Research Paper: Bridging the Domain Gap in Historic Building Data Analysis

1. Introduction: The Institutional and Policy Framework

This paper presents research conducted by ICOMOS Ireland's **NSCES+CC** (National Scientific Committee on Energy, Sustainability, and Climate Change). The committee's primary focus is on the usage and conservation of energy within historic structures, specifically identifying potential for energy improvements while monitoring for adverse impacts on physical fabric, character or cultural significance.

The results detailed here represent several years of investigation. The research has been presented both nationally (May 2024) and internationally (Estonia December 2025) and has contributed to the international legislative process. The research was instrumental in providing Irish data, informing discussions within a specialised sub-committee of **ISCES** (the International mirror Scientific Committee of the NSCES+CC) dedicated to the Energy Performance of Buildings Directive (**EPBD**). This sub-committee evaluated iterative drafts of the EPBDs to understand the implications of the "recast" **EPBD**—which came into force in May 2024—ensuring that energy reduction goals do not compromise heritage integrity or leave 'heritage buildings behind'.

A key milestone in this process was the involvement of Ciarán Cuffe, then a Member of the European Parliament and Rapporteur for the EPBD recast. As a trained architect, his dual perspective helped bridge the gap between policy and practice. Since the EPBD was ratified the subcommittee published its research and findings on January 26th[1] 2026. The author has been credited as a co-author.

2. Methodology: The Architect's Digital Notebook

The technical foundation of this study is the application of **Jupyter Notebooks**[2] and this paper also serves as a case study and a showcase of the technology for the architectural profession. While "data science" may seem removed from traditional practice of the architect, the Jupyter environment is remarkably similar to the notebooks architects have utilised throughout their careers. Just as an architect might use a 'Black Daler Pad' notebook/scrapbook to record, notes, calculations, observations and sketches, the Jupyter Notebook serves as a digital equivalent where notes, text, and calculations can also coexist but with vastly more powerful properties that the profession may not realise are available to them for free.

The selection of this technology was driven by a core contention: architectural domain expertise is essential for meaningful interrogation of building datasets. Currently, the architectural profession is often at the mercy of questions asked by data scientists who may lack the architectural or historical understanding of the built environment to ask the right questions.

By adopting open-source tools like Python (interestingly named after Monty Python)[3] and the **PANDAS** library (short for **Python And Data Analysis Library**), architects can move from being the consumers of data interrogation to being the interrogators.

In short the intelligence needed with the availability of these new tools is not artificial but architectural.

[1] https://publ.icomos.org/publicomos/jlbSai?html=Pag&page=Pml/Not&base=technica&ref=9A61DF78FF5F60C28BC202947943B6ED

[2] https://jupyter.org/

[3] https://python.plainenglish.io/the-monty-python-story-behind-python-a-tale-of-comedy-and-code-e2adb9004608

3. Data Scale and the Python Advantage

The research utilises two large-scale government open datasets:

1. **The Building Energy Rating (BER) Database:** Over 1 million records and 211 columns (effectively 211 million data points). https://data.gov.ie/dataset/ber-research-tool
2. **The National Inventory of Architectural Heritage (NIAH):** Approximately 55,000 records and 20 columns or over 1 million data points. https://data.gov.ie/dataset/national-inventory-of-architectural-heritage-niah-national-dataset

Managing these datasets in standard spreadsheet software like Excel is virtually impossible; the sheer volume often leads to system instability or reaching row limits. The PANDAS library within Python gives the users the ability to process millions of variables simultaneously. A Python library is analogous to Excel's 'function' libraries built in under the $\mathbf{f_{(x)}}$ symbol simplifying complex financial, trigonometric and statistical calculations. Python libraries allow the user to process data, create plots and run many different types of analyses putting in the hands of architects, tools that allow for a level of interrogation that has, until now, been absent in the study of the national building stock.

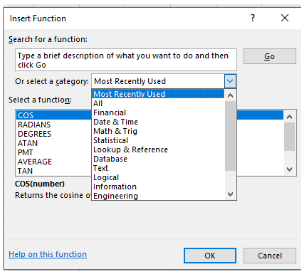


Figure 1Python Libraries are analogous to the built in function libraries in Excel

4. Correcting the "1900" Statistical Anomaly

One significant statistical discovery in this research was identifying a severe data distortion within the BER database: a massive, anomalous spike of G-rated properties defaulted to a construction year of 1900. This indicates a fallback entry by surveyors lacking the expertise to accurately date historic structures. To resolve this, the raw BER data was synthesized with the high-fidelity age distribution of the National Inventory of Architectural Heritage (NIAH). By redistributing the artificial "1900 spike" along this expert historical timeline, a critical reality emerged.

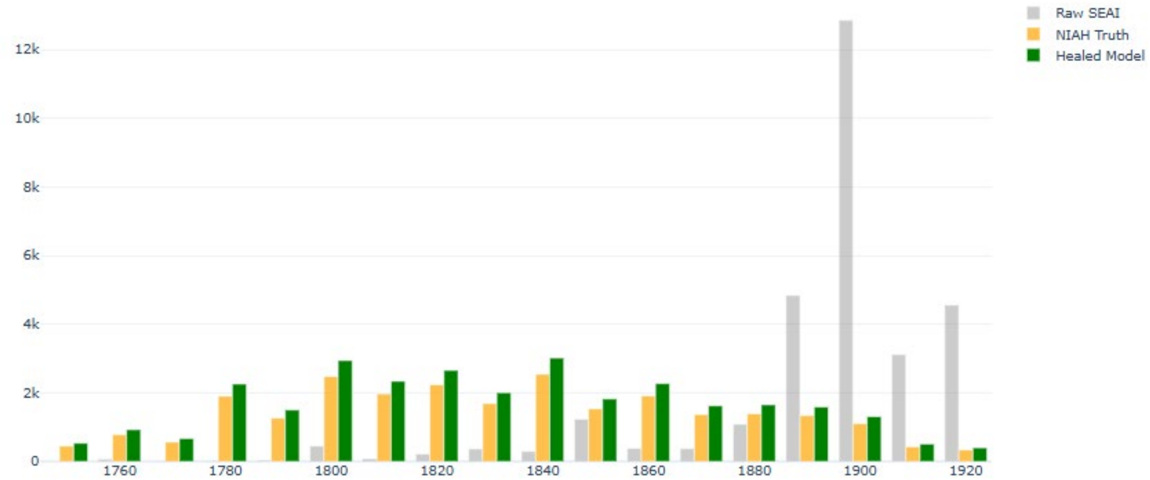


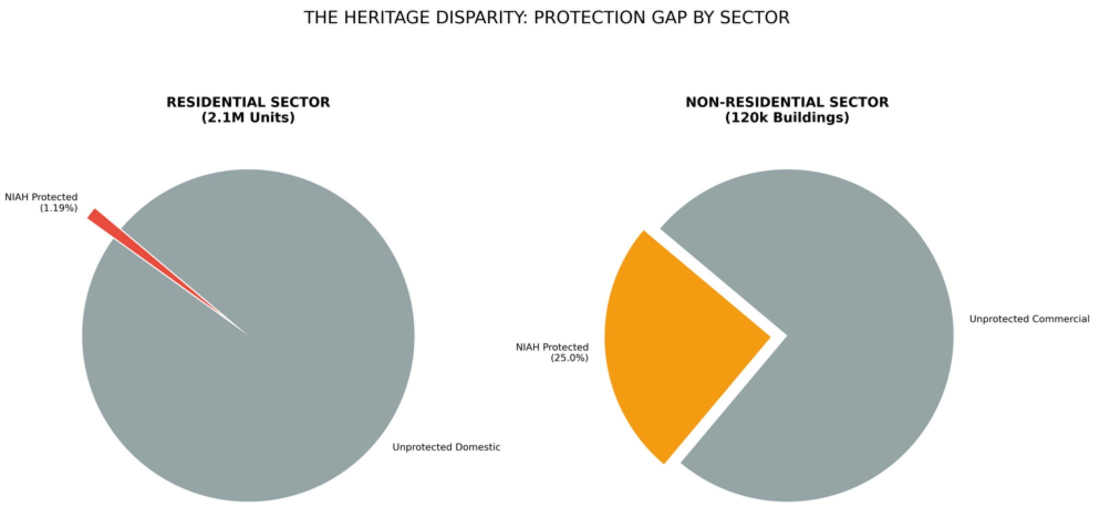
Figure 2 The grey bars show the dates the BER Assessors entered for the property showing most without the expertise to assess chose a default value of 1900. The NIAH distribution shown in yellow was applied to correct the Raw SEAI data grey bars into a healed model green bars.

The corrected model visually demonstrates that protected heritage properties account for a highly significant proportion—approximately 70% to 80%—of the surviving pre-1930 G-rated housing stock. This proves that the bulk of these poorly rated older structures are recognized heritage assets, demonstrating to policymakers that resolving the G-rated crisis requires specialised, conservation-led strategies rather than standard mass-market retrofits.

5. Key Findings: Residential and Non-Residential Disparities

While the previous section demonstrates that heritage assets heavily dominate the worst-performing 'G' tier, zooming out to the national level reveals just how concentrated this problem is. A cross-analysis of these two datasets reveals a marked difference in how 'recorded' heritage is represented:

- **Residential Sector:** Of the 1 million+ residential BER entries out of 2 million + residential structures nationally, only 25,000 align with the NIAH residential records. This means only **1.25%** of the overall residential building stock has formal protected/recorded status.
- **Non-Residential Sector:** Nationally, there are approximately 120,000 commercial/public structures. The NIAH lists 30,000 non-residential structures, meaning **25%** of the total non-residential stock is identified as heritage.



This is a significant finding: the impact of the recast EPBD will be felt most acutely in the commercial sector, where every fourth building is likely to have architectural significance.

6. A Profile of National Residential Building Stock

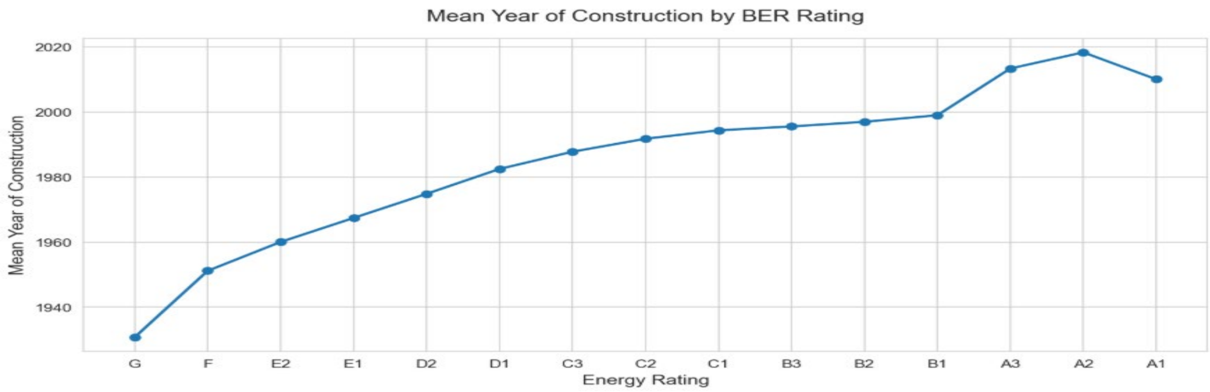


Figure 3 illustrates the direct correlation between building age and energy performance.

The tables on the following page offer a **multidimensional analysis** of key building parameters. Notable is the remarkably small glazing-to-wall ratio in G-rated properties, a defining characteristic of this cohort.

EnergyRating	Year of Const	BerRating	GFA (sq m)	UVal Wall	UVal Roof	UVal Floor	UVal Win	UVal Door	Wall Area	Roof Area	Floor Area	Win Area	Door Area	Num Storeys	CPC	EPC
A1	2010.10	15.17	152.30	0.18	0.14	0.18	1.22	1.42	128.07	91.28	88.51	30.42	3.01	1.95	0.10	0.10
A2	2018.39	42.29	134.02	0.18	0.12	0.12	1.19	1.23	101.29	65.12	63.48	24.48	2.56	1.92	0.28	0.29
A3	2013.42	58.43	129.16	0.19	0.13	0.15	1.26	1.51	108.08	71.06	70.00	23.45	2.81	1.89	0.36	0.38
B1	1999.04	89.61	155.72	0.30	0.15	0.25	1.74	2.00	118.06	87.84	83.24	27.86	3.08	1.81	0.62	0.60
B2	1997.02	115.03	147.54	0.36	0.17	0.25	2.04	2.25	106.08	80.38	73.13	26.09	2.82	1.82	0.82	0.79
B3	1995.59	139.64	137.70	0.40	0.20	0.33	2.28	2.47	104.16	79.01	72.90	23.81	2.90	1.88	0.98	0.94
C1	1994.40	163.72	125.76	0.44	0.23	0.37	2.48	2.55	98.93	73.76	69.35	21.42	2.93	1.87	1.12	1.06
C2	1991.81	188.10	118.66	0.49	0.25	0.41	2.61	2.57	96.52	72.26	68.64	20.38	2.92	1.81	1.27	1.19
C3	1987.82	213.33	112.29	0.55	0.27	0.44	2.72	2.56	93.47	71.82	68.65	19.70	2.93	1.74	1.41	1.32
D1	1982.52	242.69	104.94	0.64	0.31	0.47	2.82	2.51	88.28	68.92	66.45	18.60	2.87	1.67	1.56	1.45
D2	1974.86	280.21	99.72	0.81	0.38	0.52	2.92	2.50	85.60	67.54	65.83	17.91	2.83	1.62	1.77	1.63
E1	1967.44	320.36	96.01	1.00	0.49	0.56	3.04	2.50	83.83	64.81	64.36	16.93	2.83	1.60	2.04	1.85
E2	1960.06	360.13	94.47	1.19	0.62	0.61	3.14	2.52	85.26	64.06	63.76	16.48	2.86	1.62	2.31	2.07
F	1951.18	413.72	92.84	1.38	0.82	0.66	3.30	2.56	87.36	64.73	63.97	15.82	2.94	1.61	2.67	2.35
G	1930.75	648.03	83.02	1.71	1.41	0.74	3.68	2.62	87.79	62.23	59.37	12.74	3.08	1.52	4.12	3.31

Leveraging the PANDAS library to process over **211 million data points** allows for **exhaustive precision** within each rating band. By further isolating the 'F' and 'G' cohorts by storey height, the research moves beyond broad national averages to construct the specific architectural archetypes required for accurate energy modelling

EnergyRating	Storey_Category	Count	Year of Const	BerRating	GFA (sq m)	UVal Wall	UVal Roof	UVal Floor	UVal Win	UVal Door	Wall Area	Roof Area	Floor Area	Win Area	Door Area	Num Storeys	CPC	EPC
F	1 storey	20422	1961	414.05	74.36	1.06	0.47	0.60	3.13	2.46	70.82	66.25	68.86	12.79	2.72	1.00	2.37	2.04
	2 storeys	24347	1945	413.44	103.42	1.62	1.04	0.70	3.42	2.63	97.74	62.14	59.50	17.74	3.09	2.00	2.82	2.50
	>2 storeys	1961	1920	413.86	154.04	1.80	1.70	0.68	3.57	2.70	130.86	81.04	68.39	23.43	3.49	3.07	3.10	2.87
G	1 storey	34901	1937	672.37	65.65	1.56	1.17	0.70	3.56	2.50	71.29	61.92	60.87	10.53	2.76	1.00	3.88	3.09
	2 storeys	30018	1926	622.53	97.82	1.87	1.65	0.78	3.80	2.75	102.79	61.53	57.17	14.68	3.36	2.00	4.31	3.49
	>2 storeys	2368	1904	612.37	151.29	1.90	2.00	0.76	3.93	2.80	140.80	75.76	65.15	20.81	4.28	3.09	4.69	3.88

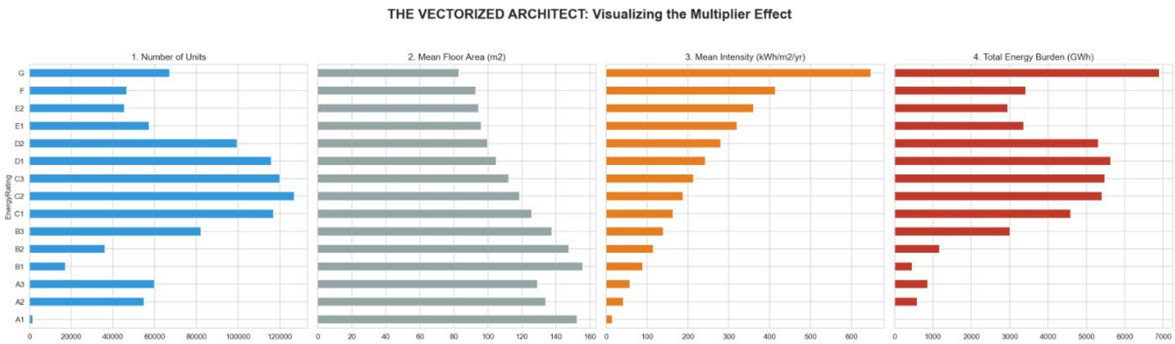
Detailed sub-stratification of the G-rated tier reveals a critical logistical insight: the overwhelming majority are detached structures. This evidence confirms that for the most vulnerable portion of the housing stock, party-wall consents and neighbour-related legal barriers to retrofitting are largely non-existent.

.

EnergyRating	DwellingTypeDescr	Number	Year	BER	GFA(m2)	U Wall	U Roof	U Floor	U Win	U Door	Wall Area	Roof Area	Floor Area	Win Area	Door Area	Storeys	CPC	EPC
G	Detached house	27110	1930	676.91	98.88	1.73	1.46	0.82	3.68	2.79	112.50	83.10	80.12	15.28	3.33	1.38	4.59	3.55
	Semi-detached house	10293	1944	621.52	85.74	1.70	1.32	0.75	3.67	2.73	86.77	59.87	58.13	14.54	3.35	1.66	4.13	3.34
	Mid-terrace house	9756	1921	612.70	79.11	1.86	1.74	0.67	3.89	2.83	61.37	48.82	46.42	10.99	3.27	1.94	4.17	3.37
	End of terrace house	5695	1934	612.80	80.00	1.79	1.45	0.79	3.72	2.80	87.07	49.09	47.38	12.13	3.43	1.90	3.92	3.27
	Top-floor apartment	5416	1927	646.53	49.12	1.53	1.57	0.44	3.45	1.77	50.07	46.96	18.83	6.49	1.77	1.24	2.65	2.56
	Ground-floor apartment	3474	1931	624.81	41.47	1.49	0.71	0.73	3.46	2.02	47.70	15.29	37.57	6.23	1.98	1.10	2.20	2.11
	House	2515	1931	742.00	93.85	1.72	1.38	0.80	3.78	2.75	105.50	75.26	72.41	15.52	3.67	1.44	5.64	4.13
	Mid-floor apartment	1974	1919	602.79	39.44	1.51	0.44	0.49	3.51	1.63	45.24	7.69	19.57	6.01	1.70	1.06	2.21	2.14
	Maisonette	743	1926	615.91	75.06	1.50	1.47	0.85	3.51	2.54	83.47	56.13	48.98	10.05	2.33	1.89	3.37	3.12
	Apartment	222	1946	671.12	55.46	1.35	0.97	0.65	3.51	2.29	58.24	36.33	37.09	8.24	1.84	1.38	7.09	6.34
	Basement Dwelling	89	1899	596.44	36.05	1.67	0.78	0.72	3.75	1.92	35.14	10.33	41.60	3.66	1.56	1.10	2.08	2.07

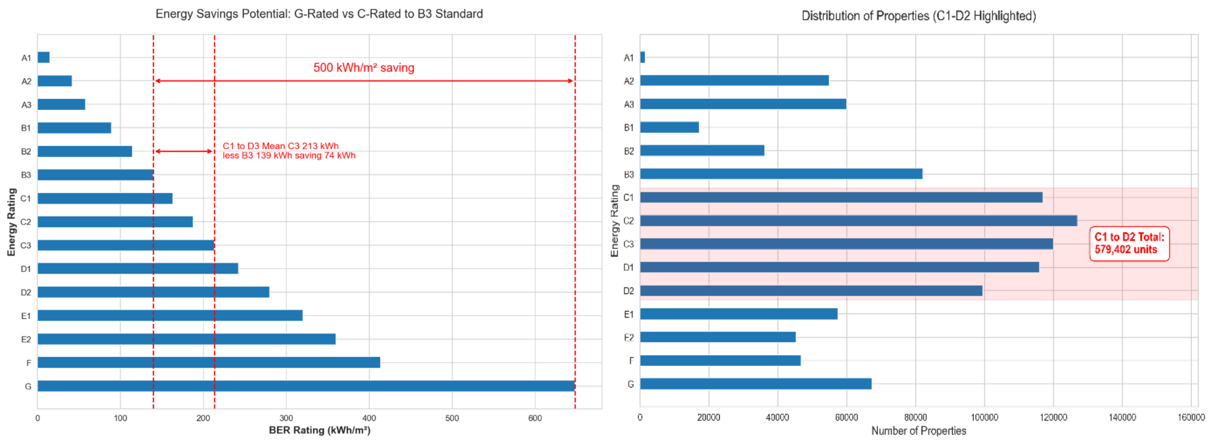
7. The "Worst-First" Mandate and the "Energy Black Hole"

The recast EPBD targets the "worst-performing 16%" of the non-residential stock first but only states that in the residential sector there would be an overall reduction of 16% by 2030.




```
PROJECTED NATIONAL G-RATED BURDEN: 6,904.2 GWh per year
```

Through the use of **vectorised multiplication** (calculating averages across the entire dataset simultaneously), the research establishes a national energy baseline. The data reveals that while there are only 60,000 G rated units in the BER dataset (and thus 120,000 nationwide) as G-rated property consumes an average of **650 kWh/m² per year,** they actually represent the single largest energy consumers in absolute terms.



By visualizing this as "energy bands" in horizontal bars from G to A, the research proves that upgrading a single G-rated historic building yields a greater reduction in national carbon emissions than upgrading dozens of modern C-D rated buildings. **In short upgrading 120,000 G rated structures to B3 would match upgrading 500,000+ structures in the C1 to D2 cohort**. Thus while the EPBD for residential sector targets an overall 16% improvement by 2030, the "worst-first" approach of focusing on the 120,000 G rated properties first remains the most mathematically efficient path to reaching that goal in the residential sector also but as can be seen this will largely intersect with heritage properties.

8. Policy Proposal: Pro-Rata Grant Funding

This conclusion in the research highlights a flaw in current fiscal policy. The SEAI currently provides a flat-rate grant of approximately €25,000 for energy upgrades, regardless of whether the building moves from D to B or G to B.

The author argues for a **pro-rata grant model**. Historic buildings are more expensive to retrofit due to the need for breathable materials and specialised craftsmanship but also provide a much larger reduction in the national energy load. Grant provision should therefore be scaled to the actual energy savings (kWh/m²/yr reduction) achieved, providing the necessary financial bridge for specialized architectural interventions.

9. Conclusion This paper demonstrates that when architects drive the narrative of data interrogation, the results remain grounded in professional reality. By demystifying tools like Jupyter Notebooks and PANDAS we can ensure that the drive toward the 2030 energy targets does not come at the cost of our cultural identity. The "black notebook" of the future architect must include these data-driven insights to protect and preserve our historic built environment in a decarbonizing world.