CORA Cork Open Research Archive
Cartlann Taighde Oscailte Chorcaí

Title	A review of data interoperability as a barrier to non-residential building energy policy: Evidence from Ireland
Authors	Mitra, Ankur;Bhalla, Kriti;Ahern, Ciara
Publication date	2026-09-15
Original Citation	Mitra, A., Bhalla, K. and Ahern, C. (2026) "A review of data interoperability as a barrier to non-residential building energy policy: Evidence from Ireland," Energy and Buildings, 367, (23pp), https://doi.org/10.1016/j.enbuild.2026.117734.
Type of publication	Article (peer-reviewed);journal-article
Link to publisher's version	10.1016/j.enbuild.2026.117734
Rights	© 2026, authors. Published by Elsevier B.V. This is an open access article under the CC BY license - http://creativecommons.org/licenses/by/4.0/
Download date	2026-11-10 23:50:49
Item downloaded from	https://hdl.handle.net/10468/19575



UCC
University College Cork, Ireland
Coláiste na hOllscoile Corcaigh

Energy & Buildings 367 (2026) 117734

Contents lists available at ScienceDirect

Energy & Buildings

journal homepage: www.elsevier.com/locate/enb

A review of data interoperability as a barrier to non-residential building energy policy: Evidence from Ireland

Ankur Mitra [a,b,*], Kriti Bhalla [a,b], Ciara Ahern [a,b]

[a] Irish Building Stock Observatory, Built Environment Research and Innovation Centre, Technological University Dublin, Dublin, Ireland
[b] MaREI, Research Ireland Centre for Energy, Climate and Marine, University College Cork, Cork, Ireland

ARTICLE INFO

Keywords:
Commercial
Energy Performance of Building Directive (EPBD)
Articles 16 and 22
Building stock energy modelling
Building data governance
Energy Performance Certificates (EPCs)
Reference building

ABSTRACT

Non-residential buildings account for a significant share of energy use and greenhouse gas emissions across the EU, yet evidence-based policy development remains hindered by fragmented and non-interoperable building data systems. This study evaluates the suitability of publicly available electronic datasets for characterising Ireland's non-residential building stock and supporting bottom-up energy modelling in the context of the recast Energy Performance of Buildings Directive (EPBD 2024/1275/EU).

A systematic scoping review of national government and institutional datasets was conducted using an adapted PRISMA-ScR approach. The analysis identified the national Energy Performance Certificate database and GeoDirectory as the only datasets with sufficient scale and detail to support national non-residential stock analysis. However, substantial limitations were identified, including incompatibilities between economic activity classifications (NACE) and building energy typologies (NEAP), incomplete stock representation, and EPC-triggered sampling bias. Frequently transacted building categories such as retail and warehouses are over-represented, while public-sector, healthcare, recreational, and small service buildings remain under-represented.

The findings demonstrate that the principal barrier to effective non-residential stock modelling in Ireland is not data availability, but deficiencies in data governance, interoperability, standardised building identifiers, and energy-relevant classification systems. Addressing these structural limitations will be critical for effective implementation of the EPBD and for developing reliable national building stock models. While focused on Ireland, the methodological approach and identified challenges are transferable to other EU member states.

1. Introduction

As shown in Fig. 1, the building and construction sector remains one of the largest sources of global greenhouse gas (GHG) emissions, accounting for approximately 37% of total global emissions in 2020 [1]. Around 10% of these emissions are associated with construction and demolition (embodied emissions), while 27% arise from building operations (operational emissions). Within operational emissions, non-residential buildings contribute about 10%, and residential buildings approximately 17% [1].

Mitigating emissions from the built environment is central to meeting the climate and energy objectives of the EU Green Deal and the Energy Performance of Buildings Directive (EPBD) [2,3]. Improving the energy performance of existing buildings, primarily through retrofit, has been identified as a critical strategy for achieving these goals [1].

1.1. Policy context

Fig. 2 is an excerpt from the Energy Performance of Buildings Directive (EPBD) that presents a "possible architecture of an energy performance of building database" within Europe [4,5]. The figure reflects the recast EPBD's vision for a digitally integrated building governance framework, whereby interoperable national databases combine Energy Performance Certificate (EPCs), Building Renovation Passports (BRPs), HVAC inspection reports, Smart Readiness Indicators (SRIs), and administrative datasets into machine-readable systems linked to the EU Building Stock Observatory (BSO). This architecture underpins the EU's transition toward evidence-based and data-driven building decarbonisation policy through enhanced interoperability, monitoring, and policy evaluation capabilities under Articles 16 and 22 [4,5] of the EPBD. Referring to Fig. 2;

* Corresponding author at: TU Dublin, Bolton Street, Rotunda, Dublin 1, D01K822, Ireland.
E-mail address: ankur.mitra@tudublin.ie (A. Mitra).

https://doi.org/10.1016/j.enbuild.2026.117734
Received 14 April 2026; Received in revised form 14 May 2026; Accepted 28 May 2026
Available online 4 June 2026
0378-7788/© 2026 The Author(s). Published by Elsevier B.V. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

Nomenclature

BAMB	Building As Material Banks
BER	Building Energy Rating
BRP	Building Renovation Passport
BSEM	Building Stock Energy Modelling
BSO	Building Stock Observatory
CIF	Construction Industry Federation
CPSS	Commercial Property Statistical System
CSO	Central Statistics Office
DETE	Department of Enterprise, Trade and Employment
DGNB	Deutsche Gesellschaft für Nachhaltiges Bauen
DRI	Digital Repository of Ireland
EPBD	Energy Performance of Buildings Directive
EPC	Energy Performance Certificate
EU	European Union
EUBUCCO	European Union Building stock Characteristics in a Common and Open database
EV	Essential Variables
GHG	Greenhouse Gas
GIS	Geographic Information System
GWP	Global Warming Potential
HVAC	Heating, Ventilation and Air Conditioning
IBSO	Irish Building Stock Observatory
IGBC	Irish Green Building Council
JBI	Joanna Briggs Institute
LDA	Land Development Agency
MS	Member State
NACE	Nomenclature of Economic Activities
NEAP	Non-Domestic Energy Assessment Procedure
NGO	Non-Governmental Organisation
openDBL	Open Digital Building Logbook
PCC	Population-Concept-Context
PRISMA	Preferred Reporting Items for Systematic Reviews and Meta-Analyses
PRISMA-ScR	PRISMA for Scoping Reviews
PSRA	Property Service Regulatory Authority
RB	Reference Building
RTB	Residential Tenancies Board
SEAI	Sustainable Energy Authority of Ireland
SRI	Smart Readiness Indicator

- Article 16 [4] strengthens the role of Energy Performance Certificates (EPCs) by promoting their integration with BRPs, enabling EPC-based performance assessments to directly inform staged renovation planning.
- Article 22 [5];
  - o obliges Member States (MSs) to establish integrated national building data repositories that host EPCs, BRPs, SRI data, inspection reports and both calculated and metered energy consumption. These databases must support machine-readable and standardised formats, to ensure interoperability with other administrative systems including digital building logbooks, land registries and Geographic Information Systems (GIS).
  - o critically requires MSs to ensure that building owners, tenants and facility managers have direct access to their buildings' systems data, including technical building system performance, metering data, control and automation systems, building system installations and, where applicable, electric vehicle charging infrastructure.
  - o permits for third-party access—for example, for financial institutions, accredited service providers or researchers— under transparent, non-discriminatory conditions, with the consent of the building owner.

    - Local authorities are to be granted access to relevant building data within their jurisdiction via these national databases to inform local renovation planning, financing and monitoring.

    - Additionally, aggregated and anonymised datasets—such as the share of buildings covered by EPCs, energy consumption statistics or life-cycle global warming potential (GWP), where available—must be published at least twice per year.

Together, these provisions establish a harmonised, accessible and data-rich digital infrastructure that underpins building stock energy modelling (BSEM) leading to evidence-based policymaking, financing mechanisms and large-scale renovation strategies across the EU. MSs must have one single or interconnected national databases in place by May 2026 and transfer data (manually or automatically) to the EU Building Stock Observatory (BSO) by March 2027 [6].

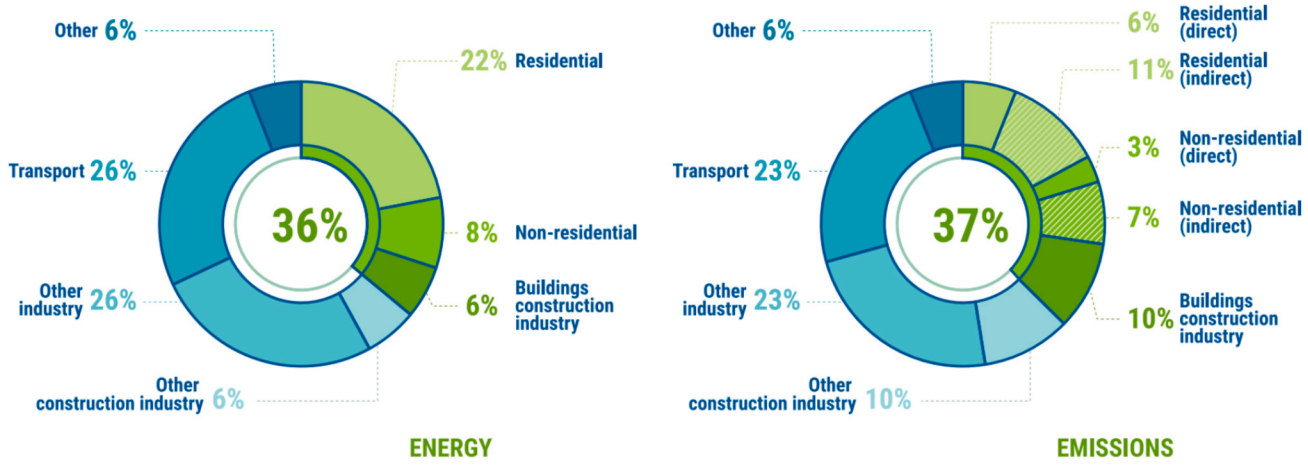


Fig. 1. Energy consumption by sector [1].

1.2. Building energy stock modelling

Building Stock Energy Models (BSEMs) for building stock are used to assess energy consumption patterns and implement targeted retrofit policies to the building stock. Bottom-up BSEMs are widely used to estimate emissions, evaluate retrofit measures, and support policy development [7]. These models depend on detailed and representative building data. Because modelling each individual building is computationally intensive, national building stocks are typically represented through Reference Buildings (RBs) or archetypes [8]. As shown in Fig. 3, RBs represent groups of buildings with comparable characteristics, enabling aggregated analysis of energy performance and retrofit potential [7].

As illustrated in Fig. 4 and established by Corgnati et al. [9], hereafter referred to as the 'Cognati RB Framework'; Reference Buildings (RBs) are characterised using data drawn from multiple sources and organised across four key domains: form (e.g., geometry, building type, size), envelope (e.g., construction materials, thermo-physical properties), system (e.g., heating, cooling, ventilation), and operation (e.g., occupancy patterns, usage schedules).

1.3. EU building data initiatives: applicability to non-residential building stock energy modelling

Several European-level initiatives aim to improve building-stock data integration and interoperability, though their applicability to non-residential building stock energy modelling (BSEM) remains limited. The EU Building Stock Characteristics in a Common and Open database (EUBUCCO v0.1) compiles building-footprint data for approximately 202 million buildings across Europe harmonising limited physical attributes such as footprint geometry, height, construction year, and building type. However, the database does not contain energy performance or consumption data, and attribute completeness varies substantially by country [10]. For Ireland, the coverage is uneven, reporting 1,610,614 buildings from OpenStreetMap [11], containing building-use classification (67%), height information (13%), and no construction-year data, limiting its utility for energy modelling.

Residential-focused typology initiatives, including the EPISCOPE project [12] and its TABULA framework [13], provide structured archetype classifications based on construction period and building type. While these approaches demonstrate the value of harmonised

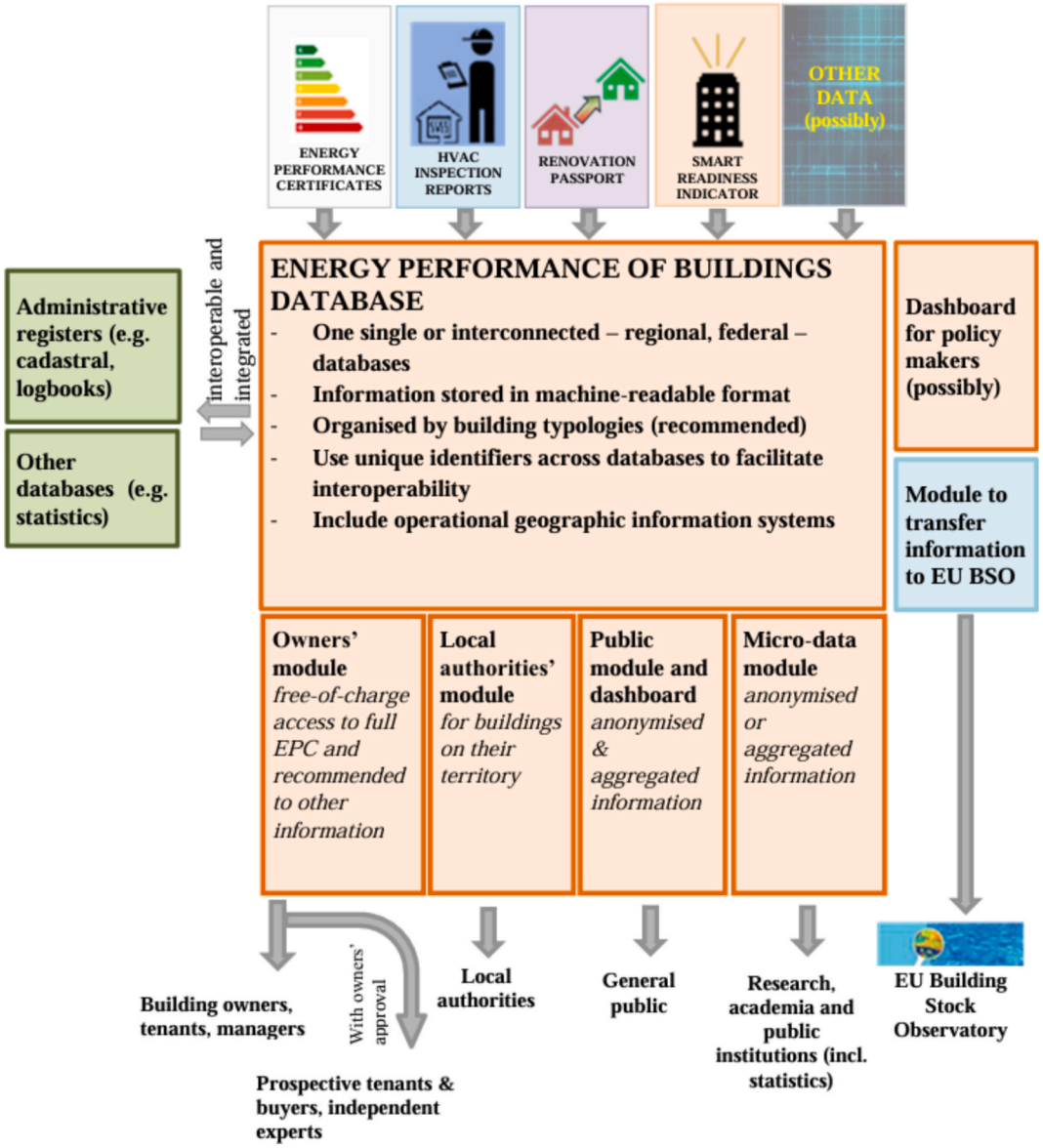


Fig. 2. Possible architecture of an energy performance of building databases: .
Source EPBD [5]

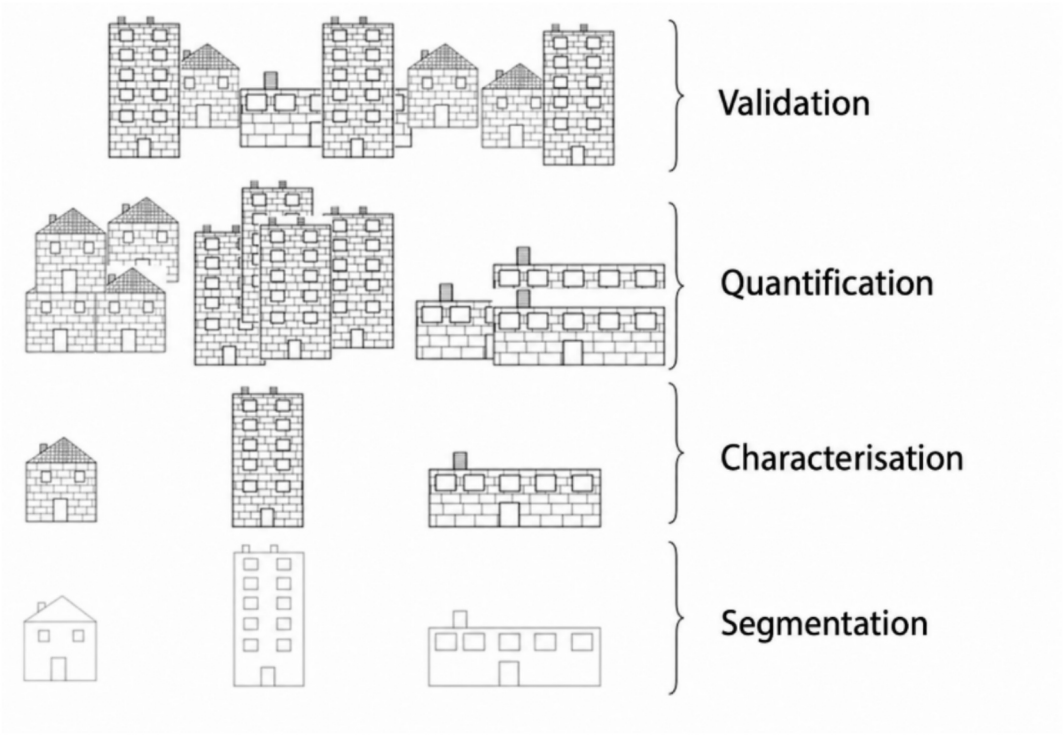


Fig. 3. Bottom-up Building Stock Modelling Approach [7].

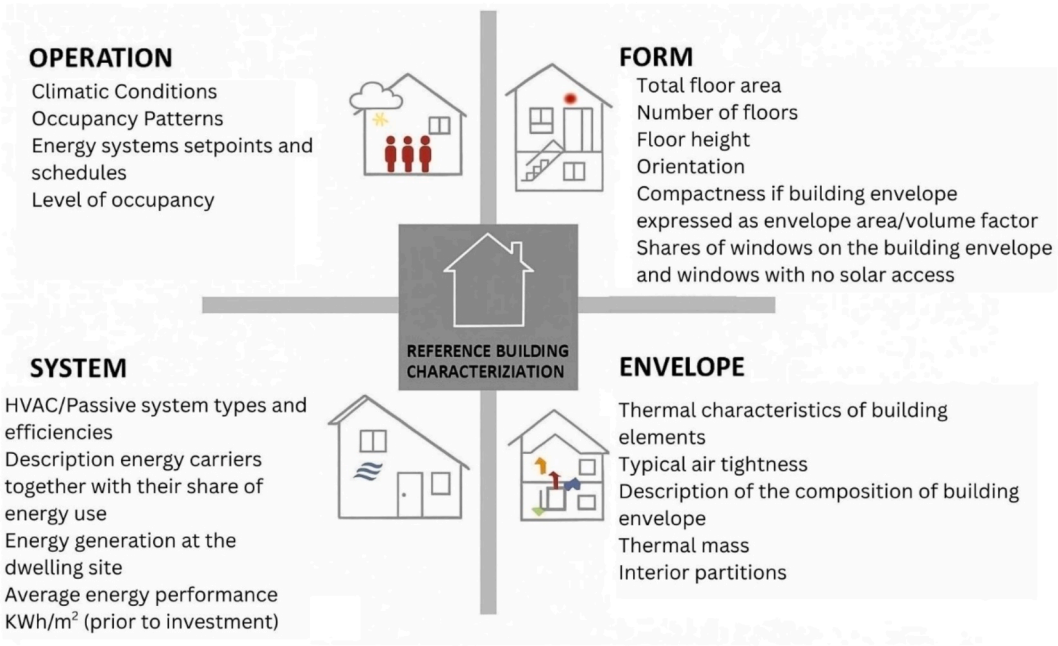


Fig. 4. Reference Building (RB) Characterisation for bottom-up Building Stock Energy Modelling (.
Adapted from Corgnati et al. [9])

typological modelling, their exclusive focus on residential dwellings renders them unsuitable for characterising the heterogeneous non-residential sector.

For energy efficiency tracking and policy monitoring scale, ODYSSEE framework [14] offers harmonised indicators on energy consumption, fuel mix, and efficiency metrics, while MURE [15] catalogues energy-efficiency policies and their estimated impacts. However, ODYSSEE–MURE [16,17] operates exclusively at the sectoral level, providing aggregated indicators for residential (households) and tertiary (services) buildings, but no building-level information. Similarly, the ENTRANZE project (2012–2014) [18], co-funded by the European Commission through the Intelligent Energy Europe (IEE) programme, provided analytical tools and scenario-based modelling using aggregated national datasets. Its approach focused on macro- and *meso*-scale policy analysis through sectoral aggregated outputs designed to inform strategic policy development rather than detailed, bottom-up BSEM or RB modelling. In the Irish context, ENTRANZE (2013) relied on data from the Sustainable Energy Authority of Ireland (SEAI) and the Central Statistics Office

(CSO), inheriting the structural limitations of those sources as detailed in section 1.4 [19].

Recent digital integration efforts emphasise interoperability infrastructure rather than dataset provision. The openDBL initiative proposes a standardised digital building logbook architecture capable of linking lifecycle and system data across platforms, although implementation remains developmental [20]. National open Energy Performance Certificates (EPC) database, such as that available in the United Kingdom (2023) [21], demonstrate how accessible certification data can support large-scale modelling, yet equivalent interoperability frameworks remain unevenly established across Member States.

1.4. Non-residential building data landscape in Ireland for stock energy modelling

The Sustainable Energy Authority of Ireland (SEAI) Commercial Buildings Survey (2015) provided the first comprehensive national overview of Ireland's non-residential building stock [22]. Based on a representative sample of approximately 1,500 properties, the survey classified non-residential buildings into five archetypes: retail, office, hotel, restaurant/public house, and warehouse. It estimated a national stock of approximately 109,000 non-residential buildings, with retail and office uses accounting for around 75% of the total. The analysis was underpinned by address-level data from Ireland's national postal service – An Post [23] and Ordnance Survey Ireland (now Tailte Éireann) [24], which jointly maintain the national GeoDirectory database.

Established in 1999, GeoDirectory provides a geocoded register of all residential and non-residential addresses in the Republic of Ireland, offering a verified national view of buildings, businesses, and small-area identifiers (Eircodes) [25]. Available under licence, GeoDirectory classifies non-residential premises using the Nomenclature of Economic Activities (NACE) system. Under this approach, buildings are categorised according to the dominant economic activity undertaken by their primary occupant, rather than their physical form, construction characteristics, or energy attributes. NACE codes are assigned through administrative processes linked to business registration and are subsequently associated with specific building addresses.

While NACE enables consistency and comparability across European statistical systems, its application to buildings introduces several limitations for building stock energy modelling. In multi-tenant or mixed-use buildings, classification is typically determined by the dominant activity, often defined by floor area, employment, or economic significance, which can obscure functional heterogeneity. Vacant buildings may retain the classification of previous occupants, and buildings within business parks or mixed-use developments may be used in ways that differ from their planning designation. Consequently, NACE-based categorisation does not reliably reflect building geometry, construction period, fabric characteristics, or energy-use profiles.

The CSO Commercial Property Statistical System (CPSS) feasibility study (2017) examined the potential for developing a unified national dataset for Ireland's non-residential buildings and identified several structural barriers [26]. These included:

- Fragmentation of data across custodians.
- Duplication and inconsistency in address-level records.
- Absence of standardised building classifications; and
- Variations in nomenclature for describing building use and business activity.

The Decarbonising the Commercial Built Environment roadmap (2024) published by the Department of Enterprise, Trade and Employment (DETE) reports that commercial, public-sector buildings, and industrial emissions account for approximately 11.17% of Ireland's total greenhouse gas emissions [27]. However, successive analyses by the Sustainable Energy Authority of Ireland (SEAI) [22] and Central Statistics Office (CSO) [26] have identified persistent deficiencies in available non-residential building data, including incomplete coverage, non-standardised classifications, and weak alignment across data custodians. The DETE roadmap [27] confirms the absence of a coherent national framework for monitoring and benchmarking energy performance, emphasising that effective decarbonisation depends on improved data governance and inter-agency coordination.

Data is the first step towards digitisation and although these initiatives represent important steps towards understanding Ireland's non-residential building stock, no comprehensive synthesis of electronic non-residential building data sources currently exists. As a result, the data infrastructure required to support national building stock energy modelling necessary to operationalising Articles 16 and 22 of the recast Energy Performance of Buildings Directive [4,5] remains heterogeneous and fragmented. This is further validated in recent research that proposes establishing a minimal set of "Essential Variables" (EVs) for accurate energy modelling at scale [28].

1.5. Research objectives

This study conducts a systematic scoping review of electronic datasets relevant to Ireland's non-residential building stock to identify, categorise, and evaluate existing data resources across public, institutional, and semi-state domains. The objective is to establish an evidence base for developing a unified, interoperable, machine-readable national non-residential building database.

Through application of a generalisable analytical model for database assessment, the review evaluates the completeness, attribute coverage, and interoperability of available datasets. Central to these efforts is the use of the Nomenclature of Economic Activities (NACE) classification system, which is mandated across all EU Member States (MSs) as the standard for categorising commercial activity. However, while NACE enables harmonisation and comparability across national datasets, its application to a building stock introduces a structural challenge: it classifies buildings by economic function rather than by their physical, thermal or energy-related characteristics. This limits its usefulness for energy performance assessment, renovation planning and stock modelling required under Article 22 [5]. As such, this study also evaluates the implications of relying on NACE within non-residential building datasets and explores how this misalignment might be addressed to better support energy-relevant data integration.

Although grounded in the Irish context, the study's methodological approach and findings are broadly relevant to other EU Member States (MSs) seeking to operationalise Articles 16 and 22 of the recast Energy Performance of Buildings Directive [4,5].

Through this approach, the study:

i. maps existing electronic data repositories relevant to Ireland's non-residential building stock, providing a replicable framework for identifying dispersed data sources;
ii. examines the recorded data attributes and their potential for integration, establishing criteria that can inform similar efforts in other national contexts;
iii. identifies existing unique identifiers and evaluates their suitability for linking datasets across administrative systems;
iv. highlights key data gaps and inconsistencies hindering the generation of reliable building-stock models; and
v. assesses the suitability of NACE as a classification framework for non-residential buildings in the context of Articles 16 and 22, and its implications for energy-relevant database interoperability across EU MSs.

The review is guided by the following questions:

i What electronic databases currently exist that provide information on Ireland's non-residential building stock?

ii What attributes relevant to non-residential buildings are recorded?
iii. What attributes relevant to non-residential building energy or retrofit characterisation are included?
iv. What unique identifiers are used across databases enabling data integration?
v. What are the main data gaps and limitations among the identified sources?
vi. How does the use of the NACE classification system influence the organisation, interoperability and energy relevance of non-residential building datasets, and what adaptations might be necessary to align it with Article 16 and 22 objectives?

By addressing these questions, the review provides a structured overview of Ireland's non-residential building data landscape while offering a transferable diagnostic framework that can be adopted by other MSs undertaking similar integration challenges. The resulting synthesis supports the creation of an interoperable national building database as a foundation for non-residential renovation planning, stock energy modelling and evidence-based climate policy across Europe.

2. Methodology and data

This study applies a scoping review to systematically map electronic datasets relevant to non-residential building stock characterisation in Ireland. Scoping reviews are particularly effective for mapping heterogeneous evidence, especially administrative records, government portals, and licensed datasets, which are insufficiently addressed by conventional systematic reviews of academic databases [29,30]. Previous national studies in Ireland did not apply a structured or replicable search protocol when examining non-residential building data; Milojevic-Dupont et al. [10], for example, used a three-step retrieval process informed by, however not fully aligned with, PRISMA for Scoping Reviews (PRISMA-ScR) To address this gap, a modified Joanna Briggs Institute (JBI) protocol [31] was adopted in this study, reported here in line with PRISMA-ScR. An internal protocol was developed with the Irish Building Stock Observatory (IBSO) [32] to define review boundaries, inclusion parameters, and validation procedures. The review followed five sequential steps: preparation, search, selection, extraction, and synthesis (as illustrated in Fig. 5).

2.1. Preparation phase (PCC framework)

Eligibility criteria were structured using the Population–Concept–Context (PCC) framework [31]. The *population* comprised the non-residential building stock of the Republic of Ireland, defined as buildings used for business, retail, education, healthcare, leisure, manufacturing, industrial, and other service activities, excluding housing. The *concept* encompassed both open and closed electronic data sources, with the latter accessible only in aggregated form, under licence, or via formal request, to ensure that restricted but evidentially central datasets (notably GeoDirectory and the Sustainable Energy Authority of Ireland EPC dataset) were not omitted. The *context* covered Irish national and regional datasets maintained by governmental, statistical, institutional, and private bodies recording attributes of non-residential buildings.

2.2. Search phase

A two-stage iterative search strategy, aligned with PRISMA-ScR, was implemented to capture both traditional and grey-literature sources, including institutional repositories and government portals.

Stage 1 – Initial Exploration: Broad Boolean combinations of core PCC-derived terms (*commercial / non-residential / non-domestic / industrial / public* combined with *building / property / stock*) were executed via Google and institutional platforms to characterise the data landscape

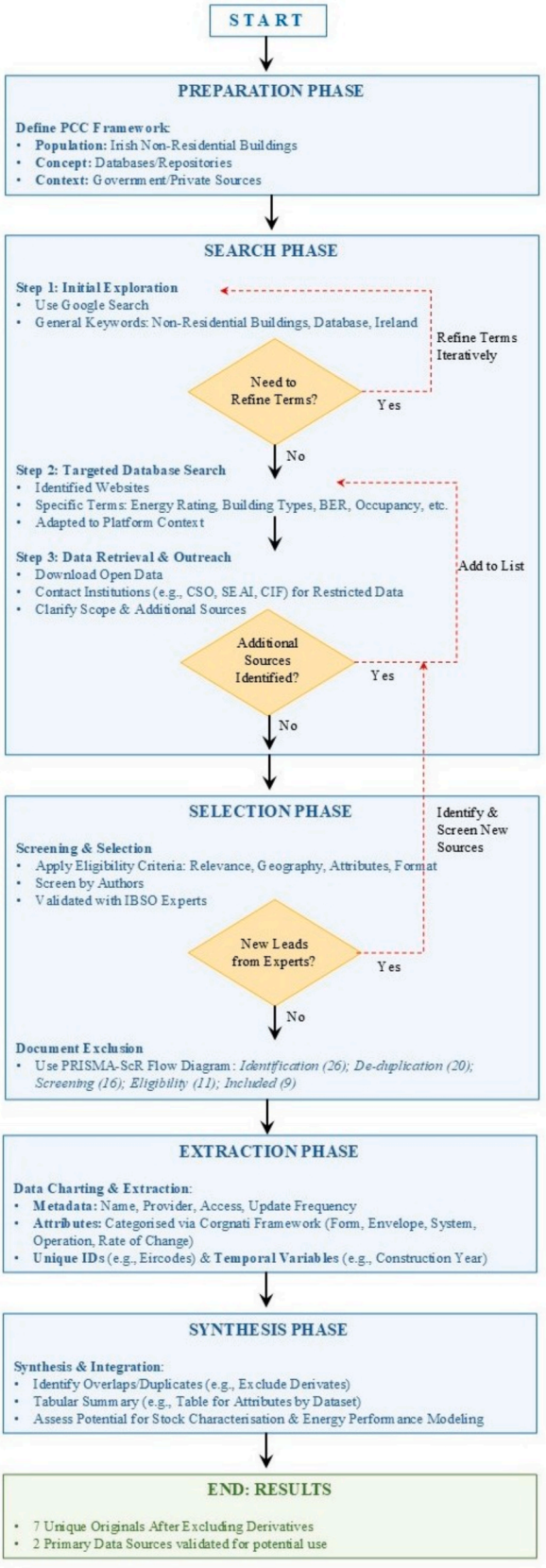


Fig. 5. Structural Framework of the Database Scoping Review.

and to test the sensitivity of search terms before refinement (as illustrated in Fig. 6).

Stage 2 – Targeted Search and Verification: Queries within identified repositories were refined using attribute-specific keywords aligned with the four domains of the Corgnati Reference Building (RB) framework [9], including form, envelope, system, and operation. These included building typology (hotels, retail, office, healthcare, educational, industrial), energy performance (Building Energy Rating (BER) / EPC, metered consumption), occupancy, valuation, lease, planning permissions, and retrofit activity. No temporal or language restrictions were applied; English was the working language. Where datasets were restricted, custodians such as An Post GeoDirectory, CSO, SEAI, the Construction Industry Federation, and the Department of Housing, Local Government and Heritage were contacted directly to verify metadata and, where appropriate, obtain access under licence or non-disclosure agreement.

Inclusion criteria, summarised in Fig. 7, required that a source:

1. Contain information on non-residential buildings as defined in the population;
2. Pertain to the Republic of Ireland;
3. Include at least one attribute relevant to building characterisation (e.g., use, size, geometry, or energy performance); and
4. Available in electronic format (structured database, spreadsheet, GIS layer, or interactive dashboard).

2.3. Selection phase

Identified sources were catalogued and assessed against the eligibility criteria through a structured, iterative screening process. Initial screening examined websites and repositories to confirm relevance; for each candidate, metadata covering provider, accessibility, data format, and update frequency were logged. The compiled inventory was then reviewed by experts from the Irish Building Stock Observatory (IBSO) [32], who flagged additional institutional sources that were subsequently screened under the same protocol.

The screening outcome is summarised in Fig. 8 and Table 1. The initial search identified 26 candidate repositories. De-duplication and exclusion of generic aggregators (Google Dataset Search, GitHub, Kaggle) that contained no Irish non-residential datasets reduced this to 20. A further four sources were excluded for lacking downloadable non-residential building datasets, and five were excluded at the access-check stage (Land Development Agency (LDA) and Construction Industry Federation (CIF) referred to CSO/data.gov.ie [33]; Digital Repository of Ireland (DRI) non-responsive; Zenodo non-relevant). A final inclusion review excluded Mendeley (residential-only) and the Housing Agency Data Hub (planning permissions only), yielding nine datasets for full review (Table 2). The complete screening register for all 26 candidate repositories is provided as supplementary material in Appendix A.

2.4. Extraction and synthesis

For each of the nine included datasets, recorded attributes were extracted and characterised against the four-domain Corgnati RB framework [9]: form (geometry, building type, floor area), envelope (construction and thermo-physical properties), system (HVAC and services), and operation (occupancy and use patterns). Datasets were additionally classified by access type (open, licensed, or restricted), coverage (population vs. sample), spatial granularity (national, county, or building-level), and update frequency.

Synthesis proceeded in three steps designed to address the research objectives and the requirements of EPBD Article 22, which are as follows:

1. A qualitative quality assessment evaluated each dataset's suitability for inclusion in an interoperable national non-residential building database.
2. The two datasets with the broadest coverage are: the GeoDirectory building register [34] and the SEAI EPC dataset [35], which were tested for interoperability through bidirectional mapping between the activity-based NACE classification [36] and the building-energy NEAP typology [37], using both bottom-up (NEAP → NACE) and top-down (NACE → NEAP) approaches.
3. The EPC sample was quantified against the GeoDirectory population to identify systematic sampling biases and coverage gaps by building type. This three-step synthesis underpins the analyses in Section 3 and the policy implications drawn in Section 4.

3. Results and analysis

3.1. Data charting and extraction

From the nine datasets identified in Table 2, the data extraction process was structured to systematically document and evaluate the recorded attributes, with the aim of assessing their suitability for integration and the development of Reference Buildings.

3.2. Data synthesis and integration

Fig. 9 presents a star schema illustrating the relationships among the nine selected electronic databases in Table 2, and their collective role in supporting characterisation of Ireland's non-residential building stock for energy performance modelling and retrofit planning.

Referring to Fig. 9:

- At the upper level of the schema are two spatial datasets – the Tailte Éireann Prime 2 base map and the GeoHive GIS platform. These provide the spatial foundation for mapping and visualising non-residential buildings in Ireland. The Prime 2 dataset supplies detailed building footprints and height information, while GeoHive includes geocoded planning application points that allow the tracking of development activity and spatial growth patterns.
- The core analytical datasets occupy the centre of the schema (see Fig. 9) form the backbone for energy performance assessment. These include the national SEAI EPC dataset [35], which contains data necessary to RB classification (see Fig. 4); the GeoDirectory Business Address Database [34], which provides business address-level data and NACE-based commercial classifications for all buildings in Ireland; and the Central Statistics Office (CSO) Energy Consumption dataset, which offers both aggregated and micro-level statistics on energy use by sector. The data.gov.ie portal functions as a national aggregator for many of these sources, republishing datasets from government departments, including planning and commercial lease records.
- At the lower tier of the schema are contextual datasets, including the National Planning Application dataset, the Property Services Data Catalogue (Lease Register), and the Valuation Office Property Valuation dataset. While not central to energy performance analysis, these sources enhance the broader data ecosystem by providing regulatory, financial, and transactional context. They also offer

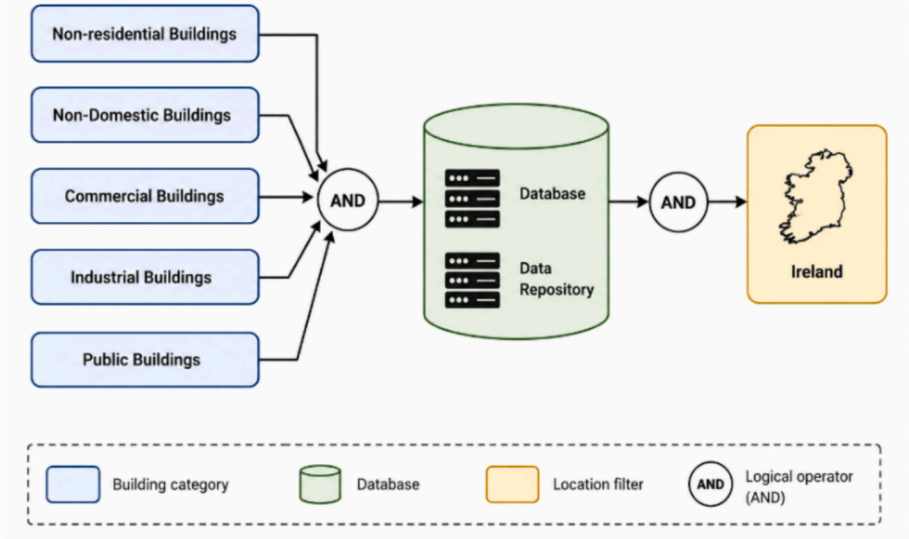


Fig. 6. Boolean String.

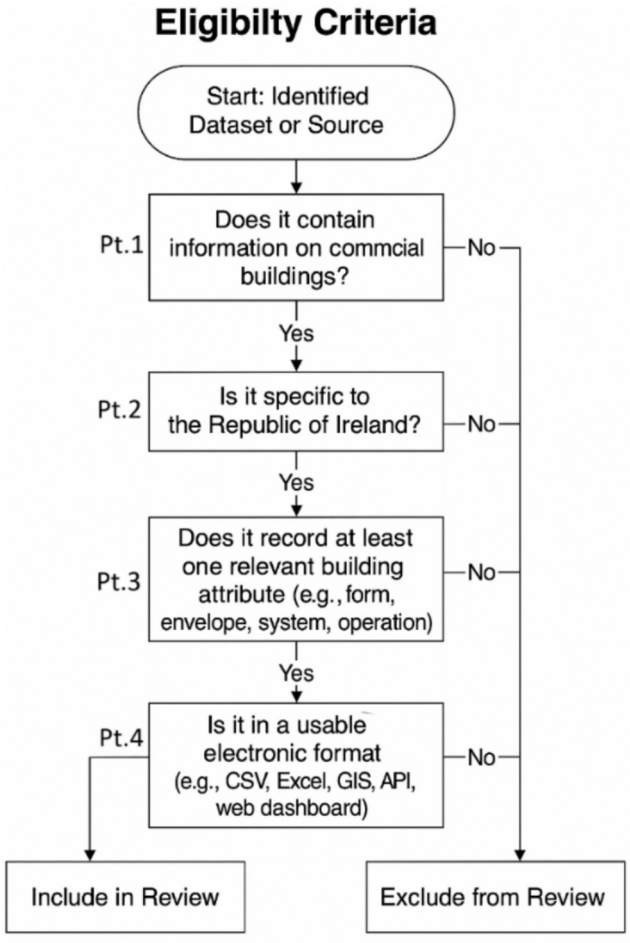


Fig. 7. Inclusion and Exclusion Criteria.

potential for future integration into digital building logbooks consistent with Article 16 of the EPBD [4].

Two datasets were subsequently identified as derivatives of others: the Planning Applications Granted dataset from data.gov.ie is a subset of the National Planning Application database; and the CSO Energy Consumption dataset, which draws on the SEAI EPC data but employs an alternate building identifier. After excluding these derivatives, seven original datasets remained for final analysis. The remaining 7 datasets are classified in Table 3 by access type – *open* or *restricted* – along with data attributes.

The seven datasets listed in Table 3 were accessed through the relevant channels. Each dataset's column attributes were examined and documented in Table 4. Data attributes are organised according to the Corgnati RB Framework [9] (see Fig. 4). Unique identifiers present in each dataset were recorded to support future data integration and cross-referencing. Attributes containing temporal variables – such as *Construction Year,* in the SEAI EPC dataset, or *Lease Date* and *Lease Length* in the Commercial Lease Register – were grouped under a separate category titled "Rate of Change." This category captures variables indicative of building conversions, ownership transitions, and renovation frequency, offering insights into the evolution and lifecycle dynamics of Ireland's non-residential building stock. Datasets are evaluated in Section 3.3.

3.3. *Evaluation of seven selected datasets*

Referring to Table 4, assessment reveals notable gaps across key domains. Building construction year as well as attributes describing building form, such as year of construction, floor area, and number of storeys, are comparatively well represented in datasets including GeoDirectory, the SEAI EPC database, and Tailte Éireann sources. Nevertheless, additional harmonisation is required to achieve consistency of these attributes across datasets – for instance SEAI bin construction year by construction period relating to thermal eras in Ireland. In contrast, building envelope characteristics (e.g. insulation levels, air tightness, and construction materials) and system-level attributes (e.g. HVAC configurations and renewable energy installations) are recorded only within the SEAI EPC dataset. Within the SEAI EPC dataset, operational attributes, including occupancy type, vacancy status, and intensity of use, are limited. In other datasets, these operational attributes can perhaps be inferred indirectly (e.g. through NACE classifications in GeoDirectory Dataset or lease-duration data in Lease Register). This constrained attribute representation restricts the modelling of behavioural and operational effects, which are critical for robust energy demand modelling, archetype development, and retrofit planning.

Understanding data lineage, overlap, and integration pathways is critical when considering the establishment of a single authoritative national building dataset. As shown in Fig. 10, the attributes recorded

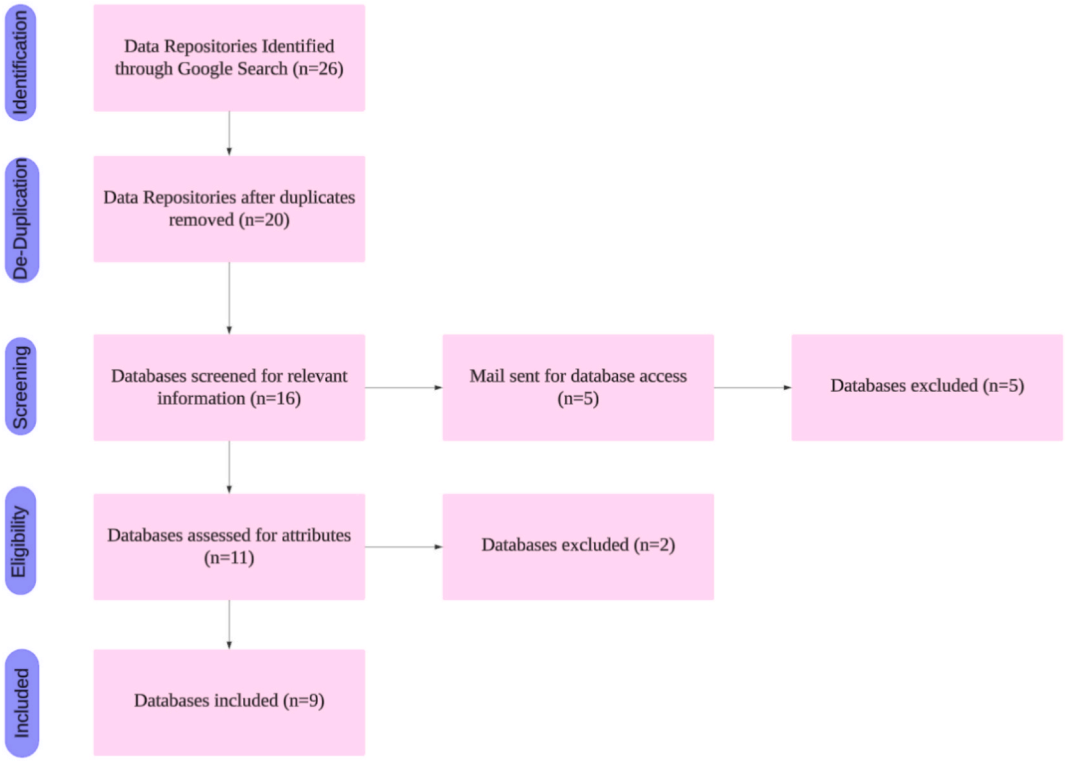


Fig. 8. PRISMA flow diagram for scoping review process.

across the seven datasets in Table 4 originate from a smaller number of authoritative data producers, primarily the Sustainable Energy Authority of Ireland (SEAI), Tailte Éireann, and An Post.

Repositories such as the Central Statistics Office (CSO) and data.gov.ie function mainly as data aggregators, republishing selected attributes from these primary sources for secondary analytical purposes. For example, the GeoHive GIS platform is a service of Tailte Éireann's Registration, Valuation and Surveying Department and is built on the Prime 2 spatial dataset, while commercial property valuation data are integrated within Tailte Éireann's mapping infrastructure to support spatially enabled valuation functions. As illustrated in Fig. 10, energy-related attributes, such as heating fuel type and energy consumption, originate from the national EPC dataset and are subsequently republished across multiple repositories. Similarly, geometric attributes including building footprint, height, and floor area are derived from Tailte Éireann's Prime 2 dataset before being integrated into GeoDirectory or GeoHive.

Data on property prices and lease terms are provided by the Property Services Regulatory Authority (PSRA), while valuation information is drawn from a combination of sources, including the Land Registry, the Property Deeds Register, and local authorities, based on owner declarations or survey-based collection.

3.3.1. *Database quality assessment for integration into an interoperable building database*

To provide a basis for determining the readiness of existing datasets to support the creation of a consolidated, interoperable national non-residential building database, a structured data quality assessment was undertaken to evaluate completeness, attribute coverage, classification consistency, granularity, update frequency, and interoperability across the seven selected datasets.

3.3.1.1. Data completeness. As Ireland lacks a definitive master list of all non-residential buildings, assessing dataset completeness in absolute terms is challenging. The GeoDirectory database provides the most comprehensive coverage of the building stock in Ireland. GeoDirectory identifies 156,011[1] units[2] designated as non-residential,[3] while the SEAI EPC dataset contains 95,745 non-residential entries (as of November 2024).

Statistically, the SEAI EPC dataset constitutes a highly significant sample size, achieving a 99% confidence level (1% margin of error) relative to the GeoDirectory population. This indicates strong overall representation at an aggregate level. Section 3.6 examines dataset completeness by building typology (e.g. office, retail).

3.3.1.2. Data granularity. Most datasets, except the Planning Application dataset, provide unit-level data suitable for detailed spatial and energy modelling. The Tailte Éireann Prime 2 and GeoHive maps offer high-resolution geospatial data, while the SEAI EPC and Lease Register datasets provide functional and temporal variables such as construction year, energy rating, and lease duration. GeoDirectory additionally includes form data such as area and operational data including building type, business activity, and vacancy status, making it highly granular data source. However, the Planning Application dataset records information such as building type, construction type, and submission/grant dates only at application level (by submission and approval event) rather than building unit level, limiting its analytical potential for stock

[1] This figure excludes 58,028 units labelled as commercial by unit-use but lacking an associated NACE classification. In addition, a further 89,044 units classified under Agriculture, Forestry and Fishing (NACE Category A)—predominantly farm and livestock buildings—are excluded as they are not represented in EPC assessments.

[2] A unit is the lowest denominator for building classification which refers to the specific purpose or category for which an individual unit within a building is used.

[3] Strictly commercial/non-residential building are denoted a 'C' in the GeoDirectory Dataset, while mixed use buildings/units are denoted as 'B' for 'both'.

Table 1
Data repository Screening Process (based on PRISMA-ScR (Fig. 8)).

Stage	Description	Repositories/ Datasets	Count Remaining
Identification	Initial search identified websites containing data on non-residential buildings in Ireland.	26 websites identified	26
De-duplication	Removed duplicate entries where platforms redirected to other repositories. Removed popular data repositories – no dataset available • *Department of Housing Open Data* → redirected to data.gov.ie • *Construction Information Services* → redirected to CSO / data.gov.ie datasets • *Google Dataset search* • *Kaggle* • *Github*	6 removed	20
Screening	Excluded websites with no non-residential building datasets in Ireland. • *Irish Green Building Council* • *City Council websites* • *National Inventory of Architectural Heritage* • *Google Explorer*	4 removed	16
Eligibility (Access Check)	Reviewed accessibility of remaining repositories. • *Land Development Agency (LDA)* – referred to CSO/ Open Data • *Construction Industry Federation (CIF)* – referred to CSO/Open Data • *Digital Repository of Ireland (DRI)* – no response • *Zenodo* – no relevant building datasets	5 removed / redirected	11
Inclusion	Final screening for relevance and data content. • *Mendeley Database* – residential energy data only • *Housing Agency Data Hub* – planning permissions only	2 removed	9 final datasets included (see Table 2)

modelling.

3.3.1.3. Update frequency and dataset currency. Update frequency varies significantly between datasets. The Planning Application dataset (particularly Dublin City Council) is updated daily, while SEAI EPC and GeoDirectory are updated quarterly, reflecting continuous additions of new records. Conversely, the Lease Register and Property Valuation datasets are updated on an ad-hoc or irregular basis, reducing their value for time-series or real-time analyses.

3.4. Selection of non-residential datasets useful to building stock energy modelling

Referring to Fig. 11 and Table 4, the comparative attribute analysis identifies the SEAI EPC and GeoDirectory datasets as the most suitable sources for supporting non-residential building energy stock analysis in Ireland. The SEAI EPC dataset offers the most comprehensive coverage of technical and operational attributes, including thermal performance, building form, system characteristics, and standardised occupancy assumptions aligned with the Cognati RB framework [9]. This rich attribute set, combined with the statistical significance of the sample, enables robust characterisation and segmentation of non-residential Reference Buildings within a bottom-up building stock energy modelling approach.

Complementing this, the GeoDirectory dataset provides the most complete representation of the non-residential building stock at the national scale, including geocoded spatial information and business classification data. While it lacks detailed energy performance attributes, GeoDirectory is uniquely suited to quantifying the overall extent and spatial distribution of non-residential buildings across Ireland. The federation of the SEAI EPC and GeoDirectory datasets therefore enables both detailed building-level energy characterisation and population-level stock quantification. In addition, integration with ancillary sources, such as property valuation records, lease registers, and planning application databases, offers further potential to infer rates of construction, demolition, and renovation. On this basis, the SEAI EPC and GeoDirectory datasets are selected as the core data sources for subsequent non-residential building energy stock modelling. Section 3.5 examines the interoperability of the two selected data sources.

3.5. Interoperability of selected datasets (SEAI EPC dataset VS GeoDirectory dataset)

A persistent challenge across the two selected datasets is the inconsistency in building classification systems, which significantly constrains cross dataset interoperability. This section examines the nature and implications of this issue.

3.5.1. Characteristics of the GeoDirectory population dataset [34]

Referring to Table 5 and Fig. 12 of the 156,011 non-residential units in Ireland assigned a Nomenclature of Economic Activity (NACE) code of business activity, the largest share belongs to;

- 'Wholesale and Retail Trade' (including motor vehicle repair), accounting for 21.85% of all non-residential units.

This is followed by;

- 'Accommodation and Food Service Activities' at 14.40%, and
- 'Other Service Activities' at 12.19%.
- 'Human Health and Social Work Activities' represent 9.63%, while
- 'Professional, Scientific and Technical Services' account for 7.82%.

Smaller but notable sectors include Arts, Entertainment and Recreation (5.83%), Construction (4.72%), Education (4.62%), and Manufacturing (4.08%). The remaining shares are distributed across Administrative and Support Services (3.47%), Transportation and Storage (2.61%), Financial and Insurance Activities (2.47%), Public Administration and Defence (2.13%), and Information and Communication (1.62%). Less frequent sectors include Real Estate (1.55%), Water Supply, Waste and Remediation (0.58%), Electricity and Gas Supply (0.22%), Mining and Quarrying (0.18%), and Activities of Extraterritorial Organisations (0.05%).

3.5.2. Characteristics of the SEAI EPC sample set [35]

The SEAI non-residential EPC dataset, obtained under licence from SEAI [35], classifies non-residential buildings using the Non-Domestic Energy Assessment Procedure (NEAP) [37]. NEAP assessments evaluate building energy performance based on envelope and system characteristics, including space heating, cooling, water heating, and lighting. The procedure adopts a building-type classification framework derived from the Simplified Building Energy Model [38], whereby each building type is assigned standardised operating parameters, such as temperature set-points and occupancy schedules, that directly influence calculated energy demand. As shown in Table 6 and Fig. 13, the EPC dataset contains 46 building typologies (as opposed to 19 NACE classifications). The largest proportion of non-residential buildings are classified as:

Table 2
Nine selected datasets.

Sr. No.	Source Name	Category	Data Type	Rationale/Sub-Details	Input (Key Words)	Output (Datasets)	Citation
1	GeoDirectory	Commercial/ Private (Grey/ Restricted)	Tabular (CSV, Excel); Spatial (GIS/Shapefiles)	Partnership (An Post/OSI); requires purchase or license for full access; granular business addresses for non-residential mapping.	Non-residential/ Commercial Building; Ireland	GeoDir-ectory Business Dataset [34]	GeoDirectory Ltd. (n.d.). GeoDirectory Database [Data set]. https://www.geodirectory.ie/. Contains GeoBuilding Intel under licensing agreement Accessed May 21, 2025.
2	Sustainable Energy Authority of Ireland (SEAI) EPC Dataset	Government/ Open Data Portals (Grey)	Tabular (CSV, Excel/XLSX)	From Sustainable Energy Authority of Ireland; publicly accessible via filtered downloads, but raw/full data require request due to inconsistencies.	Non-residential/ Commercial Building; Energy Performance; Ireland	SEAI EPC Non-Domestic Dataset [35]	SEAI (n.d.). SEAI EPC Dataset [Dataset]. https://data.gov.ie/dataset/seai-epc-dataset. Contains Irish Public Sector Data licensed under CC BY 4.0. Accessed February 07, 2025.
3	Planning Application Database	Government/ Open Data Portals (Grey)	Tabular (CSV); Spatial (ZIP, GEOJSON, KML)	National government database; freely available on data.gov.ie with aggregated spatial/tabular downloads.	Non-residential/ Commercial Building; Ireland; National Planning Data	National Planning Applicat-ion Dataset	Department of Housing, Local Government and Heritage. (n.d.). Planning Application Database [Data set]. https://data.gov.ie/dataset/planning-application-database. Contains Irish Public Sector Data licensed under CC BY 4.0. Accessed October 07, 2024.
4	Central Statistics Office – CSO	Government/ Open Data Portals (Grey)	Tabular (PX/ Statistical formats, CSV); Time-series	National stats office; freely downloadable datasets via CSO.ie and open data portal.	Non-residential/ Commercial Building; Ireland; Database	Energy consumpt-ion Dataset	Central Statistics Office. (n.d.). CSO Databases [Data set]. https://www.cso.ie/en/databases/. Accessed October 07, 2024.
5	data.gov.ie	Government/ Open Data Portals (Grey)	Various (CSV, JSON, XML, Shapefiles, GEOJSON)	National open data portal; all datasets freely reusable under open licenses.	Non-residential/ Commercial Building; Ireland; Database	Planning Application granted Dataset	Government of Ireland. (n.d.). data.gov.ie [Data set]. https://data.gov.ie/. Contains Irish Public Sector Data licensed under CC BY 4.0. Accessed October 07, 2024.
6	Property Service Data Catalogue		Tabular (CSV, JSON); API-accessible	Government regulatory authority; open catalogue for property/lease data.	Non-residential/ Commercial Building; Ireland; Database	Lease Register Dataset	Property Services Regulatory Authority. (n.d.). Property Service Data Catalogue [Data set]. https://www.psr.ie/data-catalogue. Accessed October 07, 2024.
7	Commercial Property Valuation Database		Tabular (CSV exports); Map-based/Spatial	Valuation Office Ireland; map-based and downloadable data freely available.	Non-residential/ Commercial Building; Ireland; Database	Property Valuation Dataset	Tailte Éireann – Valuation Office. (n.d.). Commercial Property Valuation Database [Data set]. https://data.gov.ie/dataset/commercial-property-valuation-database. Contains Irish Public Sector Data licensed under CC BY 4.0. Accessed October 07, 2024.
8	GeoHive Planning Application Points	Government/ Open Data Portals (Grey)	Spatial (GIS/ Shapefiles, GEOJSON)	Tailte Éireann; free GIS downloads via GeoHive portal.	Non-residential/ Commercial Building; Ireland; Map	Planning Application Points/sites Map	Tailte Éireann. (n.d.). GeoHive Planning Application Points [Data set]. https://geohive.ie/. Contains Irish Public Sector Data licensed under CC BY 4.0. Accessed March 25, 2025.
9	Tailte Prime 2 dataset		Spatial (Vector/ Geodatabase, Shapefiles)	Ordnance Survey Ireland; base map data licensed (some open, full access may require agreement).	Non-residential/ Commercial Building; Ireland; Map	Tailte Éireann Prime 2 dataset (OSi base map)	Tailte Éireann. (n.d.). Tailte Prime 2 Dataset [Data set]. https://tailte.ie/services/open-data/. Contains Irish Public Sector Data licensed under CC BY 4.0. Accessed March 25, 2025.

- Retail (26.97%), followed by
- Office (16.21%) and
- Retail/Financial and Professional Services (10.67%).

This is followed by:

- Offices and workshop businesses account for 7.64% of the stock,
- Warehouses and storage for 7.21%, and
- Restaurants/public houses for 6.60%,

while all remaining categories individually represent less than 3%.

As shown in Table 6, the distinction between the categories *retail*, *retail/financial and professional services*, and *retail warehouse* is not clearly defined within the descriptors provided in the NEAP guidance document [37]. Similarly, *office* and *office and workshop business* are treated as separate categories, yet the basis for this differentiation is not explicitly articulated. The NEAP guidance document [37] *identifies 'Offices and Workshop Business'* as general office type buildings in addition to light industry or research and development of products appropriate to be carried out in residential areas. *'Office'* does not appear as a separate

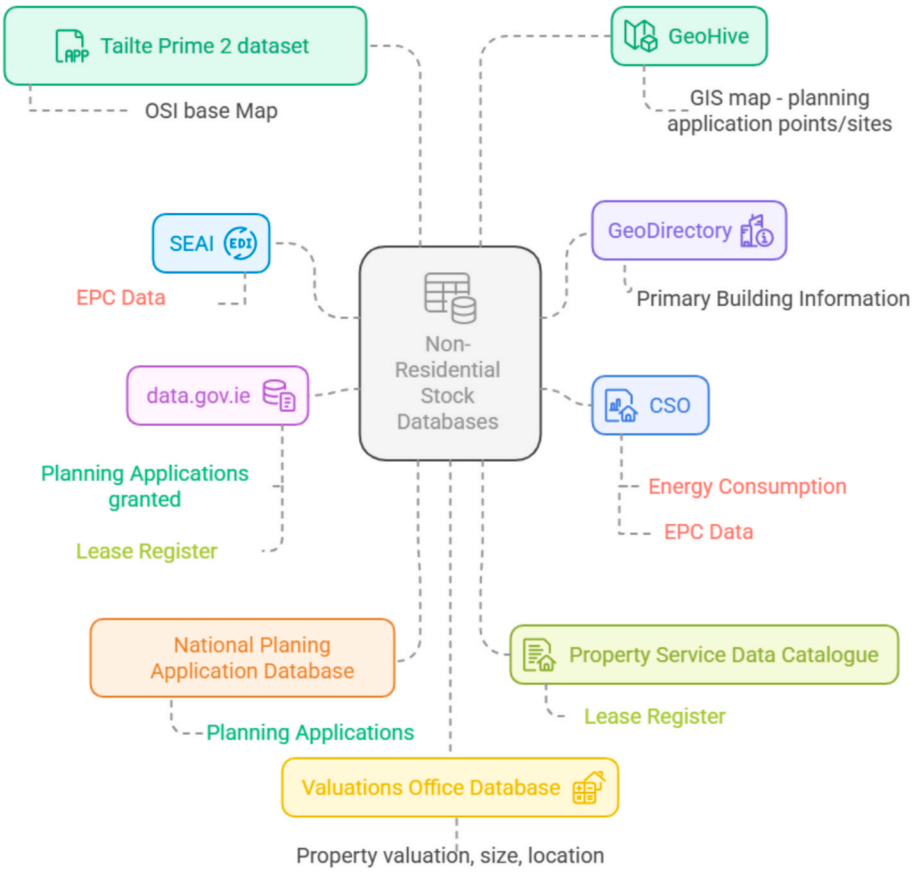


Fig. 9. STAR Schema of Non-residential Building Stock Databases in Ireland.

building type category under NEAP classification. Within the current SEAI EPC dataset, *retail* – when aggregated across retail, retail/financial and professional services, and retail warehouse – accounts for approximately 40% of non-residential buildings, while a further 24% are classified as *office* or *office and workshop business*, representing 64% of the total non-residential stock. This distribution diverges from the findings of the SEAI survey [22], which reported that approximately 75% of non-residential buildings classified as retail or office.

The SEAI EPC classification system provides a high level of granularity by differentiating buildings with broadly similar functions into distinct sub-types. For example, educational buildings are disaggregated across multiple categories, including:

- **Non-residential Institutions:**
  - Primary Education (0.95%),
  - Post-primary Education (0.56%),
  - Community/Day Centre (0.80%),
  - Primary Health Care Building (0.68%),
- **Residential Institutions:**
  - Universities and Colleges (0.26%),
  - Residential Primary Schools (0.04%), and
  - Residential Post-primary Schools (0.03%).

3.6. Alignment between building energy typologies and economic activity classifications

As shown in Fig. 11, a key ambition of this study is to assess whether the SEAI EPC dataset, classified by NEAP building typologies can be used as a representative sample for characterising the national non-residential building population classified in GeoDirectory under NACE economic activity codes.

This approach presents significant methodological challenges. The EPC dataset categorises buildings according to physical form and functional typology (e.g., office, warehouse, school), whereas NACE classifications reflect the economic activity of buildings. Consequently, relationships between these systems are neither hierarchical nor one-to-one: individual NEAP categories may correspond to multiple NACE sectors, and individual NACE sectors may occupy multiple building types. This lack of direct interoperability complicates aggregation, scaling, and assessments of representativeness. To evaluate these relationships, both bottom-up (NEAP-to-NACE) and top-down (NACE-to-NEAP) mapping exercises were undertaken in an attempt to align EPC NEAP building categories with GeoDirectory NACE economic activity classes.

3.6.1. NEAP to NACE (bottom-up)

Bottom-up mapping was undertaken to associate the 46 NEAP building typologies contained within the EPC dataset with corresponding NACE activity categories. This was achieved through rule-based matching grounded in functional similarity between typology descriptions and NACE classifications, as summarised in Fig. 14 and detailed in Table 7.[4] Two governing assumptions underpinned this allocation process:

1. GeoDirectory counts were treated as the population upper bound for each NACE category.
2. EPC/NACE records were considered a subset of the NEAP population and were assigned either fully or proportionally, provided that category-level population limits were not exceeded.

[4] Table 7 describes the mapping of the ten predominant NEAP categories following their aggregation.

Table 3
Dataset Accessibility of 7 selected datasets.

		Dataset Name	Description
Access Type	**Open**	GeoHive Planning Application Points	ArcGIS-based dataset showing geographic locations of planning applications. Includes site area and construction type attributes.
		Planning Application Granted	Dataset containing records of approved planning applications with details on building type and construction activity (e.g., new builds, extensions).
		Commercial Lease Register	Data on commercial property leases including address, property type, lease start date, average rent, and lease length.
	Restricted	Tailte Éireann Prime 2	Base spatial dataset in ArcGIS format offering property boundaries, building use types (e.g., school, hospital), and polygon geometry; available on request.
		SEAI EPC Dataset	Contains energy performance characteristics including BER ratings, U-values, energy system details, and operational attributes; access available via formal request.
		GeoDirectory Business Dataset	Jointly managed by An Post and Tailte Éireann, includes detailed address, business classification (NACE codes), and location data for over 200,000 business premises; available for purchase.
		Commercial Property Valuation Database	Managed by the Valuation Office; includes data on property category, use, floor area, floor level, and geographic identifiers. An API is available for access.

NEAP *Retail* and *Retail/Financial and Professional Services* categories amounting to 36,042 (37.64%) buildings were primarily mapped to NACE category G (Wholesale and retail trade; repair of motor vehicles and motorcycles) having with 34,086 units classified as category G in GeoDirectory. From the remaining count, 27.5% of *Retail/Financial and Professional Services* (2,814 out of 10,219 units) were mapped to NACE category K (Financial and Insurance Activities), as this category includes banks, estate and employment agencies, professional services, and betting offices (3,849 units). In addition, 56.5% of *Offices and Workshop Businesses* (4,131 out of 7,318 units) was mapped to G, reflecting the presence of light industry and small-scale product development activities permitted in residential areas. These allocations fully populated NACE category G.

For, Office-related allocations (15,523 units) of *Office* and *Offices and Workshop businesses* building type, these were fully allocated to NACE category K, along with M (Professional, Scientific and Technical Activities), N (Administrative and Support Service Activities), J (Information and Communication), and L (Real Estate Activities).

Storage NEAP categories *Offices and Workshop Businesses* (3,187 out of 7,318 units) were mapped to NACE category H (Transportation and Storage), while NEAP categories *Warehouse and Storage* (6,908), *Retail Warehouses* (2,431), and *Storage or Distribution* (2,299) showed strong correlation to NACE category H. The combined EPC count of these storage categories exceeded NACE category H count, making it an exception to Rule 1 due to the strong functional correspondence between them.

NEAP categories associated with industrial activity—*Workshop/Maintenance Depot* (2,817), *Industrial Process Buildings* (1,515), and *General and Special Industrial Buildings* (529) were mapped fully to NACE category C (Manufacturing), reflecting their role in heavy or specialised industrial processes.

Hospitality-related NEAP categories (10,547 units) – *Restaurants/Public Houses*, *Cafes/Drinking Establishments/Takeaways*, and *Hotels* – were mapped fully to NACE category I (Accommodation and Food Service Activities), while all education-related NEAP categories (2881 units) were fully mapped to NACE category P (Education), and all healthcare and community-based categories (4,843 units) were mapped entirely to NACE category Q (Human Health and Social Work Activities). Recreational and cultural buildings (1,583 units) were mapped to NACE category R (Arts, Entertainment and Recreation).

Smaller private service buildings were assigned to NACE category S (Other Service Activities), and government-owned or administrative buildings (e.g. courts, prisons, transport terminals) were mapped to NACE category O (Public Administration and Defence).

Overall, bottom-up (NEAP to NACE) mapping confirms the feasibility of broad categorical alignment but reveals structural ambiguity and residual unmapped capacity in healthcare, education, public administration, and industrial buildings, reflecting conceptual differences between economic and morphological classification systems. For this reason, a top-down approach was tested wherein the full GeoDirectory NACE population was reclassified using NEAP building typologies.

3.6.2. NACE to NEAP (top-down)

The top-down mapping exercise – from NACE activity classifications in the GeoDirectory dataset to NEAP building types in the SEAI EPC dataset – was conducted using the same core assumptions applied in the forward NEAP-to-NACE allocation. All 19 NACE categories were mapped to one or more NEAP building types based on descriptive and functional similarity as shown in Fig. 15. In this approach, observed NEAP counts were treated as capacity limits, and NACE sectors were sequentially allocated to compatible building typologies based on descriptive similarity.

Six NACE categories achieved complete (100%) mapping within the NEAP building typology, indicating strong alignment between economic activity and building type (Fig. 15 and Table 7);

- G – Wholesale and retail trade; repair of motor vehicles and motorcycles (34,086 units) mapped to *Retail* and *Retail/Financial and Professional Services* (36,042 units).
- H – Transportation and Storage (4,072 units) mapped to *Warehouse and storage* (6,908 units).
- K – Financial and Insurance Activities (3,849 units) mapped to *Office* (15,523 units) and *Retail/Financial and Professional Services* (10,219 units).
- J – Information and Communication (2,525 units), N – Administrative and Support Service Activities (2,416 units), L – Real Estate Activities (5,406 units), together mapped to the *Office* type (15,523 units).

Beyond these six categories, only 76% (4,861 units) of NACE category C (Manufacturing) was mapped to *Workshops/Maintenance Depot*, *Industrial Process Buildings*, and *General Industrial and Special Industrial Groups*, while only 60% (7,318 units) of M (Professional, Scientific and Technical Activities) mapped to *Office* and *Offices and Workshop Businesses*, assuming a predominantly office-based physical form.

Moderate correspondence was observed within I (Accommodation and Food Service Activities: 47% (10,547 units)) to hotels and restaurant related NEAP categories, P (Education: 40% (2,881 units)) to residential and non-residential schools and universities, and Q (Human Health and Social Work Activities: 32% (4,843 units)) to similar healthcare building typologies.

For the categories, R (Arts, Entertainment and Recreation: 17% (1,583)), S (Other Service Activities: 2% (446)), and O (Public Administration and Defence; Compulsory Social Security: 2% (63)) mapping was extremely poor indicating that buildings associated with these activities are largely absent from, or weakly represented within, the

Table 4
Dataset Attribute Categorisation along with Temporal Dynamics and Population Coverage of the 7 identified datasets.

				Corgnati Attribute Categories [9]						
		Publisher	**Purpose**	**Form**	**Envelope**	**System**	**Operation**	**Rate of Change**	**Unique Identifiers**	**Population Coverage**
Dataset	GeoDirectory Business Dataset	GeoDirectory (An Post & Tailte Éireann)	Comprehensive address and business database to support location-based decision-making.	Area			Building Type, Business Name, NACE Classification Type, Activity, Vacancy Data		Address, Eircode, Northing/ Easting	100%
	SEAI EPC Dataset	Sustainable Energy Authority of Ireland	Supports research on energy use and policy for energy-efficient buildings.	Area*	U Value^, BER Rating	Building Environment, Main Fuel^, Primary Energy Consumption, CO2 Emission	Business Type	Construction Year	Address, Eircode, MPRN	61%
	Planning Application Database	Local Planning Authorities via data.gov.ie	Enhance transparency in planning; support policy and development oversight.				Building Type, Construction Type˙	Year/Quarter permission granted	Not specified	County Wise Details
	Property Valuation Database	Valuation Office Ireland	Property valuation for tax and planning; supports economic and urban analysis.	Floor Area, Floor Level			Property Category, Property Use‡, Floor Use		Address, Property Number, Northing/ Easting	Low, user input dependent
	Lease Register	Property Services Regulatory Authority	Provides commercial lease transparency; supports market and economic analysis.				Building Type≈	Lease Date, Lease Length	Address, Postcode	Low, short-term properties
	Tailte Éireann Prime 2 Dataset	Ordnance Survey Ireland	Base spatial reference dataset for governmental and commercial use.	Area			Building Type#, Use		Address, Polygon, Northing/ Easting	Map based 100%
	GeoHive Planning Application Points	GeoHive (Tailte Éireann)	Facilitates spatial analysis and planning oversight.	Area, Map Shape Length/ area			Construction Type		Address, Postcode, Northing/ Easting	Map based

*SEAI Area includes both internal and external building area.
^U-Values for doors, windows, floors, and roofs.
˙Fuel types include electricity grid, natural gas or oil.
˙Construction types refers to classification into new construction, extension, or renovation.
‡ Property Use refers to certain form details like 'over the shop', 'own door', 'purpose built', etc.
≈ Building Type here refers to the type of business such as retail, office, agriculture, etc.
Building Type here refers to 'terraced', 'detached', or 'semi-detached.

current SEAI EPC dataset. Finally, NACE categories: B (Mining and Quarrying), D (Electricity, Gas, Steam and Air Conditioning Supply), E (Water Supply; Sewerage, Waste Management and Remediation Activities), F (Construction), and U (Activities of Extraterritorial Organisations and Bodies) were completely unmapped, reflecting a lack of corresponding EPC-classified building typologies within the existing SEAI EPC dataset.

3.6.3. Limitations of the bottom-up and top-down mapping approaches

Both bottom-up and top-down mapping exercises are constrained by category-level allocation rules that do not establish building-level correspondence between datasets. Consequently, mapped proportions reflect structural compatibility rather than verified record linkage. Allocation across multiple categories required interpretive judgement based on descriptive similarity and relative category scale, introducing uncertainty that cannot be resolved without shared building identifiers.

3.7. EPC sample set quantification to GeoDirectory population dataset

Referencing Fig. 11, this section assesses how representative the EPC dataset is of the GeoDirectory population, with representativeness quantified using confidence level and margin of error (SD = 0.1). Overall, the comparison reveals a structurally uneven relationship between the SEAI EPC sample and the GeoDirectory population, indicating that EPC records reflect transactional and regulatory drivers rather than a uniform cross-section of the national non-domestic building stock.

Retail-related NEAP categories are markedly over-represented, comprising 40% of EPC records (36,042 units) but only 22% of the GeoDirectory population (34,086 units). This disparity reflects market-driven certification practices, whereby retail premises undergoing sale

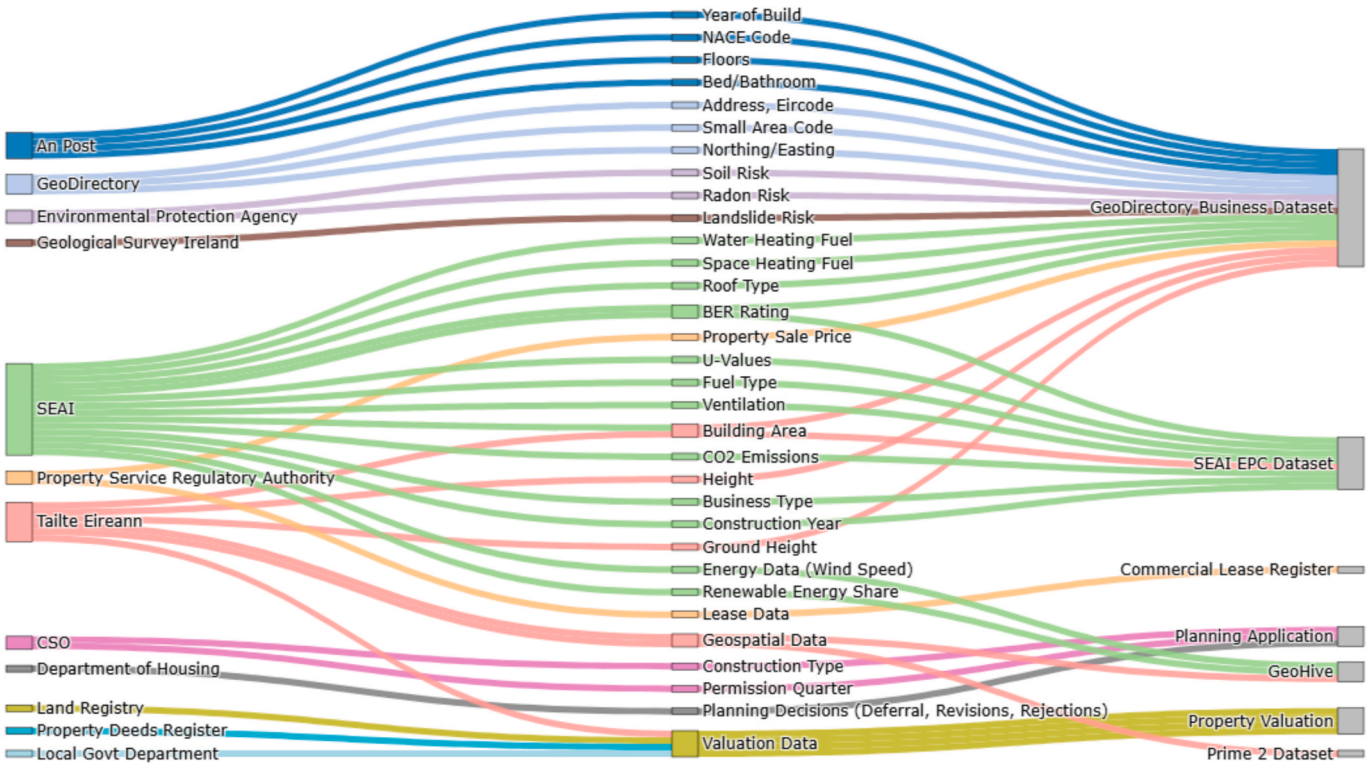


Fig. 10. Data Sources for key attributes of selected datasets (Source −> Attribute −> Dataset).

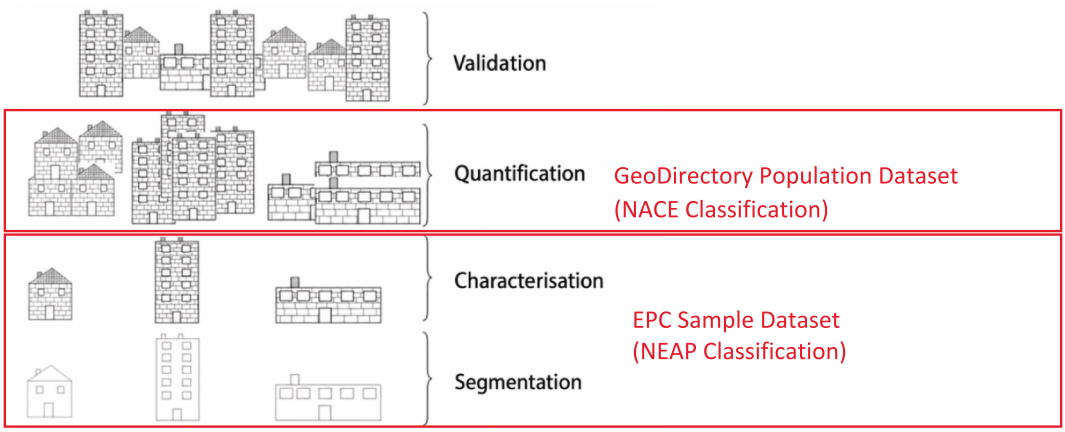


Fig. 11. Sample and Population Datasets useful to Bottom-up BSEM Approach [7].

or lease are more likely to require EPC assessment.

Office buildings exhibit comparatively strong statistical alignment. Office-related NEAP categories account for approximately 24% of the EPC dataset (22,841 units) and 22% of the GeoDirectory population (26,389 units), corresponding to a 99% confidence level with a 0.2% margin of error. This indicates good proportional representativeness, although some skew towards actively transacted offices likely persists.

Storage and warehouse buildings are also substantially over-represented, representing 12.2% of the EPC sample (11,638 units) compared with 4.1% of the GeoDirectory population (4,072 units). This distortion arises from the presence of four distinct storage-oriented NEAP categories that inflate their relative weight within the EPC classification structure.

Industrial buildings, by contrast, demonstrate strong proportional consistency; when heavy and specialised facilities, industrial buildings are included, this category comprises approximately 5% of the EPC dataset (4,861 units) compared with 4% of the GeoDirectory population (6,371 units), indicating good overall representation. Manufacturing and light industrial NEAP categories achieve a 99% confidence level, with margins of error of 0.54%.

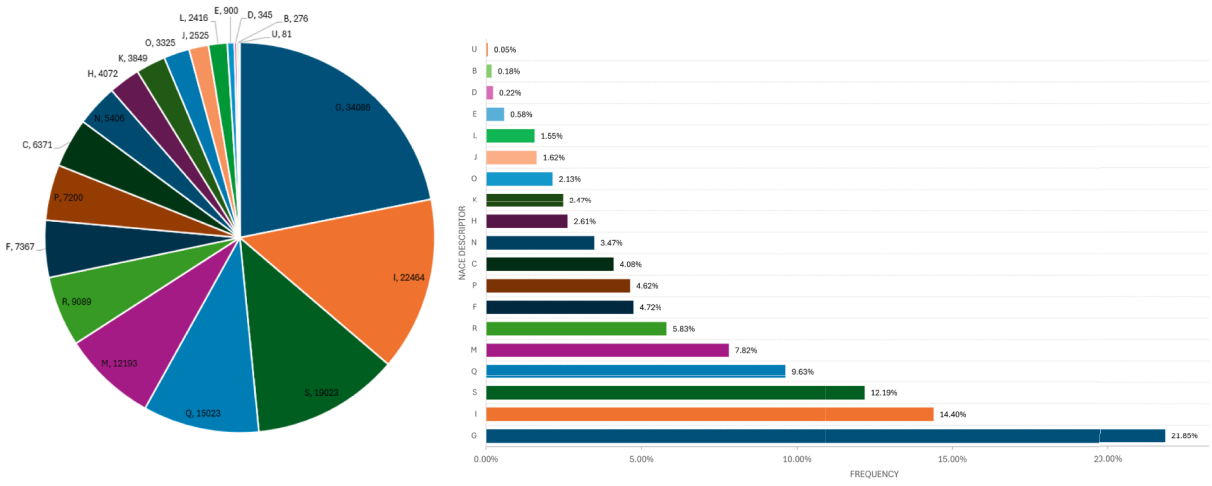
Accommodation and food service buildings also show relatively strong alignment, accounting for 11% of EPC records (10,547 units) and 14.4% of the GeoDirectory population (22,464 units), achieving a 99% confidence level with a 0.55% margin of error. However, quantitively only around half of hotels and restaurants appear in the EPC dataset, highlighting partial absolute coverage despite proportional similarity.

Educational buildings are under-represented in absolute terms but broadly proportional, representing 3% of EPC records (2,881 units) and 4.6% of the GeoDirectory population (7,200 units), with a 99% confidence level and 1.12% margin of error. This suggests limited penetration of EPC assessments rather than categorical bias.

Healthcare buildings are more significantly under-represented. They

Table 5
Relative Share of Non-residential Buildings classified by NACE Codes in Ireland in 2025 (GeoDirectory, 2025).

			Frequency %	Number of Buildings
NACE Description	G	WHOLESALE AND RETAIL TRADE; REPAIR OF MOTOR VEHICLES AND MOTORCYCLES	21.85%	34,086
	I	ACCOMMODATION AND FOOD SERVICE ACTIVITIES	14.40%	22,464
	S	OTHER SERVICE ACTIVITIES	12.19%	19,023
	Q	HUMAN HEALTH AND SOCIAL WORK ACTIVITIES	9.63%	15,023
	M	PROFESSIONAL, SCIENTIFIC AND TECHNICAL ACTIVITIES	7.82%	12,193
	R	ARTS, ENTERTAINMENT AND RECREATION	5.83%	9089
	F	CONSTRUCTION	4.72%	7367
	P	EDUCATION	4.62%	7200
	C	MANUFACTURING	4.08%	6371
	N	ADMINISTRATIVE AND SUPPORT SERVICE ACTIVITIES	3.47%	5406
	H	TRANSPORTATION AND STORAGE	2.61%	4072
	K	FINANCIAL AND INSURANCE ACTIVITIES	2.47%	3849
	O	PUBLIC ADMINISTRATION AND DEFENCE; COMPULSORY SOCIAL SECURITY	2.13%	3325
	J	INFORMATION AND COMMUNICATION	1.62%	2525
	L	REAL ESTATE ACTIVITIES	1.55%	2416
	E	WATER SUPPLY; SEWERAGE, WASTE MANAGEMENT AND REMEDIATION ACTIVITIES	0.58%	900
	D	ELECTRICITY, GAS, STEAM AND AIR CONDITIONING SUPPLY	0.22%	345
	B	MINING AND QUARRYING	0.18%	276
	U	ACTIVITIES OF EXTRATERRITORIAL ORGANISATIONS AND BODIES	0.05%	81



(A) by absolute frequency (B) by relative frequency

Fig. 12. (A & B) Relative Share of Non-Residential Buildings in Ireland classified by NACE codes in absolute and relative frequency in 2025.

comprise 5% of the EPC dataset (4,843 units) compared with 9.6% of the GeoDirectory population (15,023 units). While statistical significance is achieved (99% confidence; 0.92–1.53% margin of error), the shortfall likely reflects the exclusion of hospitals and long-term care facilities from NEAP processes or their treatment within alternative reporting frameworks [39].

Recreational buildings are also substantially under-represented, accounting for 1.6% of EPC records (1,583 units) versus 5.8% of the GeoDirectory population (9,089 units). Despite achieving statistical confidence at the (99% level; 1.77% margin of error), the absolute deficit indicates a structural absence of these typologies within EPC-driven data collection.

Other service buildings—including gyms, laundrettes, and non-governmental organisations—are among the least represented categories, comprising just 0.5% of the EPC dataset (446 units) compared with 12.2% of the GeoDirectory population (19,023 units). This indicates that small service enterprises, while numerous in the national stock, rarely trigger EPC requirements.

Government and administrative buildings are almost entirely absent from the EPC dataset, representing only 0.07% of records (63 units) compared with 2.1% of the GeoDirectory population (3,325 units). The disparity is sufficiently large that statistical representativeness is not achieved, confirming that public-sector buildings are not adequately captured through EPC-based data.

3.8. *Interoperability of temporal datasets*

The SEAI EPC dataset, Lease Register, and Planning Applications datasets are the only data sources that record some form of temporal attributes. The EPC dataset lists the construction year of the building, while the lease register records the lease date and lease length of buildings. Planning applications only capture application timelines of when the application for existing or new construction was applied and then granted. This discrepancy complicates the interoperability of temporal dynamics among the datasets. Even with the limited data available from the above-mentioned datasets, the Commercial Lease Register dataset (Table 2; Pt. 7) employs partially overlapping but non-standardised building typologies such as *retail*, *office*, and *industrial*, while the planning applications only has the application number without any significant identifier of individual buildings. These differences in classification terminology hinder typology mapping and prevent consistent aggregation temporal attributes of non-residential building data across various sources.

Table 6
Relative Share of Non-Residential Buildings in Ireland in 2024 classified by NEAP "building types" (EPC, 2025).

		Frequency %	Number of Buildings
EPC/NEAP Building Types	Retail	26.97%	25,823
	Office	16.21%	15,523
	Retail/Financial and Professional services	10.67%	10,219
	Offices and Workshop businesses	7.64%	7318
	Warehouse and storage	7.21%	6908
	Restaurant/public house	6.60%	6315
	Workshops/maintenance depot	2.94%	2817
	Restaurants and Cafes/ Drinking Establishments/ Takeaways	2.85%	2731
	Retail warehouses	2.54%	2431
	Storage or Distribution	2.40%	2299
	Industrial process building	1.58%	1515
	Hotels	1.57%	1501
	Primary health care buildings	1.23%	1181
	Community/day centre	1.20%	1149
	Non-residential Institutions: Primary Education	0.95%	907
	Non-residential Institutions: Community/Day Centre	0.80%	768
	Non-residential Institutions: Primary Health Care Building	0.68%	651
	Primary school	0.67%	638
	Nursing residential homes and hostels	0.65%	620
	Sports centre/leisure centre	0.62%	590
	Non-residential Institutions: Post-primary Education	0.56%	537
	General Industrial and Special Industrial Groups	0.55%	529
	General Assembly and Leisure Night Clubs and Theatres	0.52%	497
	Residential Institutions: Hospitals and Care Homes	0.40%	380
	Secondary school	0.27%	261
	Residential Institutions: Universities and colleges	0.26%	253
	Further education universities	0.23%	223
	Launderette	0.22%	209
	Social clubs	0.18%	171
	Theatres/cinemas/music halls and auditorium	0.13%	122
	Libraries/museums/galleries	0.11%	107
	Others: Emergency services	0.11%	104
	Hospital	0.10%	94
	Non-residential Institutions: Libraries Museums and Galleries	0.08%	75
	Telephone exchanges	0.07%	68
	Others: Miscellaneous 24hr activities	0.06%	56
	Residential Institutions: Residential Primary schools	0.04%	35
	Residential Institutions: Residential Post-primary schools	0.03%	27
	Crown and county courts	0.03%	25
	Airport terminals	0.02%	22
	Sports ground arena	0.02%	21
	Others: Car Parks 24 hrs	0.01%	9
	Non-residential Institutions: Law Courts	0.01%	5
	Bus station/train station/ seaport terminal	0.01%	5
	Prisons	0.00%	3
	Others: Passenger terminals	0.00%	3

4. Discussion and policy recommendations

This scoping review diagnoses the structural data barriers limiting Ireland's ability to implement evidence-based non-residential building energy policy. While seven datasets relevant to the non-residential building stock were identified, fragmented custodianship, incompatible classification systems, and the absence of shared building identifiers prevent integration into a coherent national register.

Although the SEAI EPC and the GeoDirectory dataset provide sufficient depth and coverage for building stock analysis, combined use is constrained by misalignment between activity- and building-based classifications, as well as sampling bias. These limitations undermine development of representative typologies, tracking renovation progress, and meeting EPBD Articles 16 and 22 requirements. The findings indicate that Ireland's primary challenge is not data availability, but data governance.

4.1. *Establishing a unified data framework*

A key outcome of this review is the recognition that Ireland lacks a coherent framework for integrating existing datasets into a single, authoritative non-residential building register. To enable interoperability, a building-level unique identifier – derived from or extending the existing *Eircode* system – should be universally adopted across agencies. For the non-residential sector, this Eircode based identifier does not pose data privacy issues as all businesses and commercial buildings list their addresses publicly. In alignment with Article 16 of the EPBD [4], this would allow one-to-one building matching, longitudinal tracking, and spatial validation while supporting linkage to valuation, lease, and planning datasets.

Parallel to this, a harmonised building-use typology is required to reconcile current inconsistencies between economic classifications (e.g., NACE) and physical or operational typologies (e.g., SEAI property types/NEAP building types). A dual-layer schema combining *activity* and *form* attributes would enable both economic and energy analyses and should be aligned with European typology standards such as TABULA/ Episcope.

Designating authoritative or "truth-originator" datasets for each major attribute domain would strengthen governance and data reliability. For example,

- Tailte Éireann / GeoDirectory could be the authority for spatial form and geolocation;
- SEAI EPC for envelope and system data; and
- Central Statistics Office (CSO/PSRA for operational and market-related attributes.

This model would ensure traceability, prevent duplication, and clarify custodianship. These priorities collectively form the conceptual basis for an integrated data architecture illustrated in Fig. 16. As shown in Fig. 16, the key steps towards building a unified building data framework require the following:

- Ensure complete and standardised postcode recording across all non-residential building datasets.
- Establish a unique building identifier schema capable of distinguishing individual non-residential units within multi-unit buildings, potentially through an extension of the postcode incorporating floor level and unit number.
- Develop a harmonised building-use typology integrating both economic activity classifications and energy-relevant building typologies.
- Maintain authoritative "truth-originator" datasets for key domains including form, envelope, operational, and system characteristics, which can subsequently feed into an integrated national building register for stock energy modelling and policy analysis.

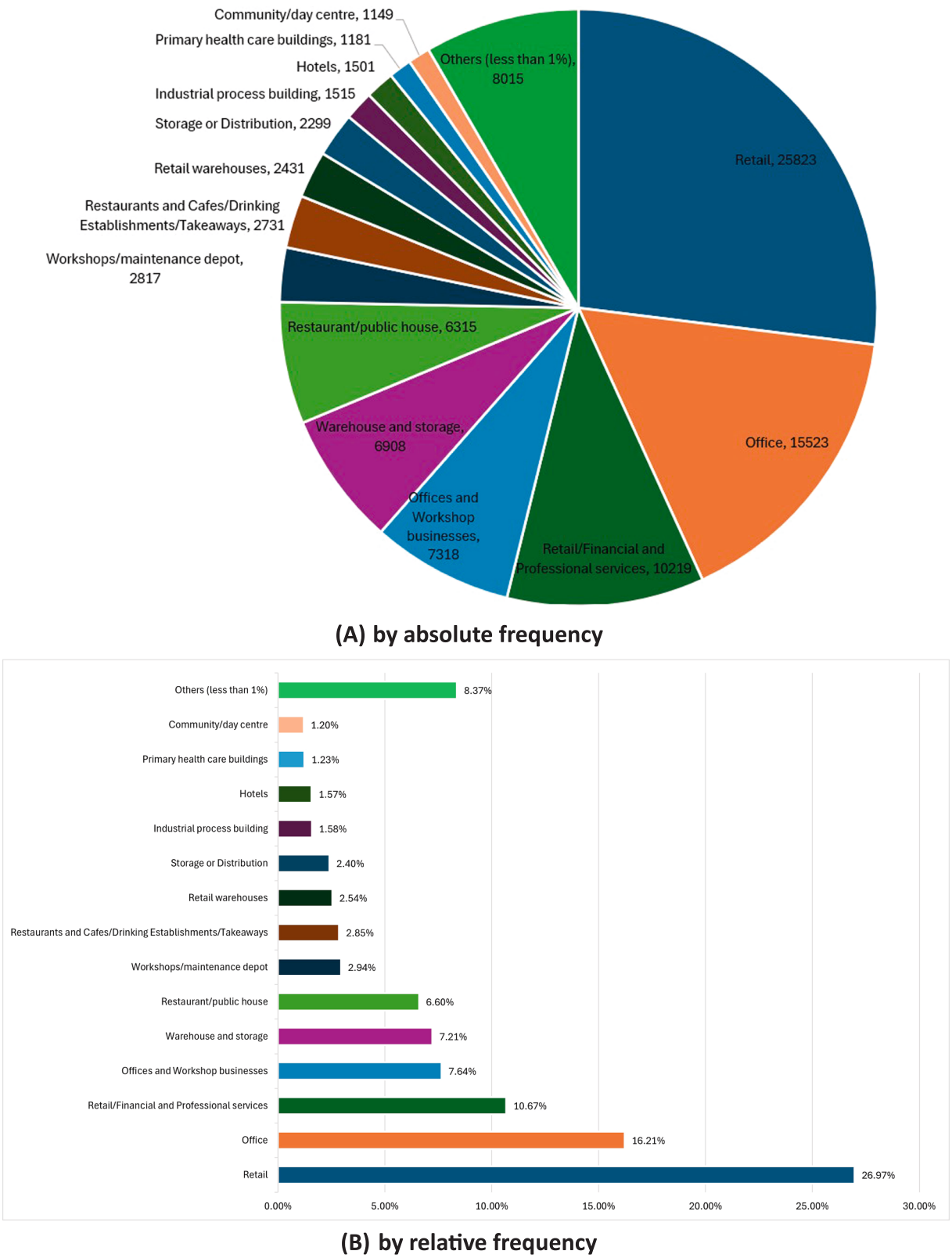


Fig. 13. (A & B) Relative Share of Non-Residential Buildings in Ireland classified by NEAP Building Types in absolute and relative frequency in 2024.

4.2. Toward a transferable EU framework for building data governance

The specific interoperability and governance challenges identified in this study are shaped by the Irish administrative, institutional, and data-governance context, particularly the absence of a universally adopted building identifier and the distribution of non-residential building data across multiple custodians and classification systems.

The findings of this review reveal that the fragmentation, inconsistent custodianship, and incomplete coverage of Ireland's non-residential building datasets are not isolated conditions but part of a broader

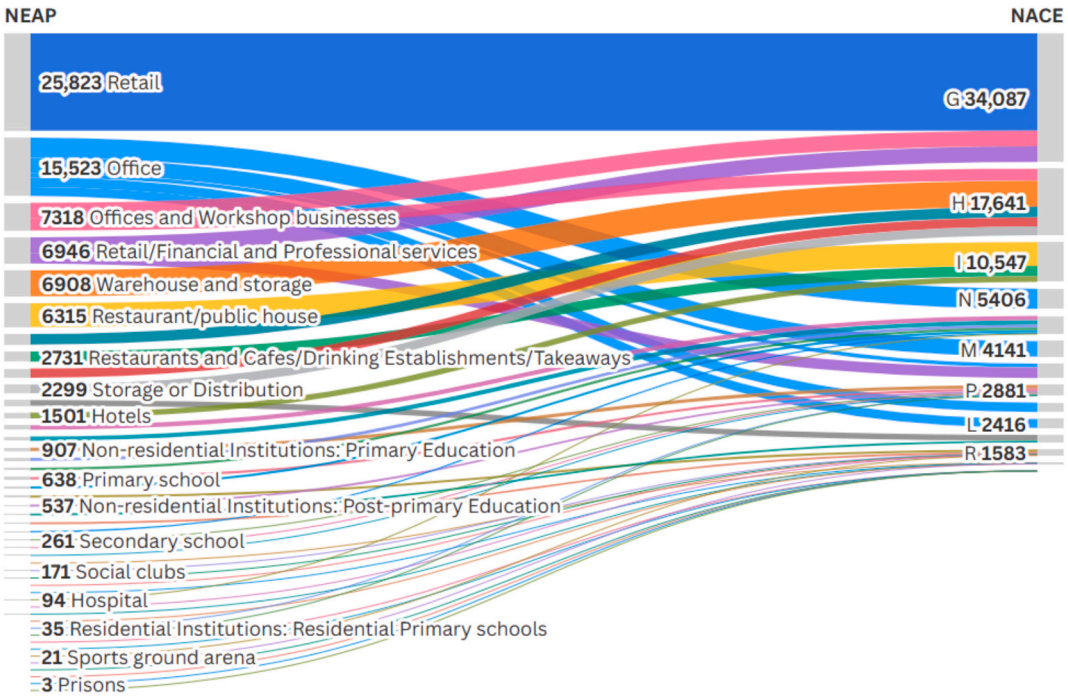


Fig. 14. Quantitative Mapping: SEAI EPC (NEAP) to GeoDirectory (NACE) – cross reference Table 7.

European challenge in aligning statistical, spatial, and energy data systems [28]. To move beyond a national diagnosis, these insights can be structured into a transferable framework for evaluating and improving data interoperability in building-stock databases. This framework comprises four analytical dimensions that together describe the readiness of national data infrastructures for integration:

- **Institutional dimension** – clarity of custodianship, mandates, and update responsibilities across agencies;
- **Technical dimension** – compatibility of attributes, formats, metadata, and data-exchange protocols that enable cross-system interoperability;
- **Semantic dimension** – consistency of variable definitions, typologies, and classifications relevant to building form, use, and performance; and
- **Operational dimension** – temporal resolution, completeness, and accessibility of datasets that determine their suitability for longitudinal and spatial modelling.

Applying this framework provides a replicable diagnostic approach that other MSs can adapt when assessing their own building data ecosystems.

4.2.1. *Semantic limitations: the case of NACE classification*

One prominent semantic constraint identified, is the widespread reliance on the NACE (Nomenclature of Economic Activities) classification to describe non-residential building functions. The study does not assume a one-to-one relationship between economic activity, building typology, and operational energy demand; rather, the focus is on the suitability of classification systems for representative building segmentation and interoperability within stock energy modelling. While operational energy consumption is influenced by behavioural, economic, climatic, and occupancy-related factors, representative typologies remain necessary for segmentation and stock-level energy modelling within bottom-up BSEM approaches. Although NACE provides harmonisation across MSs and facilitates alignment with national accounts, it is too broad and activity-oriented, when used in isolation, to support meaningful RB characterisation or detailed operational energy modelling [36]. This semantic mismatch represents a systemic constraint on data-driven building policy across the EU, rather than an issue unique to Ireland. For instance, the NACE category "G – Wholesale and Retail Trade" spans buildings ranging from small retail units to large distribution centres with markedly different physical and operational energy-use profiles. This lack of typological granularity limits alignment between NACE economic activity categories and NEAP building typologies, thereby reducing model precision and constraining policy targeting. Considering the EPBD's emphasis on national building databases and representative RBs, the suitability of NACE as a standalone descriptor merits critical review. A shift toward a building-oriented typology, harmonised at EU level but grounded in physical and operational attributes, would better support energy-performance assessment, renovation tracking, and emissions forecasting.

4.2.2. *Benchmarking European practice*

Comparison with other Member States and the UK highlights that Ireland's current approach remains behind European best practice, particularly in terms of data completeness, accessibility, and institutional coherence. Mature national registers demonstrate that effective building-stock databases are underpinned by four interrelated diagnostic dimensions:

i. Clear and accountable custodianship, as exemplified by Denmark's Building and Dwelling Register (BBR) [40], ensures national consistency through mandatory reporting and stable institutional ownership.
ii. Legislative and regulatory integration plays a critical enabling role, as illustrated by France's digital building logbook [41] and Germany's iBRoad [42] and DGNB Building Resource Passport pilots [43], where statutory embedding supports longitudinal data use and policy alignment.
iii. The United Kingdom's open EPC database [21] demonstrates the value of machine-readable, standardised data structures, with open APIs enabling reuse in analytical tools such as the London Building Stock Model. Finally,
iv. EU-level initiatives including openDBL [44] and Building As Material Banks (BAMB) 2020 [45] emphasise the operational

Table 7
Statistical Significance of SEAI EPC Sample Dataset over GeoDirectory Population Dataset.

		Sample Dataset			Population Dataset					
		SEAI EPC			GeoDirectory			(p = 0.1; Min S.D.)	(p = 0.5; Max S.D.)	
		Quantity	%	Closest NACE Category/ Equivalent	Quantity	%	Sample Size Mapped	Margin of Error at 99% confidence level (z = 2.58)	Margin of Error at 99% confidence level (z = 2.58)	Comment
EPC/NEAP Building Category*	Retail	36,042	37.64%	G – Wholesale and Retail Trade; Repair of Motor Vehicles and Motorcycles	34,086	21.85	34,086	0.00%	0.00%	Full coverage of Population Dataset; All combined retail NEAP categories over-performs Population dataset; NEAP heavily biased toward EPC-assessed retail units (shops being sold/leased).
	Office	22,841	23.86%	M – Professional, Scientific and Technical Activities - 12,193	26,389	7.82	22,841	0.19%	0.31%	Many offices are not EPC-rated unless being sold/ leased → missing stock.
				N – Administrative and Support Service Activities – 5406		3.47				
				K – Financial and Insurance Activities – 3849		2.47				
				J – Information and Communication – 2525		1.62				
				L – Real Estate Activities – 2416		1.55				
	Storage	11,638	12.16%	H – Transportation and Storage	4072	2.61	4072	0.00%	0.00%	Excess in SEAI Dataset with 4 separate NEAP categories.
	Industrial	4861	5.08%	C – Manufacturing	6371	4.08	4861	0.54%	0.90%	Manufacturing + light workshops aligns well; Combined with Heavy/specialised industrial building –> SEAI EPC exceeds coverage.
	Hotel/Restaurant	10,547	11.02%	I – Accommodation and Food Service Activities	22,464	14.4	10,547	0.55%	0.91%	Good alignment in representation within each dataset (10–15%); however only 50% represented in sample dataset as compared to population.
	Education	2881	3.01%	P – Education	7200	4.62	2881	1.12%	1.86%	Under-represented but broadly proportional.
	Healthcare	4843	5.06%	Q – Human Health and Social Work Activities	15,023	9.63	4843	0.92%	1.53%	Under-represented; Hospitals and care homes often excluded from NEAP or assessed under different systems.
	Recreation	1583	1.65%	R – Arts, Entertainment and Recreation	9089	5.83	1583	1.77%	2.95%	Slightly lower representation in NEAP.
	Other Services	446	0.47%	S – Other Service Activities	19,023	12.19	446	3.62%	6.04%	Highly under-represented; Includes salons, gyms, laundrettes, NGOs, charities → NEAP rarely captures these.
	Government/ Administrative	63	0.07%	O – Public Administration and Defence; Compulsory Social Security	3325	2.13	63	9.66%	16.10%	NEAP does not capture state-owned buildings well.

*NEAP Categories Aggregated, Full table available as supplementary material (Appendix B).

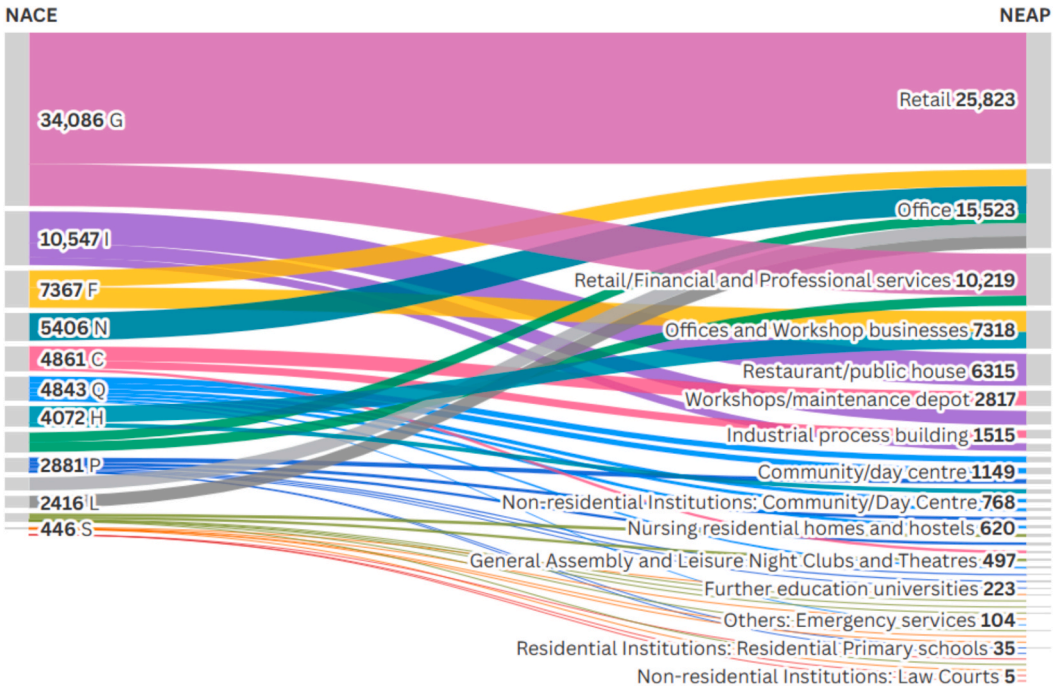


Fig. 15. Quantitative Mapping: GeoDirectory (NACE) to SEAI EPC (NEAP).

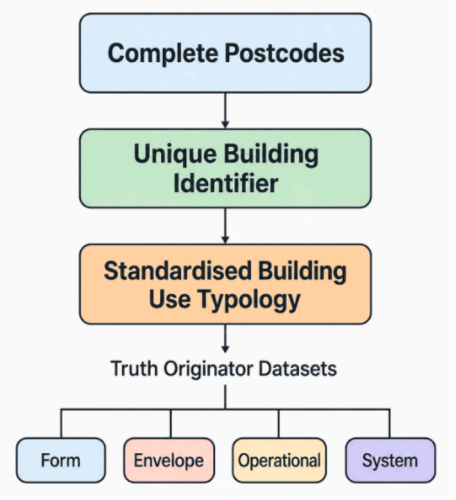


Fig. 16. Conceptual framework for dataset integration.

importance of sustained updating mechanisms and open interfaces that allow national systems to evolve and interoperate.

Viewed through these four dimensions, Ireland's fragmented custodianship arrangements and inconsistent identifier systems indicate persistent weaknesses across the institutional and semantic layers, despite the presence of underlying technical capacity.

4.3. *Recommendations for policy implementation*

Based on the findings of this scoping review, four priority actions are recommended to enable effective implementation of EPBD Articles 16 and 22 [4,5] and strengthen evidence-based non-residential building energy policy in Ireland. The interoperability and governance challenges identified in this study are shaped by the Irish administrative and data-governance context, particularly the absence of a universally adopted building identifier and the distribution of non-residential building data across multiple custodians and classification systems. While the specific institutional arrangements may differ across Member States, the broader methodological and interoperability challenges are relevant to other national contexts where integrated building-data infrastructures remain under development.

Introduce a mandatory building-level identifier system: establish a universal, persistent identifier for every non-residential building in Ireland, preferably derived from or linked to the existing Eircode system but unique at structure or unit level. This identifier should be mandated through regulation and adopted consistently across public agencies to enable reliable dataset linkage, longitudinal tracking, and spatial validation of building attributes.

Develop a harmonised building-use typology for energy analysis: create an integrated classification framework that reconciles economic activity classifications (e.g. NACE) with building-based energy typologies (e.g. NEAP). This harmonised schema should prioritise physical, operational, and occupancy-related building characteristics relevant to representative energy performance assessment, align with European typology initiatives (e.g. TABULA/Episcope), and support consistent Reference Building characterisation across datasets.

Define clear governance and data stewardship responsibilities: formally assign custodianship ("truth-originator" status) for each major building attribute domain (e.g. spatial form, envelope and systems, operational use), supported by standardised metadata, version control, and validation protocols. Clear governance arrangements are essential to ensure data reliability, accountability, and coordinated updating across institutions.

Enable interoperable data access and continuous updating: implement open, machine-readable interfaces (e.g. APIs), automated data exchange, and synchronised update cycles across building-related datasets. These measures should be supported by minimum interoperability standards and long-term resourcing commitments to ensure that national building-data infrastructures remain current, scalable, and suitable for policy monitoring and stock modelling purposes.

Although focused on Ireland, the methodological approach developed in this study may be transferable to other EU Member States seeking to strengthen non-residential building-data integration. The proposed implementation pathway involves first undertaking a national inventory of available non-residential building datasets and evaluating them according to form, envelope, system, and operational domains consistent with the Corgnati RB Framework [9]. Subsequent priorities would include establishing interoperable identifier systems, harmonising energy-relevant building typologies, and integrating datasets within coordinated national building-stock registers appropriate to each Member State's institutional and governance structure. At a broader EU level, these measures would support alignment with the data-integration objectives of EPBD Articles 16 and 22.

4.4. Limitations and future work

This scoping review achieved its aim of systematically identifying and mapping electronic data repositories and their recorded attributes relevant to Ireland's non-residential building stock (Objectives i–ii; RQs 1–3). However, limitations such as differences in classification systems and recorded attributes across datasets were documented (Objectives ii and v; RQs 2, 3, and 6). The study did not undertake schema harmonisation or reclassification. Consequently, the suitability of NACE for non-residential building energy analysis was assessed at a structural level rather than thorough empirical testing.

Different unique identifiers across datasets were identified and evaluated (Objective iii; RQ 4), but record-level linkage, de-duplication, and validation were outside the scope of the review. As a result, the practical effectiveness of identifiers such as Eircode for integrating datasets needs to be confirmed in future research.

Data gaps and inconsistencies were identified based on documented metadata, attribute availability, and custodial descriptions (Objective iv; RQ 5), rather than through quantitative testing of coverage, temporal continuity, or internal consistency. Findings related to representativeness and modelling readiness should therefore be interpreted as indicative.

Future work will move beyond mapping toward empirical integration. This will include testing identifier-based linkages across datasets, harmonising building-use classifications for energy relevance, and validating attribute completeness and temporal coverage. These steps are required to operationalise the findings of the scoping review and to support the development of representative buildings and an EPBD Article 22 [5] –compliant national non −residential building stock energy model. Policymakers should therefore interpret the findings as indicative of systemic readiness rather than as a definitive quantification of national building stock performance.

5. Conclusion

This study systematically reviewed electronic datasets relevant to Ireland's non-residential building stock to evaluate their suitability for bottom-up building stock energy modelling and implementation of the recast Energy Performance of Buildings Directive (EPBD). Seven core datasets were identified, of which the SEAI Energy Performance Certificate dataset and GeoDirectory provide the most comprehensive coverage of building, operational, and energy-related characteristics. However, their combined use is constrained by fragmented custodianship, inconsistent identifiers, incompatible classification systems, and uneven dataset completeness.

The findings demonstrate that the principal barrier to developing an interoperable national non-residential building database in Ireland is not data availability, but deficiencies in data governance and semantic interoperability. Significant incompatibilities were identified between the NACE economic activity classification system and NEAP building energy typologies, limiting integration of economic and energy datasets for stock modelling and policy monitoring. The analysis also revealed systematic over-representation of frequently transacted building categories within EPC datasets, alongside under-representation of public-sector, healthcare, recreational, and small-service buildings.

Based on these findings, the study proposes a transferable diagnostic framework for assessing building-data interoperability across institutional, technical, semantic, and operational dimensions. The results support implementation of Articles 16 and 22 of the recast EPBD by highlighting the need for harmonised building identifiers, interoperable machine-readable databases, and energy-relevant building classification systems.

Although focused on Ireland, the methodological approach and interoperability challenges identified are relevant to other EU Member States seeking to establish integrated national building data infrastructures for evidence-based retrofit planning and building stock decarbonisation.

CRediT authorship contribution statement

Ankur Mitra: Writing – original draft, Methodology, Formal analysis. **Kriti Bhalla:** Writing – review & editing, Validation, Methodology. **Ciara Ahern:** Writing – review & editing, Validation, Supervision.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Ankur Mitra reports financial support was provided by MaREI the SFI Research Centre for Energy Climate and Marine. Ciara Ahern reports financial support was provided by Sustainable Energy Authority of Ireland. If there are other authors, they declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

The authors wish to thank the Irish Building Stock Observatory (www.ibso.ie) team for their expert input to this work.

Funding

This research was part funded by the Sustainable Energy Authority of Ireland, RD&D PREDICT project [Grant No. 23/RDD/1046} and the MaREI, the Research Ireland Centre for Energy, Climate and Marine [Grant No.: 12/RC/2303-P2].

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.enbuild.2026.117734.

Data availability

The data that has been used is confidential.

References

[1] United Nations Environment Programme. 2021 Global Status Report for Buildings and Construction: Towards a Zero-emission, Efficient and Resilient Buildings and Construction Sector. https://globalabc.org/resources/publications/2021-global-status-report-buildings-and-construction, 2021 (accessed 19 September 2025).
[2] European Commission, The European Green Deal. https://commission.europa.eu/strategy-and-policy/priorities-2019-2024/european-green-deal_en, 2024 (accessed 29 September 2025).
[3] European Commission, 2030 climate targets. https://climate.ec.europa.eu/eu-action/climate-strategies-targets/2030-climate-targets_en, 2025 (accessed 10 October 2025).
[4] European Commission, Data exchange (Article 16) — Annex 6. https://energy.ec.europa.eu/document/download/6f7e3300-dee9-4c32-b73e-079676ed7d6e_en?

filename=Data%20exchange%20%28Article%2016%29%20-%20annex%206.pdf, 2025 (accessed 29 September 2025).
[5] European Commission, Databases for the energy performance of buildings (Article 22) — Annex 5. https://energy.ec.europa.eu/document/download/74388e1e-ca33-4ae0-9e4a-900953873cf2_en?filename=Databases%20for%20the%20energy%20performance%20of%20buildings%20%28Article%2022%29%20-%20annex%205.pdf, 2025 (accessed 29 September 2025).
[6] European Commission, EU Building Stock Observatory. https://energy.ec.europa.eu/topics/energy-efficiency/energy-performance-buildings/eu-building-stock-observatory_en, 2025 (accessed 6 October 2025).
[7] E. Mata, A.S. Kalagasidis, F. Johnsson, Building-stock aggregation through archetype buildings: France, Germany, Spain and the UK, Build. Environ. 81 (2014) 270–282, https://doi.org/10.1016/j.buildenv.2014.06.013.
[8] C. Ahern, B. Norton, A generalisable bottom-up methodology for deriving a residential stock model from large empirical databases, Energ. Buildings 215 (2020) 109684, https://doi.org/10.1016/j.enbuild.2020.109684.
[9] S.P. Corgnati, E. Fabrizio, V. Monetti, Reference buildings for cost optimal analysis: method of definition and application, Appl. Energy 102 (2013) 983–993, https://doi.org/10.1016/j.apenergy.2012.06.001.
[10] N. Milojevic-Dupont, F. Wagner, F. Nachtigall, J. Hu, G.B. Brüser, M. Zumwald, F. Biljecki, N. Heeren, L.H. Kaack, P.P. Pichler, F. Creutzig, EUBUCCO v0.1: European building stock characteristics in a common and open database for 200+ million individual buildings, Sci. Data 10 (2023) 147, https://doi.org/10.1038/s41597-023-02040-2.
[11] OpenStreetMap Ireland. https://www.openstreetmap.ie/, n.d. (accessed 10 October 2025).
[12] EPISCOPE Project, Energy Performance Indicators for Building Stocks. https://episcope.eu/welcome/, n.d. (accessed 10 October 2025).
[13] EPISCOPE Project, Building Typology (TABULA structure) — Concept. https://episcope.eu/building-typology/tabula-structure/concept/, n.d. (accessed 10 October 2025).
[14] ODYSSEE-MURE Project, Energy Efficiency Indicators and Policies. https://www.odyssee-mure.eu/project.html, 2026 (accessed 10 October 2025).
[15] ODYSSEE-MURE Project, Key energy and CO_2 indicators. https://www.indicators.odyssee-mure.eu/key-indicators.html, n.d. (accessed 10 October 2025).
[16] ODYSSEE-MURE Project, Energy efficiency measures database. https://www.measures.odyssee-mure.eu/, n.d. (accessed 10 October 2025).
[17] ODYSSEE-MURE Project, Ireland — Efficiency trends, policies & profiles. https://www.odyssee-mure.eu/publications/efficiency-trends-policies-profiles/ireland.html, 2026 (accessed 10 October 2025).
[18] ENTRANZE Project. https://www.entranze.eu/, n.d. (accessed 10 October 2025).
[19] ENTRANZE Project, Data Tool and Guidelines for Energy Efficiency Policies. https://entranze.enerdata.net/documents/guideline-datatool.pdf, 2013 (accessed 10 October 2025).
[20] European Commission CORDIS, "openDBL – ONE STEP OPEN DBL SOLUTION" (Grant ID 101092161). https://cordis.europa.eu/project/id/101092161, n.d. (accessed 10 October 2025).
[21] Greater London Authority, London Building Stock Model. https://data.london.gov.uk/dataset/london-building-stock-model/, 2023 (accessed 10 October 2025).
[22] Sustainable Energy Authority of Ireland, Extensive survey of the commercial buildings stock in the Republic of Ireland. https://www.seai.ie/sites/default/files/publications/Extensive-Survey-of-Commercial-Buildings-Stock-in-the-Republic-of-Ireland.pdf, 2015 (accessed 29 September 2025).
[23] An Post. https://www.anpost.com/, n.d. (accessed 19 October 2025).
[24] Tailte Éireann. https://tailte.ie/, n.d. (accessed 19 October 2025).
[25] GeoDirectory. https://www.geodirectory.ie/, n.d. (accessed 10 October 2025).
[26] Central Statistics Office, Report of the Steering Group on the development of official statistics for the Irish commercial property market. https://www.cso.ie/en/media/csoie/methods/methodologicalresearch/commercialproperty/CPSS_Steering_Group_Report_Final.pdf, 2017 (accessed 19 September 2025).
[27] Department of Enterprise, Trade and Employment, Ireland. Decarbonising the Commercial Built Environment – Draft Roadmap Work. https://enterprise.gov.ie/en/publications/publication-files/decarbonising-the-commercial-built-environment-6-march-2024.pdf, 2024 (accessed 10 October 2025).
[28] C. Franco, C. Maduta, A.M. Martínez, P. Florio, D. D'Agostino, V. Todeschi, A. Kona, Identification of essential variables for the estimation of energy demand in buildings at European scale, Journal of Building Engineering 119 (2026) 115222, https://doi.org/10.1016/j.jobe.2026.115222.
[29] S. Mak, A. Thomas, Steps for conducting a scoping review, J. Grad. Med. Educ. 14 (5) (2022) 565–567, https://doi.org/10.4300/JGME-D-22-00621.1.
[30] H. Arksey, L. O'Malley, Scoping studies: towards a methodological framework, Int. J. Soc. Res. Methodol. 8 (1) (2005) 19–32, https://doi.org/10.1080/1364557032000119616.
[31] M.D.J. Peters, C.M. Godfrey, H. Khalil, P. McInerney, D. Parker, C.B. Soares, Guidance for conducting systematic scoping reviews, Int. J. Evid. Based Healthc. 13 (3) (2015) 141–146, https://doi.org/10.1097/XEB.0000000000000050.
[32] Irish Building Stock Observatory. https://ibso.ie/, n.d. (accessed 29 January 2026).
[33] Ireland's Open Data Portal. https://data.gov.ie/, n.d. (accessed 10 October 2025).
[34] GeoDirectory Dataset, GeoDirectory & An Post. Obtained under license (accessed 21 May 2025).
[35] Sustainable Energy Authority of Ireland, Energy Performance Certificate Dataset. Obtained under license (accessed 7 February 2025).
[36] Eurostat, NACE Rev. 2 – Statistical classification of economic activities in the European Community: Methodologies and working papers (KS-RA-07-015-EN). Luxembourg: Office for Official Publications of the European Communities. https://ec.europa.eu/eurostat/documents/3859598/5902521/KS-RA-07-015-EN.PDF.pdf/dd5443f5-b886-40e4-920d-9df03590ff91?t=1414781457000, 2008 (accessed 19 October 2025).
[37] Sustainable Energy Authority of Ireland, NEAP Guidance Document Version 1.0 — Non-Domestic Energy Assessment Procedure. https://www.seai.ie/sites/default/files/ber/support-for-ber-assessors/neap/NEAP-Guidance-Document.pdf, 2022 (accessed 10 October 2025).
[38] Sustainable Energy Authority of Ireland, SBEM-ie Technical Manual Version 5.5.h. https://www.seai.ie/sites/default/files/ber/support-for-ber-assessors/neap/SBEMie_Technical_Manual_v5.5.h.pdf, 2018 (accessed 10 October 2025).
[39] Sustainable Energy Authority of Ireland. Display Energy Certificates (DEC). https://www.seai.ie/sites/default/files/publications/Public-Buildings-Technical-Manual.pdf, 2023 (accessed 29 January 2026).
[40] Statistics Denmark, The Central Register of Buildings and Dwellings (BBR). https://ncrr.au.dk/danish-registers/the-central-register-of-buildings-and-dwellings-bbr, n.d. (accessed 10 October 2025).
[41] BUILD UP, Business case for digital building logbooks (CLEA). https://build-up.ec.europa.eu/en/resources-and-tools/publications/business-case-digital-building-logbooks-clea, 2025 (accessed 10 October 2025).
[42] European Climate, Infrastructure and Environment Executive Agency (CINEA), iBRoad Project. https://cinea.ec.europa.eu/featured-projects/ibroad_en, n.d. (accessed 10 October 2025).
[43] Deutsche Gesellschaft für Nachhaltiges Bauen (DGNB), Building Resource Passport – Towards Circular Construction. https://www.dgnb.de/en/sustainable-building/circular-building/building-resource-passport, n.d. (accessed 10 October 2025).
[44] OpenDBL Project, Open Digital Building Logbook Platform. https://www.opendbl.eu/, n.d. (accessed 10 October 2025).
[45] Buildings As Material Banks (BAMB 2020) Project, Circular Building Resources. https://www.bamb2020.eu/, n.d. (accessed 10 October 2025).