopted by individual researchers during data collection.

POE LITERATURE REVIEW RESULTS

Literature reviews on POE have taken a wide-ranging view, mostly focusing on subjective data. Vásquez-Hernández and Álvarez (2017) have identified the following key aspects: Environmental, Physical and space, Psychosocial, and Socioeconomic. However, objective and subjective elements overlap: for example, thermal comfort can be subjective (user-reporting), or objective (sensor measurements). Moreover, socioeconomic aspects often do not address the building, but user circumstances (e.g. travelling time). In total, they identified that 15.1% of the studies included direct building monitoring. Guerra-Santin and Tweed (2015) provided a comprehensive review of measurement methods, classifying monitoring as embedded, reflective-engaged, and independent. Moreover, they provided descriptions of the methods for the identification of thermal comfort, energy consumption, and what they describe as ‘building operation’, referring to the effect that occupants’ actions have on the efficiency of the building technologies.

From a theoretical perspective, Mahdavi and Taheri Mahdavi and Taheri (2017) have created an extensive ontology that is intended to be applied to all types of data that can be monitored in buildings, classified in six categories. While all-encompassing ontologies are common in academic research, they often find little applicability in practice. From a Digital Curation perspective they are problematic, supporting over-collection of data without consideration of how this data will be managed, archived, processed, or retrieved. Such data structures result in highly demanding computational operations, while the associated information to make such data meaningful is overlooked. For some types of POE, this is further exacerbated by the ease of automating data collection.

The work presented here focuses on three categories, energy consumption data, indoor environment data, and/or outdoor environment data, covered by objective measurements (i.e. via equipment, and not user perception. A total of 37 research projects were identified as fulfilling those criteria. The types of environmental and energy data collected in the papers reviewed are given in Table 2.

Table 2. Types of Measured Data in the surveyed works

Data type	*Papers*
Data relating to Energy Consumption	
Electricity use [general]	(Berge and Mathisen, 2016; Filippín, Larsen, & Marek, 2015; Gupta, Barnfield, & Hipwood, 2014; Gupta and Gregg, 2016; Sodagar and Starkey, 2016; Touchie and Pressnail, 2014; Yu, Du, & Pan, 2019)

Electricity use [divided by type of activity]	(Ascione, Bianco, Böttcher, Kaltenbrunner, & Vanoli, 2016; O. Guerra-Santin, Romero Herrera, Cuerda, & Keyson, 2016)
Gas use	(Filippín, et al., 2015; Gupta, et al., 2014; Gupta and Gregg, 2016; Pretlove and Kade, 2016; Sodagar and Starkey, 2016; Touchie and Pressnail, 2014; Yu, et al., 2019)
Water use	(Gupta and Gregg, 2016; Pretlove and Kade, 2016; Sodagar and Starkey, 2016)
Photovoltaic use	(Gupta and Gregg, 2016; Pretlove and Kade, 2016; Sodagar and Starkey, 2016)
Indoor Environment Data	
Temperature	(Adaji, Adekunle, Watkins, & Adler, 2019; Adekunle and Nikolopoulou, 2016; Baja et al., 2019; Barreca and Praticò, 2018; Berge and Mathisen, 2016; Choi and Lee, 2018; Colclough, Kinnane, Hewitt, & Griffiths, 2018; Dabaieh and Johansson, 2018; Deuble and de Dear, 2014; Doctor-Pingel, Vardhan, Manu, Brager, & Rawal, 2019; Filippín, et al., 2015; Gerrish, Ruikar, Cook, Johnson, & Phillip, 2017; O. Guerra-Santin, et al., 2016; Gupta, et al., 2014; Gupta and Gregg, 2016; Gupta and Kapsali, 2016; Hua, Göçer, & Göçer, 2014; Ioannidis et al., 2017; Jentsch et al., 2017; Lee, Wargocki, Chan, Chen, & Tham, 2019; Liu, Wang, Zhang, Hong, & Lin, 2018; Loyola, 2019; Naspi, Arnesano, Stazi, D'Orazio, & Revel, 2018; Park, Loftness, & Aziz, 2018; Pastore

	and Andersen, 2019; Patlakas, Koronaios, Raslan, Neighbour, & Altan, 2017; Patlakas, Santacruz, & Altan, 2014; Pei, Lin, Liu, & Zhu, 2015; Pretlove and Kade, 2016; Silva, Maas, Souza, & Gomes, 2017; Sodagar and Starkey, 2016; Tang, Ding, Li, & Li, 2019; Touchie and Pressnail, 2014; Wang, Xue, Ji, & Yu, 2018; Wang et al., 2015)
Relative Humidity	(Adaji, et al., 2019; Adekunle and Nikolopoulou, 2016; Baja, et al., 2019; Berge and Mathisen, 2016; Choi and Lee, 2018; Colclough, et al., 2018; Dabaieh and Johansson, 2018; Deuble and de Dear, 2014; Doctor-Pingel, et al., 2019; Filippín, et al., 2015; O. Guerra-Santin, et al., 2016; Gupta, et al., 2014; Gupta and Gregg, 2016; Gupta and Kapsali, 2016; Hua, et al., 2014; Ioannidis, et al., 2017; Jentsch, et al., 2017; Lee, et al., 2019; Liu, et al., 2018; Loyola, 2019; Naspi, et al., 2018; Park, et al., 2018; Pastore and Andersen, 2019; Patlakas, et al., 2017; Patlakas, et al., 2014; Pei, et al., 2015; Pretlove and Kade, 2016; Silva, et al., 2017; Sodagar and Starkey, 2016; Tang, et al., 2019; Touchie and Pressnail, 2014; Wang, et al., 2018; Wang, et al., 2015)
CO_2	(Berge and Mathisen, 2016; Choi and Lee, 2018; Colclough, et al., 2018; Dabaieh and Johansson, 2018; O. Guerra-Santin, et al., 2016; Gupta, et al., 2014; Gupta and Gregg, 2016; Gupta and Kapsali, 2016; Hua, et al., 2014; Ioannidis, et al., 2017; Lee, et al., 2019;

	Liu, et al., 2018; Naspi, et al., 2018; Park, et al., 2018; Pastore and Andersen, 2019; Silva, et al., 2017; Tang, et al., 2019; Wang, et al., 2015)
Total Volatile Organic Compounds (TVOC)	(Choi and Lee, 2018; Park, et al., 2018; Tang, et al., 2019)
Formaldehyde HCHO	(Lee, et al., 2019)
CO [ppm]	(Lee, et al., 2019; Park, et al., 2018)
Fine particles (PM 2.5)	(Choi and Lee, 2018; Lee, et al., 2019; Park, et al., 2018; Tang, et al., 2019)
Particulate matter PM 10	(Park, et al., 2018; Wang, et al., 2018)
Lighting levels (illuminance)	(Baja, et al., 2019; O. Guerra-Santin, et al., 2016; Hua, et al., 2014; Ioannidis, et al., 2017; Lee, et al., 2019; Liu, et al., 2018; Naspi, et al., 2018; Pastore and Andersen, 2019; Tang, et al., 2019; Wang, et al., 2015)
Sound pressure	(Choi and Lee, 2018; Tang, et al., 2019; Wang, et al., 2018; Wang, et al., 2015)
Air velocity	(Jentsch, et al., 2017; Lee, et al., 2019; Park, et al., 2018; Pei, et al., 2015; Silva, et al., 2017; Tang, et al., 2019; Wang, et al., 2018)
Outdoor Environment Data	
Temperature	(Barreca and Praticò, 2018; Colclough, et al., 2018; Dabaieh and Johansson, 2018; Doctor-Pingel, et al., 2019; Filippín, et al., 2015; O. Guerra-Santin, et al., 2016; Gupta, et al., 2014; Gupta and Gregg, 2016; Tang, et al., 2019; Touchie and Pressnail, 2014; Wang, et al., 2018)

Relative Humidity	(Barreca and Praticò, 2018; Colclough, et al., 2018; Dabaieh and Johansson, 2018; Filippín, et al., 2015; O. Guerra-Santin, et al., 2016; Gupta and Gregg, 2016; Tang, et al., 2019; Wang, et al., 2018)
Atmospheric pressure	(Colclough, et al., 2018; Doctor-Pingel, et al., 2019)
Wind direction and speed	(Doctor-Pingel, et al., 2019; Filippín, et al., 2015; O. Guerra-Santin, et al., 2016)

These highlight the type of data the model needs to cover at first instance.

QUESTIONNAIRE RESULTS

The questionnaire provided insights into the data collection, storage, and management practices of leading researchers.

The POE data collection process can present a number of issues. All (100%) researchers reported user-related issues with data collection, with user concerns about data privacy being a concern in 67% of cases. User interference was also an issue during the actual data logging process for 57% of researchers. Equipment reliability affected a relatively high number of researchers (47%), while the cost has been an issue for 52% of researchers. Overall this suggests that there are still challenges regarding the POE data collection at a large scale. However, the core issues appear to be around user concerns and behaviour and part of the greater discussion around data ownership.

Data retention appears to be standard practice: 100% of researchers have kept at least some of the POE data collected. As often with data, the possibility of future use is a major driver for keeping it: 76% of researchers reported their hope that they might find some use for it in the future, with 57% intending to publish something using it at a later date. In some (14%) cases, the data is retained simply because the owner has not deleted it: however, these researchers all had recent projects and this percentage might be higher for older projects.

Researchers store the data in a range of formats (Figure 3).

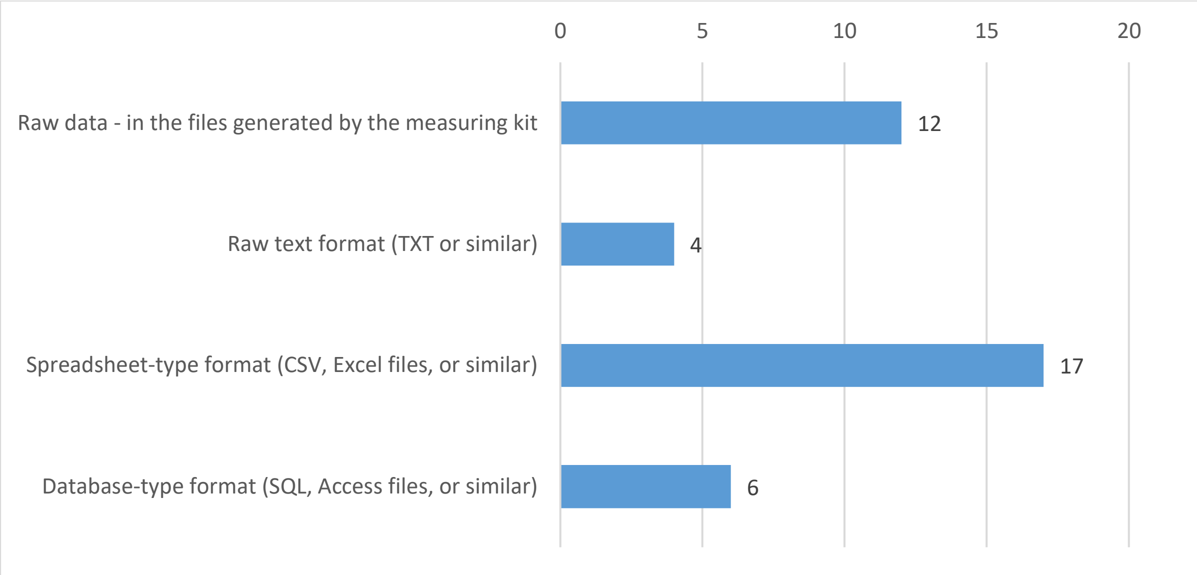

Figure 3. Formats in which researchers keep the POE data

It is interesting that the most popular (81%) format, however, is a spreadsheet type, typically the last medium in which the data was analysed. Only 19% of researchers store the data in raw text format: thus the data is vulnerable to compatibility issues across software versions (as in the case of Excel files) or areas (as in the case of CSV files). Similar issues can appear for raw data generated in proprietary formats connected to the measuring equipment.

The type of storage used by researchers is also of interest (Figure 4).

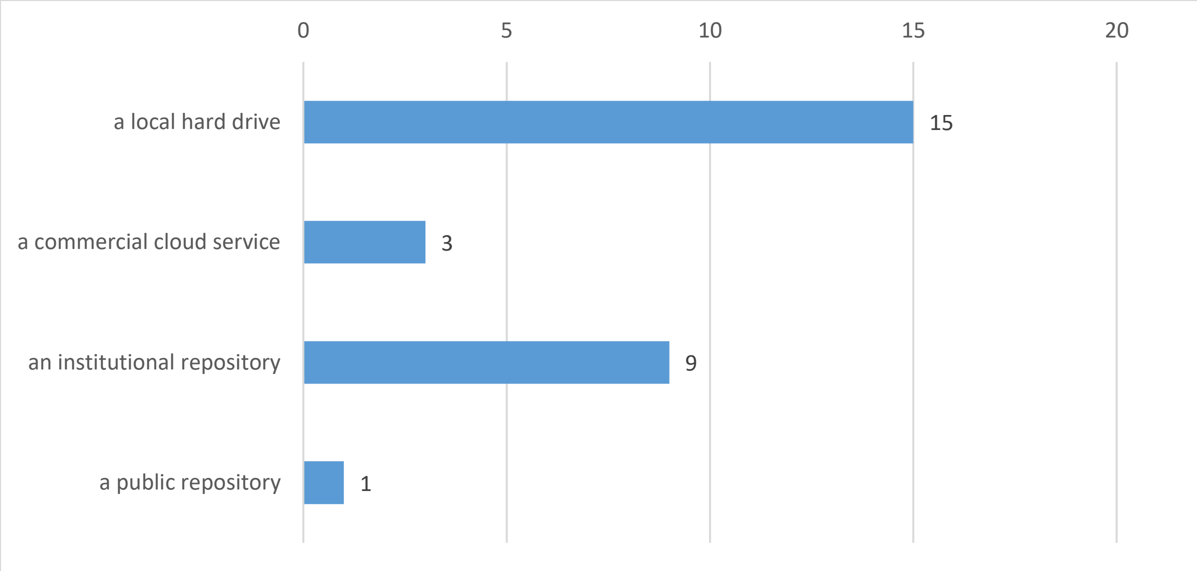

Figure 4. Storage types of POE data

Only 43% of respondents use an institutional repository, and more than half (51%) of researchers rely only on a local hard drive. Hence, almost all of the data are either subject to the access policies of an institution, or connected to a physical medium of limited lifespan, thus limiting significantly the retention options. The limited availability of data is extended further by the fact that the researcher responsible for collecting the data is often the sole person maintaining access: amongst the questionnaire respondents, this applied to 43% of cases. In 90% of cases, the only persons with access to data are the researcher and their group.

It is interesting also that 48% of researchers are unable or unwilling to share the data publicly. The most common concern cited is data privacy policies, typically set by the researchers' institutions. However, in some cases this might be a request of the building users/owners or simply the researchers considering some data to be personal in nature.

Finally, the value proposition of POE studies (i.e. the benefit they provide with respect to their cost) with the current technologies is generally viewed favourably by researchers, with 57% thinking the value proposition is excellent or good, and only 9.5% thinking of it as bad. However, given that these are typically environmental design specialists, the answers are perhaps less favourable that might have been expected. Figure 5 provides a clearer picture, showing a breakdown of the impact on the cost of individual aspects of a POE study.

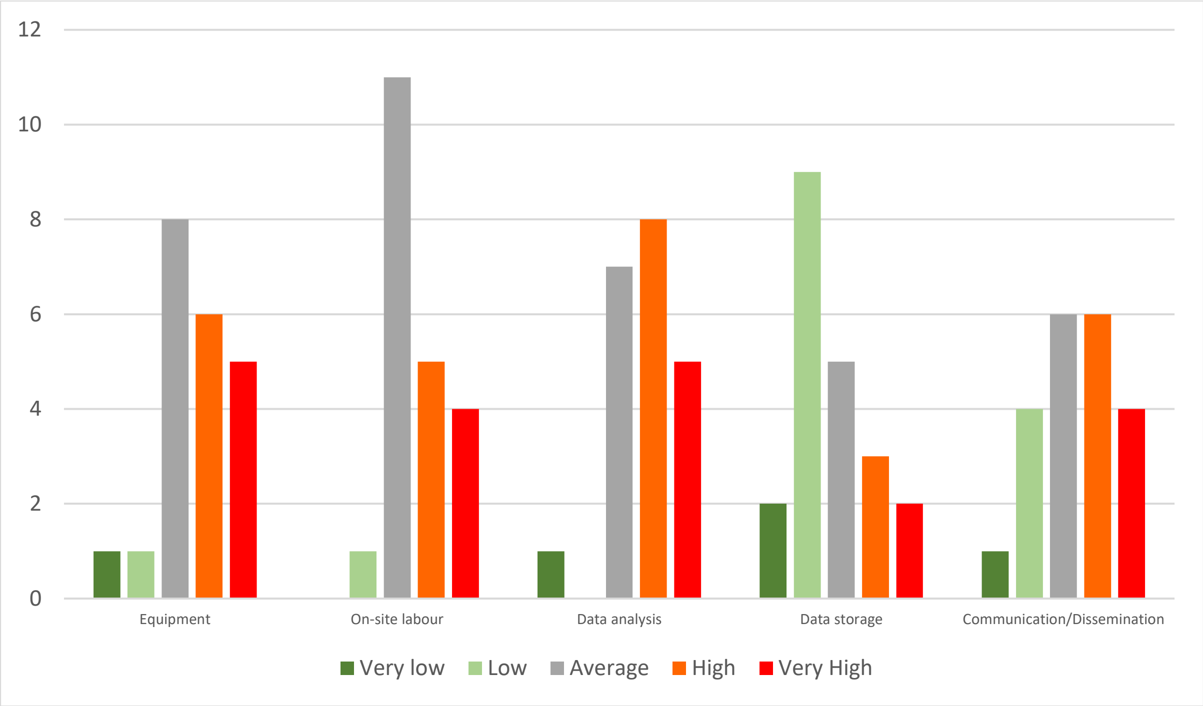

Figure 5. Impact on cost of different aspects of POE study.

The costs of the equipment and the on-site labour are to be expected, especially given the fact that these POE studies are typically ad hoc and for a specific project. What is perhaps more surprising is the impact on cost of the data analysis (62% rating it high or very high) and the communication and dissemination of results (48% rating it high or very high). Given that most of the researchers do not have a specific data management and retention strategy, it is also noteworthy that only 52% consider that the costs of the data storage is low or very low. Given the mass data collection that is typical in other fields, it is likely that the data types used by researchers has an impact on their evaluation of the cost.

DEVELOPMENT OF A DIGITAL CURATION PROTYPE MODEL

Model Requirements

The combination of the general DC principles, the literature review, and the responses collected by the questionnaire allows the development of a prototype model for environmental POE data.

A number of conflicting requirements need to be addressed in order to achieve this aim. The model needs to have low storage costs, allow easy cross-platform conversion, maximise compatibility, and minimise legacy issues across versions, while being extensible and customisable. The source data needs to be human-readable in order to allow the manual extraction and conversion if needed. Finally, the system needs to be format-, software-, and framework-independent. While there are benefits to having a system that contains as much information as possible, this comes with exponential overheads in data collection, homogenisation, processing, retention and management.

A case in point is the introduction of spatial parameters: superficially, requiring the inclusion of spatial information appears a straightforward demand with clear benefits. However, this comes with very substantial data collection overheads as the data collection mechanism needs to identify this location, which needs to be contextualised in a local coordinate system and linked to some type of universal coordinate system. However, this spatial information is unlikely to be of use independently without some information about the geometry and construction of a space. Thus this requires involvement with additional, highly complex, systems in multiple formats, which have their own major issues of compatibility, data storage etc. Therefore, the model presented here eschews such data, with the aim of remaining self-contained. It can be viewed as a *subset* of information systems that cover building construction, such as Building Information Modelling (BIM), and its structure should allow for this to be the case.

Set against the DC models and principles presented earlier, the prototype model should be:

- lightweight
- generic
- extensible
- human-readable
- machine-readable
- self-contained

- agnostic of software-specific formats and data structures

A Prototype Model for POE Data Curation

A prototype model was developed for POE data of an objective environmental nature, i.e. environmental properties that have quantifiable values, measured in scientific units. It is based on the Extensible Markup Language (XML), thus being human-readable, text-based, generic, extensible, and format agnostic. It also has the advantage of being a very widely used format, which can also be converted easily into other formats such as JSON, maximising ease of adoption.

Data Organisation and Structure

The data structure is organised around the concept of a building complex, made up of individual buildings, each containing groups of individual spaces. The root element can be either a building complex or an individual building. Data regarding external spaces are provided at a building level, with an optional subdivision for different locations of measurement. Data regarding internal spaces are provided at a space level. Data regarding energy can be provided at either a building or space level. Figures 6-8 show the data structure.

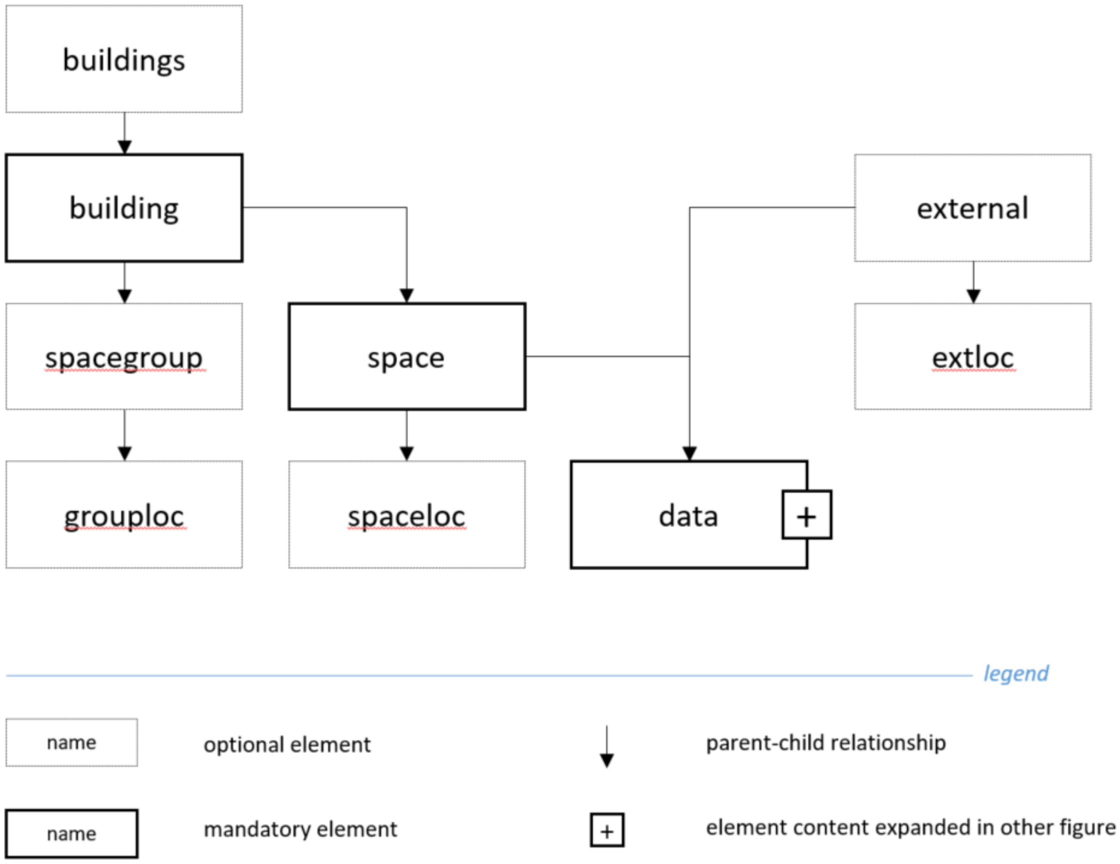

Figure 6. Space Data Structure

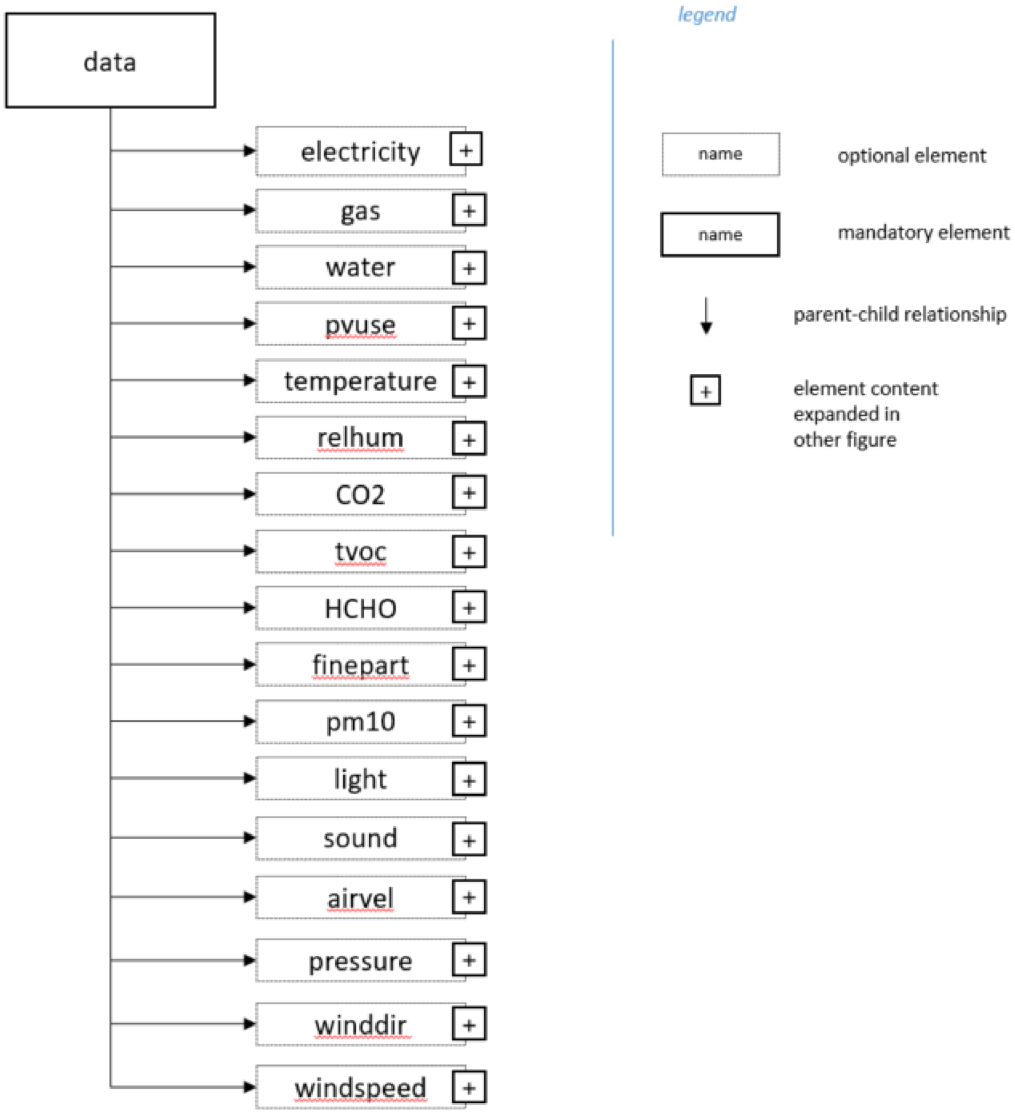

Figure 7. Monitored Data Structure

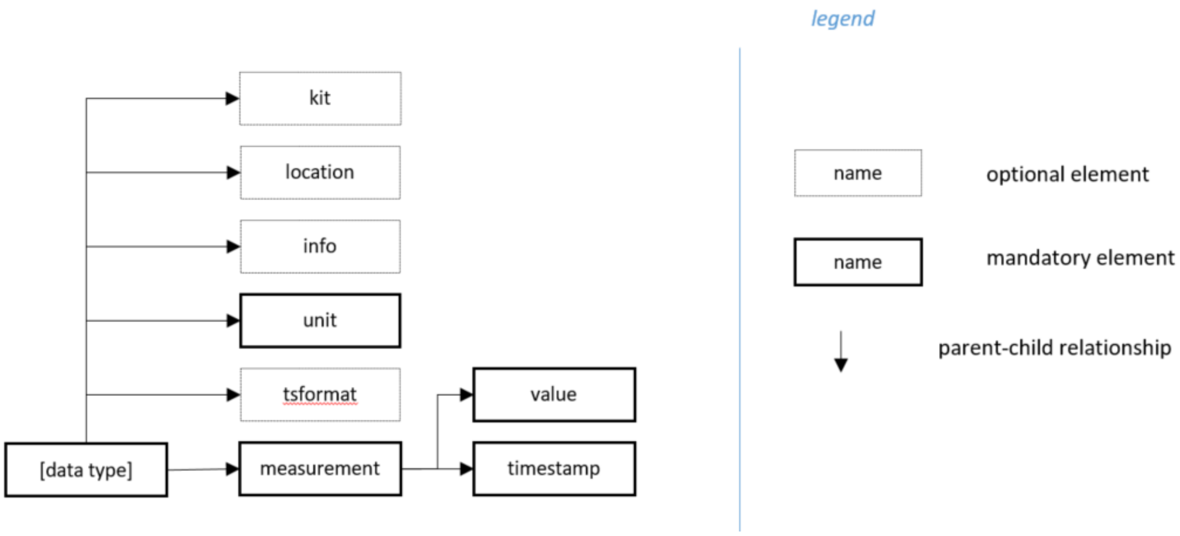

Figure 8. Internal Data Structure

Integration with BIM and BRICK

As the system is XML-based, integration within other frameworks is relatively straightforward. In BIM, the main complexity comes from the hierarchy of spaces, as different BIM systems will have different conceptual hierarchies. A conversion of the space hierarchy to Industry Foundation Classes (IFC) is included as supplementary material (Appendix I).

Similarly, each monitored element of the prototype model can be represented as a BRICK `Measurable` class. A mapping can be made between most data elements presented here (e.g. <temperature> as BRICK `Dry Bulb Temperature`), while similar relationships can be developed between location tags.

A building monitoring project was selected as a case study, in order to evaluate the effectiveness of the DC model presented here. The case study consisted of a residential building complex in the UK that was monitored over 1 year with Hobo data loggers. For the purposes of the case study, one building was selected, with 3 spaces monitored. An XML file containing the first 3,000 monitored data entries is included as supplementary material.

Model Validation

The model was validated against the Model Requirements described earlier. Specifically:

- lightweight, as the data is 307 KB compared to 839 KB of the source files (37% of file size)
- generic, as the data structure is in generic text format
- extensible, as the data structure is XML-based and thus extensible
- the data is both human- and machine-readable
- self-contained, as all building monitoring data is included
- XML, a software-agnostic, and easily convertible, format

The case study also demonstrates that the prototype model developed here supports the Digital Curation frameworks discussed above.

DISCUSSION

The advantages of implementing such a DC model for POE data are substantial. Firstly, it would allow the continuing use of data and development of long-term data series. The majority of questionnaire respondents (76%) said they plan to reuse the data. However, without a DRM strategy this is not feasible and/or cost effective in the long term and the responses suggest that such a strategy is not in place for most researchers.

A second advantage of a standardised DC model is that it would allow the sharing and cross-use of data. The questionnaire responses suggest that researchers are not generally unwilling to share the data, and it is mostly external policy constraints that prevent them from doing so. However, even if those policy constraints were addressed, POE data is currently collected and stored an ad hoc fashion. Effective sharing and reuse of such data is difficult. There is also a lack of software applications that allow the visualisation, analysis, and storage of such data. Where those exist, as in experimental projects (Patlakas, et al., 2014), the treatment is again ad hoc.

This is at a time where the collection of data is expected to increase in an exponential manner. The Internet of Things (IoT) will enable the collection of various types of data from multiple sources, and research projects are already demonstrating what is feasible (Amaxilatis, Akrivopoulos, Mylonas, & Chatzigiannakis, 2017). The potential of large-scale data to enable data savings is also beginning to be highlighted (Abrol, Mehmani, Kerman, Meinrenken, & Culligan, 2018). The continually-increasing scale of these datasets also means that new methods of analysis are required: techniques borrowed from computer science and machine learning facilitate handling such “big data” and gauging valuable insights (Geronazzo, Brager, & Manu, 2018).

A more theoretical aspect involves the beneficial effects large-scale datasets would have to our understanding of core aspects of building performance and, more generally, building physics. The typical cycle in science and engineering involves developing a hypothesis, testing it, and validating, rejecting, or adjusting the hypothesis accordingly. On the micro scale, this is generally achievable, e.g. identifying a single property of single material, or establishing the performance of a single component in a single aspect. However, the challenge becomes substantially greater in a complex system, such as the environmental performance of an entire building. Collection of data at a large scale, over long periods of time, together with the continually evolving methods of analysis in computing, appear to be the most suitable way to address this challenge. Those datasets, however, will require a framework for retention and management.

At the same time, a number of challenges remain, which act as an impediment to releasing the potential of large-scale POE analyses and thus allowing full and effective use of DC strategies such as those presented above. Questionnaire respondents were largely positive about the value proposition of POE: however, even these specialists often identified the cost of equipment and on-site labour as high or very high. Economies of scale can be expected as technology improves and POE becomes more popular: twenty years ago, few would have anticipated that a simple mobile phone can be used to collect data for the related field of structural monitoring (Noel et al., 2017). It is not unreasonable to expect that similar progress can be made for the collection of other types of data. However, some researchers also identified the cost of storage (24%), analysis (62%), and communication of results (48%) as high or very high. Compared with the sizes of datasets collected in other fields, POE is relatively modest. The main issue, then, is that building performance analysis based on monitored data has not benefited from the state-of-the-art in technology.

A major challenge is that, fifteen years after Bordass and Leaman's call for making POE routine (Bordass and Leaman, 2005), there is still a substantial lack of awareness of its techniques and benefits. The limited interest that does exist is concentrated in the industrial and commercial sector, with the residential sector relatively untouched, despite its construction volume and importance (Vásquez-Hernández and Restrepo Álvarez, 2017).

This lack of awareness partly explains two further obstacles to releasing the potential of large-scale, long-term POE studies. The first is that there is no policy support. Despite the increasing interest in sustainable buildings over the last two decades, accreditation schemes still operate primarily based on construction specification and not on monitored performance. Thus any policy support is biased towards pre-emptive component-based certification as opposed to whole-system performance in real conditions of use.

A final issue of particular importance, and a major obstacle to the effective application of such frameworks, is data ownership. This is apparent on three fronts. Firstly, building users are understandably reluctant to have data they consider personal, such as building performance, monitored and analysed. This is supported by the questionnaire findings, as two-thirds of respondents reported that user concerns about privacy interfered with data collection, a figure that is likely to be even higher in the case of residential monitoring.

Secondly, commercial organisations that acquire such data, are likely to have a vested interest in owning, retaining, mining, and monetizing such data, as well as aggressive policies to block sharing with third parties. The research presented here did not cover such organisations, but the example of other areas suggests that commercial entities are more interested in "closed silo" approaches and attempting to build data monopolies.

Thirdly, there are institutional and policy aspects. As questionnaire respondents work in large research-oriented organisations, institutional policies place strict controls on the retention, access, and analysis of data. On a more general scale, legal requirements currently appear set to increase. The

European Union’s General Data Protection Regulation (GDPR) ("Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46," 2016) is a leading example of the strictness of such regulations.

There are, however, good reasons to be optimistic. It is a modern paradox that, while data collection for scientific, technological, and other socially-beneficial purposes is more strictly controlled and regulated than ever, individuals have become accustomed to providing and sharing personal data on a constant basis. It is likely that in an IoT future, data sharing about the performance of “things” will be commonplace; buildings are no exception to this rule.

On the commercial side, while private organisations might aspire to data monopolies, the final decision remains with the State regulator. A good example is the field of Building Information Modelling: while software vendors might attempt format and platform monopolies, the introduction of IFC, and its adoption by some states as a required or recommended standard, meant that a common, open format was supported. While by no means perfect at the current stage, the existence of IFC suggests that, if the State supports a common format, commercial entities follow this lead.

Finally, the issue of legislation trends might be misleading. Research organisations, from which the questionnaire participants were recruited, tend to have very strict policies, as they push forward the state-of-the-art, and often encounter new grounds. However, there is an established legal and philosophical precedent of sharing performance data (and thus, some kind of personal information), when the greater good is served. Car sellers, for example, need to reveal data such as mileage and service checks; house buyers are encouraged (and in some cases required) to conduct surveys. There is no reason why a similar approach cannot be taken for long-term POE data.

Overall, the issue of data ownership is a complex question, touching not simply science and technology, but also philosophy and politics, and outside the focus of this work. A common framework, however, can contribute towards changing policy. At first instance, it is more likely to be applied in closed systems, where building users have a limited say in how this data is collected, such as work environments or State buildings. Once the benefits have been established, there is no inherent reason not to collect such data. What is certain, however, is that the issue of data ownership will be of major importance in all such efforts.

CONCLUSION

This work is the first, to our knowledge, to address the issue of Digital Curation in the field of Post-Occupancy Evaluation surveys that focus on objectively measured and monitored properties. Such surveys are a necessary component of the evaluation of building performance, providing datasets that allow the validation and verification of design assumptions. Advances in hardware and software, together with the increasing interest shown in sustainable buildings in the context of the climate emergency, mean that such studies are likely to increase in the future. This increase, however, means that researchers and practitioners will need to collect, store, manage, analyse, and disseminate very large datasets. The work presented here introduces DC concepts and provides a prototype model for this purpose.

This prototype can be a basis for releasing the potential of large-scale, long-term POE surveys. It can allow for cross-use, re-use, sharing, and pooling of data, as well as address issues of long-term management and storage. The increasing power of machine learning-based analytical tools enhances the argument for storing those. A limitation of this prototype is that, in order to achieve the DC aims, it focuses only on key properties relating to the conditions of a space. It does not engage with other aspects of POE, such as subjective user feedback, temporary elements (such as equipment), or aspects that would require linkage with building drawings and 3d models, (such as control systems). In common with most DC models, it should be viewed as a "made-to-measure" tool aiming to achieve a specific aim, and not as an all-encompassing system; the feasibility and applicability of the latter systems remain to be seen.

At the same time, there are major obstacles that need to be overcome so that DC strategies such as those presented here, and building monitoring more generally, can release their full potential. The industry awareness of the benefits of POE remains low, dropping to almost non-existent in the general population. Policy support is minimal. More importantly, collecting environmental building data on a large scale involves the greater issue of data privacy and protection, a complex, multi-faceted issue.

There is a number of actions that can be taken to address these issues and release the potential of DC for POE. Increasing data and information literacy in the Built Environment domain is a fundamental step: while Building Performance in the entire building lifecycle is established as a topic, it remains primarily based on tick box benchmarks and simulations. Increasing the validation and verification of these models is necessary; this comes with the associated data management issues described. A second aspect is considering this as a policy issue, and not a research or industry one. The application of common standards depends on State specification, or at least support: the experience of other areas of data management suggests that left unchecked commercial interests aim for monopolisation and monetization of data. Finally, it is important to consider Digital Curation as a core issue more generally, not only for environmental and building performance aspects, but for the whole of Architecture, Engineering, and Construction (AEC) industry.

REFERENCES

Abrol, S., Mehmani, A., Kerman, M., Meinrenken, C. J., & Culligan, P. J. (2018). Data-Enabled Building Energy Savings (D-E BES). [Article]. *Proceedings of the IEEE, 106*(4), pp. 661-679. doi:10.1109/JPROC.2018.2791405

Adaji, M. U., Adekunle, T. O., Watkins, R., & Adler, G. (2019). Indoor comfort and adaptation in low-income and middle-income residential buildings in a Nigerian city during a dry season. *Building and Environment, 162*, p 106276. doi:https://doi.org/10.1016/j.buildenv.2019.106276 Retrieved from http://www.sciencedirect.com/science/article/pii/S036013231930486X

Adekunle, T. O., & Nikolopoulou, M. (2016). Thermal comfort, summertime temperatures and overheating in prefabricated timber housing. [Article]. *Building and Environment, 103*, pp. 21-35. doi:10.1016/j.buildenv.2016.04.001 Retrieved from https://www.scopus.com/inward/record.uri?eid=2-s2.0-84962476829&doi=10.1016%2fj.buildenv.2016.04.001&partnerID=40&md5=569eb7f1d11c729ea653cdfddf16b002

Amaxilatis, D., Akrivopoulos, O., Mylonas, G., & Chatzigiannakis, I. (2017). An IoT-based solution for monitoring a fleet of educational buildings focusing on energy efficiency. [Article]. *Sensors (Switzerland), 17*(10)doi:10.3390/s17102296 Retrieved from https://www.scopus.com/inward/record.uri?eid=2-s2.0-85031279546&doi=10.3390%2fs17102296&partnerID=40&md5=9e078589f0a60a772c6e363a9243f8db

Ascione, F., Bianco, N., Böttcher, O., Kaltenbrunner, R., & Vanoli, G. P. (2016). Net zero-energy buildings in Germany: Design, model calibration and lessons learned from a case-study in Berlin. [Article]. *Energy and Buildings, 133*, pp. 688-710. doi:10.1016/j.enbuild.2016.10.019 Retrieved from https://www.scopus.com/inward/record.uri?eid=2-s2.0-84992603230&doi=10.1016%2fj.enbuild.2016.10.019&partnerID=40&md5=df0c854adb0ed33afd3823f9ad3c207c

Baja, F. D. F., Bajracharya, S., Freeman, M. A., Gray, A. J., Haglund, B. T., Kuipers, H. R., & Opatola, O. R. (2019). Leed gold but not equal: Two case study buildings. [Article]. *International Journal of Design and Nature and Ecodynamics, 14*(1), pp. 52-62. doi:10.2495/DNE-V14-N1-52-62 Retrieved from https://www.scopus.com/inward/record.uri?eid=2-s2.0-85061551387&doi=10.2495%2fDNE-V14-N1-52-62&partnerID=40&md5=12309e21aacc198ef1dc878be3c78de1

Barreca, F., & Praticò, P. (2018). Post-occupancy evaluation of buildings for sustainable agri-food production-A method applied to an olive oil mill. [Article]. *Buildings, 8*(7)doi:10.3390/buildings8070083 Retrieved from https://www.scopus.com/inward/record.uri?eid=2-s2.0-85050460531&doi=10.3390%2fbuildings8070083&partnerID=40&md5=fd3dc8a8ee06434911a3cba1b8363157

Berge, M., & Mathisen, H. M. (2016). Perceived and measured indoor climate conditions in high-performance residential buildings. [Article]. *Energy and Buildings, 127*, pp. 1057-1073. doi:10.1016/j.enbuild.2016.06.061 Retrieved from https://www.scopus.com/inward/record.uri?eid=2-s2.0-84978061541&doi=10.1016%2fj.enbuild.2016.06.061&partnerID=40&md5=2990dbeb3247bfd4dbfb15eb5ff0f849

Berners-Lee, T. F., M. (2008). *Weaving the Web: The Original Design and Ultimate Destiny of the World Wide Web by Its Inventor*: HarperCollins.

Bordass, B., & Leaman, A. (2005). Making feedback and post-occupancy evaluation routine 1: A portfolio of feedback techniques. *Building Research & Information, 33*(4), pp. 347-352.

doi:10.1080/09613210500162016 Retrieved from https://doi.org/10.1080/09613210500162016
buildingSMART. IfcSensor. Retrieved Date from https://standards.buildingsmart.org/IFC/RELEASE/IFC4/FINAL/HTML/.
Choi, J. H., & Lee, K. (2018). Investigation of the feasibility of POE methodology for a modern commercial office building. [Article]. *Building and Environment, 143*, pp. 591-604. doi:10.1016/j.buildenv.2018.07.049 Retrieved from https://www.scopus.com/inward/record.uri?eid=2-s2.0-85050875088&doi=10.1016%2fj.buildenv.2018.07.049&partnerID=40&md5=15c48d3776473293cf9662561c910491
Colclough, S., Kinnane, O., Hewitt, N., & Griffiths, P. (2018). Investigation of nZEB social housing built to the Passive House standard. [Article]. *Energy and Buildings, 179*, pp. 344-359. doi:10.1016/j.enbuild.2018.06.069 Retrieved from https://www.scopus.com/inward/record.uri?eid=2-s2.0-85054401213&doi=10.1016%2fj.enbuild.2018.06.069&partnerID=40&md5=4996ae09a5efb27cb19b2f292fc92bef
Construction, G. A. f. B. a., Agency, I. E., & Programme, U. N. E. (2019). *2019 global status report for buildings and construction: Towards a zero-emission, efficient and resilient buildings and construction sector.*
Dabaieh, M., & Johansson, E. (2018). Building Performance and Post Occupancy Evaluation for an off-grid low carbon and solar PV plus-energy powered building. A case from the Western Desert in Egypt. [Article]. *Journal of Building Engineering, 18*, pp. 418-428. doi:10.1016/j.jobe.2018.04.011 Retrieved from https://www.scopus.com/inward/record.uri?eid=2-s2.0-85046446115&doi=10.1016%2fj.jobe.2018.04.011&partnerID=40&md5=21d09375d446d450ef9e7a599d02cb25
De Wilde, P. (2014). The gap between predicted and measured energy performance of buildings: A framework for investigation. *Automation in Construction, 41*, pp. 40-49.
Deuble, M. P., & de Dear, R. J. (2014). Is it hot in here or is it just me? Validating the post-occupancy evaluation. [Article]. *Intelligent Buildings International, 6*(2), pp. 112-134. doi:10.1080/17508975.2014.883299 Retrieved from https://www.scopus.com/inward/record.uri?eid=2-s2.0-84900014056&doi=10.1080%2f17508975.2014.883299&partnerID=40&md5=a85d3d98339e26fc9b6b19ce6535bedc
Digital Curation Centre. What is digital curation? Retrieved Date from http://www.dcc.ac.uk/digital-curation/what-digital-curation.
Doctor-Pingel, M., Vardhan, V., Manu, S., Brager, G., & Rawal, R. (2019). A study of indoor thermal parameters for naturally ventilated occupied buildings in the warm-humid climate of southern India. *Building and Environment, 151*, pp. 1-14. doi:https://doi.org/10.1016/j.buildenv.2019.01.026 Retrieved from http://www.sciencedirect.com/science/article/pii/S0360132319300344
ESRC. UK Data Service. Retrieved Date from https://www.ukdataservice.ac.uk/.
Filippín, C., Larsen, S. F., & Marek, L. (2015). Experimental monitoring and post-occupancy evaluation of a non-domestic solar building in the central region of Argentina. [Article]. *Energy and Buildings, 92*, pp. 267-281. doi:10.1016/j.enbuild.2015.01.053 Retrieved from https://www.scopus.com/inward/record.uri?eid=2-s2.0-84923203908&doi=10.1016%2fj.enbuild.2015.01.053&partnerID=40&md5=2931add58ccf831680781f2929b86591
Geronazzo, A., Brager, G., & Manu, S. (2018). Making sense of building data: New analysis methods for understanding indoor climate. [Article]. *Building and Environment, 128*, pp. 260-271. doi:10.1016/j.buildenv.2017.11.030 Retrieved from

https://www.scopus.com/inward/record.uri?eid=2-s2.0-85037373527&doi=10.1016%2fj.buildenv.2017.11.030&partnerID=40&md5=86c9bfdff7be884696b32b31f327df8f

Gerrish, T., Ruikar, K., Cook, M., Johnson, M., & Phillip, M. (2017). Analysis of basic building performance data for identification of performance issues. [Article]. *facilities, 35*(13-14), pp. 801-817. doi:10.1108/F-01-2016-0003 Retrieved from https://www.scopus.com/inward/record.uri?eid=2-s2.0-85032962982&doi=10.1108%2fF-01-2016-0003&partnerID=40&md5=7c2f1fdc9407d6eaf5c1569f679990e4

Girard, J. G., J. (2015). Defining knowledge management: Toward an applied compendium *Online Journal of Applied Knowledge Management, 3*(1), pp. 1-20.

Gray, J., Gerlitz, C., & Bounegru, L. (2018). Data infrastructure literacy. *Big Data & Society, 5*(2), p 2053951718786316. doi:10.1177/2053951718786316 Retrieved from https://journals.sagepub.com/doi/abs/10.1177/2053951718786316

Guerra-Santin, O., Romero Herrera, N., Cuerda, E., & Keyson, D. (2016). Mixed methods approach to determine occupants' behaviour – Analysis of two case studies. [Article]. *Energy and Buildings, 130*, pp. 546-566. doi:10.1016/j.enbuild.2016.08.084 Retrieved from https://www.scopus.com/inward/record.uri?eid=2-s2.0-84985011810&doi=10.1016%2fj.enbuild.2016.08.084&partnerID=40&md5=97151bbbe1da227d86fdcfbc8813cf3c

Guerra-Santin, O., & Tweed, C. A. (2015). In-use monitoring of buildings: An overview of data collection methods. *Energy and Buildings, 93*, pp. 189-207. doi:https://doi.org/10.1016/j.enbuild.2015.02.042 Retrieved from http://www.sciencedirect.com/science/article/pii/S0378778815001474

Gupta, R., Barnfield, L., & Hipwood, T. (2014). Impacts of community-led energy retrofitting of owner-occupied dwellings. [Article]. *Building Research and Information, 42*(4), pp. 446-461. doi:10.1080/09613218.2014.894742 Retrieved from https://www.scopus.com/inward/record.uri?eid=2-s2.0-84901199201&doi=10.1080%2f09613218.2014.894742&partnerID=40&md5=313b8bae42f067b3cb3f8283e3012cab

Gupta, R., & Gregg, M. (2016). Do deep low carbon domestic retrofits actually work? [Article]. *Energy and Buildings, 129*, pp. 330-343. doi:10.1016/j.enbuild.2016.08.010 Retrieved from https://www.scopus.com/inward/record.uri?eid=2-s2.0-84984838229&doi=10.1016%2fj.enbuild.2016.08.010&partnerID=40&md5=74034d36ad65ba1688cddc8a9439c0fe

Gupta, R., & Kapsali, M. (2016). Empirical assessment of indoor air quality and overheating in low-carbon social housing dwellings in England, UK. [Article]. *Advances in Building Energy Research, 10*(1), pp. 46-68. doi:10.1080/17512549.2015.1014843 Retrieved from https://www.scopus.com/inward/record.uri?eid=2-s2.0-84923554587&doi=10.1080%2f17512549.2015.1014843&partnerID=40&md5=1d79cb1fc804ff3048d0fc7eb3af2b9a

Hendler, J., & Berners-Lee, T. (2010). From the Semantic Web to social machines: A research challenge for AI on the World Wide Web. *Artificial intelligence, 174*(2), pp. 156-161.

Higgins, S. (2008). The DCC curation lifecycle model. *International journal of digital curation, 3*(1)

Hua, Y., Göçer, T., & Göçer, K. (2014). Spatial mapping of occupant satisfaction and indoor environment quality in a LEED platinum campus building. [Article]. *Building and Environment, 79*, pp. 124-137. doi:10.1016/j.buildenv.2014.04.029 Retrieved from https://www.scopus.com/inward/record.uri?eid=2-s2.0-84901310477&doi=10.1016%2fj.buildenv.2014.04.029&partnerID=40&md5=c65c79e1ea63035d152f567fd820d730

Huebner, G. M., & Mahdavi, A. (2019). A structured open data collection on occupant behaviour in buildings. *Scientific data, 6*(1), pp. 1-4.

Initiative, F. Fair Principles: Go Fair. Retrieved Date from https://www.go-fair.org/fair-principles/.
Ioannidis, D., Zikos, S., Krinidis, S., Tryferidis, A., Tzovaras, D., & Likothanassis, S. (2017). Occupancy-driven facility management and building performance analysis. [Article]. *International Journal of Sustainable Development and Planning, 12*(7), pp. 1155-1167. doi:10.2495/SDP-V12-N7-1155-1167 Retrieved from https://www.scopus.com/inward/record.uri?eid=2-s2.0-85021307165&doi=10.2495%2fSDP-V12-N7-1155-1167&partnerID=40&md5=7fadbb397bf7d8ccfbdd1377c6bc74af
Jentsch, M. F., Kulle, C., Bode, T., Pauer, T., Osburg, A., Tenzin, . . . Tenzin, K. (2017). Field study of the building physics properties of common building types in the Inner Himalayan valleys of Bhutan. [Article]. *Energy for Sustainable Development, 38*, pp. 48-66. doi:10.1016/j.esd.2017.03.001 Retrieved from https://www.scopus.com/inward/record.uri?eid=2-s2.0-85016636145&doi=10.1016%2fj.esd.2017.03.001&partnerID=40&md5=143a2db4fa94aca1759f8910946e8749
Lab, M. M. Observatory for Economic Complexity. Retrieved Date from https://oec.world/en/.
Laughton, P., & Du Plessis, T. (2013). Data curation in the World Data System: proposed framework. *Data Science Journal, 12*, pp. 56-70.
Lavoie, B. F. (2004). The open archival information system reference model: Introductory guide. *Microform & imaging review, 33*(2), pp. 68-81.
Lee, J. Y., Wargocki, P., Chan, Y. H., Chen, L., & Tham, K. W. (2019). Indoor environmental quality, occupant satisfaction, and acute building-related health symptoms in Green Mark-certified compared with non-certified office buildings. [Article]. *Indoor Air, 29*(1), pp. 112-129. doi:10.1111/ina.12515 Retrieved from https://www.scopus.com/inward/record.uri?eid=2-s2.0-85056733970&doi=10.1111%2fina.12515&partnerID=40&md5=0a236d862facc8dfa8d765f995d821a2
Li, P., Froese, T. M., & Brager, G. (2018). Post-occupancy evaluation: State-of-the-art analysis and state-of-the-practice review. *Building and Environment, 133*, pp. 187-202. doi:https://doi.org/10.1016/j.buildenv.2018.02.024 Retrieved from http://www.sciencedirect.com/science/article/pii/S0360132318300957
Liu, Y., Wang, Z., Zhang, Z., Hong, J., & Lin, B. (2018). Investigation on the Indoor Environment Quality of health care facilities in China. [Article]. *Building and Environment, 141*, pp. 273-287. doi:10.1016/j.buildenv.2018.05.054 Retrieved from https://www.scopus.com/inward/record.uri?eid=2-s2.0-85048158308&doi=10.1016%2fj.buildenv.2018.05.054&partnerID=40&md5=2a8be96ba40046dc153a779c81b351db
Lord, P. M., A. (2003). *E-Science Curation Report. Data Curation for e-Science in the UK: An Audit to Establish Requirements for Future Curation and Provision*. T. J. C. f. t. S. o. Research. https://web.archive.org/web/20050225214807/http://www.jisc.ac.uk/uploaded_documents/e-ScienceReportFinal.pdf
Loyola, M. (2019). A method for real-time error detection in low-cost environmental sensors data. *Smart and Sustainable Built Environment, 8*(4), pp. 338-350. doi:10.1108/SASBE-10-2018-0051 Retrieved from https://doi.org/10.1108/SASBE-10-2018-0051
Mahdavi, A., & Taheri, M. (2017). An ontology for building monitoring. *Journal of Building Performance Simulation, 10*(5-6), pp. 499-508.
Menezes, A. C., Cripps, A., Bouchlaghem, D., & Buswell, R. (2012). Predicted vs. actual energy performance of non-domestic buildings: Using post-occupancy evaluation data to reduce the performance gap. *Applied energy, 97*, pp. 355-364.
Naspi, F., Arnesano, M., Stazi, F., D'Orazio, M., & Revel, G. M. (2018). Measuring occupants' behaviour for buildings' dynamic cosimulation. [Article]. *Journal of Sensors, 2018*doi:10.1155/2018/2756542 Retrieved from

https://www.scopus.com/inward/record.uri?eid=2-s2.0-85058285232&doi=10.1155%2f2018%2f2756542&partnerID=40&md5=60c8ab8a6e10afd9365eee5b05e0497f
NIH. (2020). GenBank: the NIH genetic sequence database. Retrieved Date from https://www.ncbi.nlm.nih.gov/genbank/.
Noel, A. B., Abdaoui, A., Elfouly, T., Ahmed, M. H., Badawy, A., & Shehata, M. S. (2017). Structural Health Monitoring Using Wireless Sensor Networks: A Comprehensive Survey. *IEEE Communications Surveys & Tutorials, 19*(3), pp. 1403-1423. doi:10.1109/COMST.2017.2691551
Park, J., Loftness, V., & Aziz, A. (2018). Post-occupancy evaluation and IEQ measurements from 64 office buildings: Critical factors and thresholds for user satisfaction on thermal quality. [Article]. *Buildings, 8*(11)doi:10.3390/buildings8110156 Retrieved from https://www.scopus.com/inward/record.uri?eid=2-s2.0-85056584580&doi=10.3390%2fbuildings8110156&partnerID=40&md5=95f4914ef986aea771381cb5f11e3d70
Pastore, L., & Andersen, M. (2019). Building energy certification versus user satisfaction with the indoor environment: Findings from a multi-site post-occupancy evaluation (POE) in Switzerland. *Building and Environment, 150*, pp. 60-74. doi:https://doi.org/10.1016/j.buildenv.2019.01.001 Retrieved from http://www.sciencedirect.com/science/article/pii/S0360132319300010
Patlakas, P., Koronaios, G., Raslan, R., Neighbour, G., & Altan, H. (2017). Case studies of environmental visualization. [Article]. *Energies, 10*(10)doi:10.3390/en10101459 Retrieved from https://www.scopus.com/inward/record.uri?eid=2-s2.0-85044381601&doi=10.3390%2fen10101459&partnerID=40&md5=071d6a7e914212dd566ce32ea964185c
Patlakas, P., Santacruz, H. B., & Altan, H. (2014). Visualising the environmental conditions of buildings. [Article]. *Proceedings of the Institution of Civil Engineers: Civil Engineering, 167*(5), pp. 56-64. doi:10.1680/cien.13.00014 Retrieved from https://www.scopus.com/inward/record.uri?eid=2-s2.0-84901774662&doi=10.1680%2fcien.13.00014&partnerID=40&md5=f2f1ae7d33a78a7a41f518e6c8332874
Pei, Z., Lin, B., Liu, Y., & Zhu, Y. (2015). Comparative study on the indoor environment quality of green office buildings in China with a long-term field measurement and investigation. [Article]. *Building and Environment, 84*, pp. 80-88. doi:10.1016/j.buildenv.2014.10.015 Retrieved from https://www.scopus.com/inward/record.uri?eid=2-s2.0-84910636935&doi=10.1016%2fj.buildenv.2014.10.015&partnerID=40&md5=0ce89751c0902ce95e5ff734560627ca
Poole, A. H. (2016). The conceptual landscape of digital curation. *Journal of Documentation*
Preiser, W. F., White, E., & Rabinowitz, H. (2015). *Post-Occupancy Evaluation (Routledge Revivals)*: Routledge.
Preiser, W. F. E. (2005). Building Performance Assessment—From POE to BPE, A Personal Perspective. *Architectural Science Review, 48*(3), pp. 201-204. doi:10.3763/asre.2005.4826 Retrieved from https://doi.org/10.3763/asre.2005.4826
Pretlove, S., & Kade, S. (2016). Post occupancy evaluation of social housing designed and built to Code for Sustainable Homes levels 3, 4 and 5. [Article]. *Energy and Buildings, 110*, pp. 120-134. doi:10.1016/j.enbuild.2015.10.014 Retrieved from https://www.scopus.com/inward/record.uri?eid=2-s2.0-84946887281&doi=10.1016%2fj.enbuild.2015.10.014&partnerID=40&md5=f56770abb2196a11e132ea7f9103bf94
Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free

movement of such data, and repealing Directive 95/46. (2016). *Official Journal of the European Union (OJ), 59*(1-88), p 294.

Silva, M. F., Maas, S., Souza, H. A. D., & Gomes, A. P. (2017). Post-occupancy evaluation of residential buildings in Luxembourg with centralized and decentralized ventilation systems, focusing on indoor air quality (IAQ). Assessment by questionnaires and physical measurements. [Article]. *Energy and Buildings, 148*, pp. 119-127. doi:10.1016/j.enbuild.2017.04.049 Retrieved from https://www.scopus.com/inward/record.uri?eid=2-s2.0-85019442795&doi=10.1016%2fj.enbuild.2017.04.049&partnerID=40&md5=b86c442989c2717834f054d9b015baa4

Sodagar, B., & Starkey, D. (2016). The monitored performance of four social houses certified to the Code for Sustainable Homes Level 5. [Article]. *Energy and Buildings, 110*, pp. 245-256. doi:10.1016/j.enbuild.2015.11.016 Retrieved from https://www.scopus.com/inward/record.uri?eid=2-s2.0-84947774302&doi=10.1016%2fj.enbuild.2015.11.016&partnerID=40&md5=ffcbd6e91292158c01246a5efb8a30b6

Tang, H., Ding, J., Li, C., & Li, J. (2019). A field study on indoor environment quality of Chinese inpatient buildings in a hot and humid region. *Building and Environment, 151*, pp. 156-167. doi:https://doi.org/10.1016/j.buildenv.2019.01.046 Retrieved from http://www.sciencedirect.com/science/article/pii/S0360132319300824

Touchie, M. F., & Pressnail, K. D. (2014). Using suite energy-use and interior condition data to improve energy modeling of a 1960s MURB. [Article]. *Energy and Buildings, 80*, pp. 184-194. doi:10.1016/j.enbuild.2014.05.014 Retrieved from https://www.scopus.com/inward/record.uri?eid=2-s2.0-84902679651&doi=10.1016%2fj.enbuild.2014.05.014&partnerID=40&md5=d5e79c4a0bd533bc34d69c9058a72024

Vásquez-Hernández, A., & Restrepo Álvarez, M. F. (2017). Evaluation of buildings in real conditions of use: Current situation. *Journal of Building Engineering, 12*, pp. 26-36. doi:https://doi.org/10.1016/j.jobe.2017.04.019 Retrieved from http://www.sciencedirect.com/science/article/pii/S2352710216302674

Wang, Z., Xue, Q., Ji, Y., & Yu, Z. (2018). Indoor environment quality in a low-energy residential building in winter in Harbin. [Article]. *Building and Environment, 135*, pp. 194-201. doi:10.1016/j.buildenv.2018.03.012 Retrieved from https://www.scopus.com/inward/record.uri?eid=2-s2.0-85043770297&doi=10.1016%2fj.buildenv.2018.03.012&partnerID=40&md5=db609a95897666cd9e887ceac94b3a69

Wang, Z., Zhao, H., Lin, B., Zhu, Y., Ouyang, Q., & Yu, J. (2015). Investigation of indoor environment quality of Chinese large-hub airport terminal buildings through longitudinal field measurement and subjective survey. [Article]. *Building and Environment, 94*, pp. 593-605. doi:10.1016/j.buildenv.2015.10.014 Retrieved from https://www.scopus.com/inward/record.uri?eid=2-s2.0-84946711254&doi=10.1016%2fj.buildenv.2015.10.014&partnerID=40&md5=81fd23c90cdea06644c60ce7ee30c723

Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J., Appleton, G., Axton, M., Baak, A., . . . Bourne, P. E. (2016). The FAIR Guiding Principles for scientific data management and stewardship. *Scientific data, 3*

Yu, C., Du, J., & Pan, W. (2019). Improving accuracy in building energy simulation via evaluating occupant behaviors: A case study in Hong Kong. *Energy and Buildings, 202*, p 109373. doi:https://doi.org/10.1016/j.enbuild.2019.109373 Retrieved from http://www.sciencedirect.com/science/article/pii/S0378778819315488

Zimmerman, A., & Martin, M. (2001). Post-occupancy evaluation: benefits and barriers. *Building Research & Information, 29*(2), pp. 168-174. doi:10.1080/09613210010016857 Retrieved from http://dx.doi.org/10.1080/09613210010016857

Zimring, C. M., & Reizenstein, J. E. (1980). Post-occupancy evaluation: An overview. *Environment and Behavior, 12*(4), pp. 429-450.