



Research Repository UCD

Title	A linked data approach to multi-scale energy modelling
Authors(s)	Hoare, Cathal, Aghamolaei, Reihaneh, Lynch, Muireann Á., Gaur, Ankita, O'Donnell, James
Publication date	2022-10
Publication information	Hoare, Cathal, Reihaneh Aghamolaei, Muireann Á. Lynch, Ankita Gaur, and James O'Donnell. "A Linked Data Approach to Multi-Scale Energy Modelling." Elsevier, October 2022. https://doi.org/10.1016/j.aei.2022.101719 .
Publisher	Elsevier
Item record/more information	http://hdl.handle.net/10197/26182
Publisher's statement	This is the author's version of a work that was accepted for publication in Advanced Engineering Informatics. Changes resulting from the publishing process, such as peer review, editing, corrections, structural formatting, and other quality control mechanisms may not be reflected in this document. Changes may have been made to this work since it was submitted for publication. A definitive version was subsequently published in Advanced Engineering Informatics (VOL 54, (October 2022)) DOI: https://doi.org/10.1016/j.aei.2022.101719
Publisher's version (DOI)	10.1016/j.aei.2022.101719

Downloaded 2026-05-01 23:35:22

The UCD community has made this article openly available. Please share how this access benefits you. Your story matters! (@ucd_oa)



© Some rights reserved. For more information

A Linked Data Approach to Multi-scale Energy Modelling

Cathal Hoare^a, Reihaneh Aghamolaei^{b,c}, Muireann Lynch^d, Ankita Gaur^d,
James O'Donnell^a,

*^aSchool of Mechanical and Materials Engineering and UCD Energy Institute,
University College Dublin, Belfield, Dublin 4, Ireland*

*^bSchool of Mechanical Manufacturing Engineering, Dublin City University Glasnevin
Campus, Dublin 9, Ireland*

^cCollege of Fine arts, University of Tehran, Tehran, Iran

^dEconomic and Social Research Institute, Dublin, Ireland

Abstract

Interactions between built infrastructure are complex and nuanced; changes to any one component can have disproportionate effects on the system as a whole. For instance, adoption of heat pumps or electric vehicles by a significant proportion of a population in an urban centre would place new demands on both electricity transmission and distribution networks. It is essential that planners - both national and local - can understand and share information about the resource demands that this type of change places on national and local infrastructure. Access to integrated sources of information - from building component to national levels - is key to supporting policy makers and decision takers. However, over time, information - and as a consequence, the software that manages it - has evolved into functional silos; this has, in turn, affected the definition of data exchange standards. This limits the ability of experts in functional areas to exchange data and implement broader decision support systems. This paper describes the use of linked data approaches to permit queries across large, diverse information sources to provide reasoning about complex questions at multiple scales. The methodology defines a central context to which various external sources can be attached. These

distributed sources are, in themselves, registered in a central catalogue; they remain, however, under the control of their source organisations. In this way a large, extensible, interconnected network of distributed data describing, for example, a built environment or electricity transmission network; this network of data resources can be queried centrally to provide customised views of subsets of the data, and so provide a richer view than one source in isolation. The approach was applied to prepare and integrate information about Ireland’s transmission grid and administrative boundaries, along with domestic housing stock into a single data source. The resulting data network is queried by a scenario exploration tool. This tool successfully allows analysis, at a national level by economists, of the effects of the adoption of new technologies on the national grid of Ireland.

Keywords: Energy assessment, data inter-operability, Linked Data

1. Introduction

Driven by the need to meet carbon emission goals, a variety of changes are being made to built infrastructure [1]. These include, for example, changes to transport through the adoption of electric vehicles, changes to buildings through the adoption of heat pumps and changes to energy intensive infrastructure such as water treatment [2, 3, 4]. The widespread adoption of even one of these technologies will have an effect on the capacity of a nation’s electricity transmission and distribution networks, which must develop in parallel to satisfy anticipated demand [5, 6]. This work describes the use of linked data approach to create a multi-model data collection that can be used to reason about the impact of such changes at various degrees of granularity, from building to national levels.

Access to accurate information is of fundamental importance to all stakeholders in this process, including planners and policy makers [7]. The information that must be managed is, however, complex and homogeneous, both in content and organization [8]. Information sources are dynamic, evolving, and multifaceted; information can be distributed across multiple sources providing related, yet homogeneous collections of data that inform utility companies, planners and other stakeholders [9]. National and district level views of the same entities are managed by separate formats and systems, requiring expensive data reorganization in order to extract relevant information for planners and policy makers [10, 11, 12]. Together these issues describe

rich, timely data sources whose utility is diminished by their disjointed and homogeneous nature. This work presents an approach that seeks to overcome these difficulties to increase value by integrating these diverse sources into a heterogeneous data source that can be extended to include new sources as they become available [13].

Several advances have enabled solutions to these problems [14]. Cloud computing has made performant, cheap processing available [15]. This has reduced the cost of maintaining processing infrastructure, eliminating the need for the purchase of a physical machine, hiring of administrators and eliminating the cost of maintenance. Cloud services, while sophisticated, can be managed as turn-key services and their providers are acutely aware of the need to maintain security and privacy standards [16, 17]. Linked data provides a mechanism to interlink disparate datasources and interrogate them using semantic queries - that is queries that are of associative and contextual nature [18]. Linked data allows all relationships in a datastore to be explicitly defined. This facilitates the development of flexible queries, and admits reasoning about both data and their relationships [19]. Linked data also permits the integration and exchange of data from a variety of sources and domains [20].

The system described here-in, the Dynamic District Information Model (DDIM) server aims to use these technologies in a manner that creates a unified information view at various scales from national to building. An example application of the server is described to motivate the server's utility; the example provides data to an exploration tool that informs analysts that seek to explore the effects of large scale adoption of electricity intensive technologies at building level on local national level electricity infrastructure. This view can be explored at varying granularity, from transmission nodes to understand the impact of novel technologies, such as heat pumps and electric vehicles, on distribution and transmission networks. This hierarchy (buildings supplied by distribution network supplied by transmission network) is used to provide context for other information sources allowing their integration into a wider data network. While a core set of information sources are provided, the schema is extensible; these extensions are themselves defined through data rather than software. The data set that results is queryable through semantic queries, defined in SPARQL. Together these capabilities allow for integration of the various data views that have developed in the construction, electricity utilities and other domains. The arrangement also overcomes issues that arise as new sources of information become available without requiring ex-

61 pensive software modifications [21]. While these sources can be integrated
62 using other technologies, such as developing a unified database schema, it is
63 the authors' experience that both the initial implementation, and effort to
64 include new sources, especially in a project with many stakeholders, requires
65 an expensive effort that may not be viable for poorly resourced projects.

66 This paper continues by reviewing concepts, technologies, data formats
67 and systems that informed this work. The theoretical underpinning of DDIM's
68 data organization concept and methodology follow, along with details of its
69 implementation. A description of the server's use to provide data to a report-
70 ing interface that allowed economists to estimate various scenarios demand
71 on the national energy grid illustrates an application case for the server.

72 2. Background

73 Data integration and exchange has, in general, proven to be a vexing prob-
74 lem. The Architecture/Engineering/Construction/Owner/Operator (AECOO)
75 industry has proven to be no different, where diverse uses, narrow margins,
76 a lack of cross-enterprise standards, and the formations of task specific data
77 silos has created both horizontal (e.g. between functional areas in projects)
78 and vertical (e.g. between individual building representations and neighbour-
79 hoods) barriers to data exchange [22]. A range of solutions developed inline
80 with software engineering advancements have been developed, including the
81 replacement of monolithic applications, the use of software APIs and data
82 management techniques including semantic and RDF graph approaches [23].
83 Even so, many solutions in the literature suggest applications that repackage
84 data and inject them into a purpose specific data management schema [24].
85 While it is possible that many of these solutions could be integrated into a
86 single web of data, little work has been done on providing an infrastructure
87 that would support this development.

88 This paper focuses on the usefulness of RDF graph based approaches, and
89 in particular, the use of distributed, federated data sources to create such an
90 information resource. The usefulness of the approach is demonstrated by
91 providing data to decision support systems in the energy domain. To this
92 end, the work described here uses linked data approaches to organise large
93 scale distributed data sources. This section describes the background to the
94 concepts and technologies used to achieve this. While many of the data
95 sources used here are provided by project participants, one set of building

energy modelling data was created by the authors for another purpose and reused here [25]; a review of the background to this work is also provided.

2.1. Decision Support and Modelling Systems

The architecture of decision support systems broadly follow that shown in Figure 1 [26]. Four general tasks, including data collection, data processing, learning and interpretation combine human intelligence and machine learning techniques to inform decisions with many dynamic variables. These types of decisions manifest in society, and as a result, a large body of work has sought to investigate and apply a variety of modelling, machine learning techniques and interpretation supports to various sub-domains including construction [27], economics [28] and energy [29]. The quality of a learning model's output is dependent on the quality and organisation of the inputs [30]. These are determined during the first two phases, data collection and data processing.

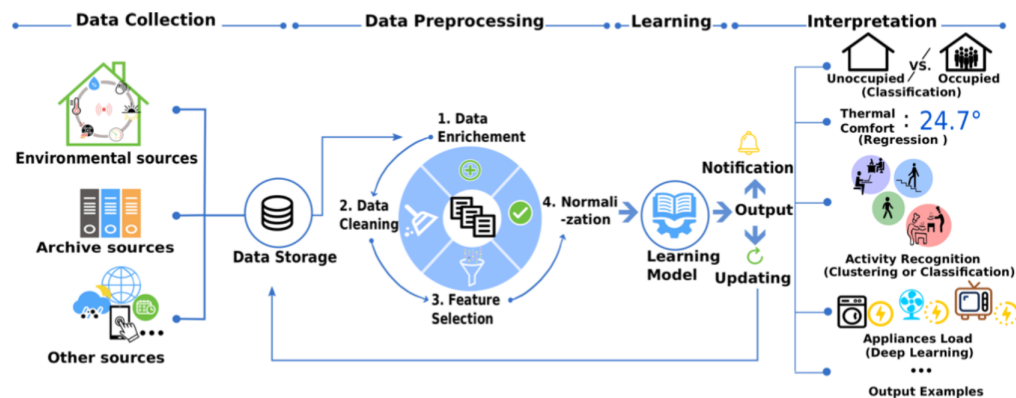


Figure 1: The general organisation of simulation and other machine learning systems [26]

Sourcing data, expressing it in a suitable format and integrating several sources are difficult tasks that have generated a large body of research. Indeed, it is at this point that systems developed in academia often encounter most difficulty in moving to a wider application in industry [31]. The reasons are many. Besides the aforementioned challenges in organising data, scalability of learning models, commercial and legal sensitivities, and price of adoption in an industry with narrow margins are amongst the greatest barriers to data collection [8].

117 2.2. Semantic and Linked Data

118 Linked data [32] is a set of design principles for sharing machine-readable
119 data on the Web [33]. It is made up of named data that can be exchanged and
120 reasoned about. Data can be expressed statically as Resource Description
121 Framework (RDF) or dynamically embedded in the format. These data are
122 named using a unique Uniform Resource Identifier (URI) that allows it to
123 be found and shared across the web. The data can contain links to other
124 named data sets and meta data about the relationship between these linked
125 datasets. Given a rich enough network of data, a computer algorithm can
126 reason about entities based on the contexts provided by their related data.

127 While data are encoded as RDF, meaning is imbued by providing on-
128 tologies define the concepts and relationships of a domain, and so provide
129 meaning to the encoded data [34]. These meanings provide a common vocab-
130 ulary between different data providers, ensuring that data can be exchanged
131 and integrated in a structured way. These ontologies are represented using
132 Web Ontology Language (OWL). This approach has been used extensively in
133 both industry and research to facilitate data integration and exchange. The
134 use of ontologies to facilitate data exchange and integration was explored by
135 [35], while the ability to reuse, and so broaden the adoption of these ontolo-
136 gies was explored by [36]. Linked data has been used to support complex
137 reasoning. For example, Baumgärtel et al [37] used semantic approaches to
138 reason about the application of building performance regulations to design
139 and operation; in doing so, they improved the efficiency and effectiveness of
140 a complex form of analysis.

141 The DDIM uses these linked data formats to admit queries across in-
142 tegrated data sources. This, in effect, allows these diverse sources to be
143 integrated and their information to be treated as a single source.

144 Several protocols, tools and languages have been developed as part of the
145 overall effort to deliver linked data. SPARQL, a semantic query language was
146 developed to create, retrieve, update and delete data stored in Resource De-
147 scription Framework (RDF) based datastores [33]. Data Catalogue (DCAT)
148 is an RDF vocabulary that allows reasoning about online data catalogues
149 [38]. Its use promotes discoverability of data sources and provides a mecha-
150 nism to promulgate the data sources schema. The vocabulary also provides
151 metadata that can be used to reason about a dataset's provenance. The
152 vocabulary provides a rich set of attributes that can provide context about
153 a dataset and its versions.

154 The DDIM server uses these ontologies to integrate distributed data
155 sources, and so allow data controllers associated with each source to maintain
156 control of their data. It also ensures that the server is extensible and that
157 the server can scale.

158 2.3. *The Role of Context*

159 Information can be seen as some fact about some thing at some time;
160 the thing and the time together provide a context that allows all sources of
161 information about that *thing* to be interconnected [39]. Context also pro-
162 vides a framework where the *things* can be related to one another, and so
163 expands the network of interconnected information. Linked data, with its use
164 of a URI to identify entities and its ability to define relationships between
165 entities, is ideally suited to represent this type of information. *Things* can
166 be represented in a multitude of ways, but in built environments, context is
167 related to physical concepts such as buildings and areas.

168 The concept of using a common context has been used by several works to
169 provide a mechanism to integrate and share AEC information. ISO 21597's
170 approach [40] provides a framework that expresses a common context by cre-
171 ating a document container that represents some context; documents and
172 other data related to that context can be grouped together in this container.
173 The embedded documents are not modified and the notion of relatedness
174 between them is communicated through their common association with the
175 container. This approach is, in itself, a combination of the Multi-model [41]
176 and Linked Building Data (LBD) approaches [42]. The former created a
177 common context for distributed information collections that form a feder-
178 ated database. A series of ID-based links are formed between collections,
179 creating an interlinked collection of data that can then be subject to query.
180 These federated queries - queries that can be distributed amongst several
181 linked data sources - are submitted to a single authority, which responds
182 by querying the distributed members of the database and returning an ag-
183 gregation of those results. The approach is beneficial because data can be
184 maintained in its original format, though some compatible API must be pro-
185 vided to make each source queryable [43, 44]. LBD also admits the concept
186 of a common context; in this approach, relationships between represented
187 entities are captured through relationships that are defined as part of some
188 ontology defined using the Web Ontology Language (OWL). Objects within
189 the resulting model are uniquely identified using a URI and related objects

190 can be retrieved by querying across relationships using the SPARQL lan-
191 guage which supports federated queries; as a result, data can be distributed
192 across multiple servers.

193 The DDIM uses the relationships that exist between geographical areas
194 and locations to provide a common context between data sources. This con-
195 text is captured through an RDF graph where relationships conform to a
196 context ontology. This would allow queries at one level of geographical gran-
197 ularity (for instance a CityGML source describing a district) to aggregate
198 data stored at a different granularity (for instance an IFC format describing
199 buildings); this approach allows data to remain in its context, reducing data
200 duplication and its associated issues. Furthermore, data sources remain gen-
201 eral and are not specific to any one purpose; they can be reused for other
202 interrogation tasks.

203 While the ability to reason about distributed data is in itself sufficient
204 motivation to embrace the concept of linked data, the schema has further
205 advantages. Because the data are distributed, maintenance is managed by
206 each organisation and its impact on any one participant is reduced; only
207 expertise about data related to an organisation’s speciality need be main-
208 tained, while this simple structure allows new data to be easily integrated
209 [45]. Furthermore, the data exposed need only be the subset that is relevant
210 and permissible for an organisation’s remit. While there is an overhead in
211 re-expressing data in RDF format, the process is becoming easier as software
212 tools improve [46].

213 A methodology and software architecture for creating a common context
214 for a project is now described including a linked data implementation. The
215 common context is used to include other information sources. These sources
216 are distributed, and queried through federated queries.

217 **3. Methodology**

218 Energy Systems Integration is a relatively new research domain that aims
219 to address some of the leading energy challenges required to decarbonise
220 energy systems. In pursuit of this initiative’s goals, the ESIPP Project brings
221 together a multi-disciplinary team of energy engineers, data scientists and
222 economists to analyse diverse information sources to reason about the impact
223 of various decarbonisation measures on Ireland’s electricity network. In order
224 to organise these information sources, an off-the-shelf solution, called the
225 Dynamic District Information Model (DDIM) [47], is used. This section

226 describes both the implementation of the DDIM and its application to this
227 use-case.

228 The server implementation consists of four key functional areas:

- 229 • A series of interfaces to allow client software to query (through RESTful
230 or SPARQL based interfaces) and to interact with the notifications
231 system;
- 232 • A DCAT compliant data catalogue;
- 233 • A Core Spine or RDF schema that can be used to place other informa-
234 tion in the space and time covered by the project. This spine allows
235 information to be associated with entities in either the urban or build-
236 ing information spaces, and creates a bridge to allow seamless querying
237 across these spaces;
- 238 • Subsystems to manage the native formats of both the urban and build-
239 ing spaces;

240 A Django server implements the RESTful interface. This framework also
241 provides user management and security. All queries must be made to this
242 central server. GraphDB is installed within the same domain. SPARQL
243 queries are submitted to this and its functionality is responsible for the ex-
244 ecution of any federated queries to the distributed sources. Those sources
245 implement an access policy that whitelists the central server to allow access
246 to their data sources.

247 The DDIM uses two RDF graph based techniques to integrate large, dis-
248 tributed, diversely formatted sets of data, and so inform tools that admit
249 deeper reasoning about the combined information sources. Specifically, the
250 server supports a scenario exploration tool that allows users to pose hypo-
251 thetical modifications, such as the addition of heat pumps, to various propor-
252 tions of the domestic housing stock. While supporting this type of analysis,
253 the model permits the data sources' owners to maintain complete ownership
254 and access control over their information on their own servers; this allows
255 them to contribute information and mitigates against any reluctance due to
256 commercial or regulatory concerns.

257 The two technologies are i) a *core context* that interconnects the dis-
258 tributed data sources across multiple scales and ii) a *data registry* and *feder-*
259 *ated queries* that together allow contributors to include data served by their

Table 1: DDIM Methodology describing steps to set up the server and populate it with data

Phase	Task Performed	contributor
Initiation Phase	Initial survey of information relevant to project, with collection of schema	Expert participants
	Decide on core context structure (meeting, document and disseminate)	Expert participants and admin
Preparation	Definition of core context	Admin
	Deployment of servers	Admin
	Data preparation	Expert participants
	Publication and registration of sources	Expert participants and admin
Operational	Use by authorised users and systems	Expert participants and systems

own servers for querying by other project participants. While the server is implemented as a cloud-based and turn-key system, some preparation by a project team is required in order to define required data artifacts. The DDIM methodology, shown in Table 1 describes these. The methodology is divided over three phases, including *initiation*, *preparation* and *operation*. The technique used to apply the technologies, including the application of the methodology, along with the measures taken to prepare the information sources is described in the remainder of this section.

3.1. Core Context

A graphical representation of the context model along with the organization of this project’s data through the use of the ontology are shown in Figure 2 (A). The ontology is deliberately sparse to maintain clarity. An entity represents some physical entity or space in time. Each instance of an entity is described by a class, for example, a building, region or nation. Together, these classes describe a set of layered contexts, each representing a different scale. Information sources are associated with some entity in each layer. This association with context allows sources to be intra-connected at

277 a particular scale, and inter-connected between scales. Figure 2 (B) shows
 278 the relationships used in the project described here. It should be noted that
 279 the red boxes denote the core context for the project, while the green, blue,
 280 purple and yellow boxes denote data that can be connected through the hier-
 281 archical and equivalence relationships enabled by the core context. It should
 282 also be noted that this image would be unique for each separate project.

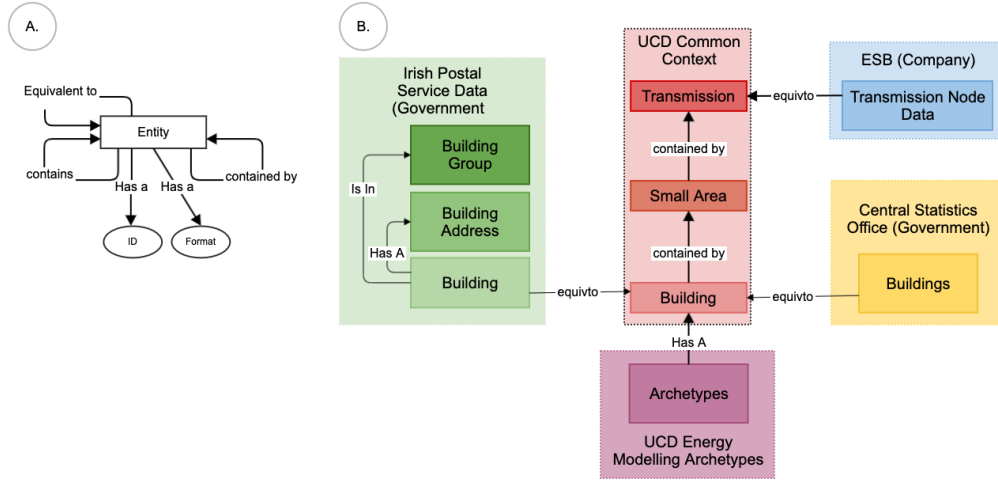


Figure 2: A. Context model and B. relationships to data sets used in this project

283 Each entity is associated with a format and an identity. Entities in the
 284 hierarchy are related to one another using one of three relationships, two
 285 hierarchical relationships, *containing*, *contained-by* and one mapping rela-
 286 tionship, *equivalent-to*. The mapping relationship is used to bridge formats
 287 where two entities are considered to represent the same physical entity. This
 288 relationship can be used as a *join query* between formats. It is the responsi-
 289 bility of the projects' data managers to maintain alignment between versions
 290 of entities and manage dependencies between them where they exist.

291 3.1.1. Setting Up Core Context

292 The first step of the methodology is applied to define the core context
 293 using custom interface. The task is carried out in three parts. Initially, a
 294 survey by project participants identifies information sources that might be
 295 relevant, and their schema are analysed to identify physical contexts such as

296 locations and timestamps. These are then analysed and organised in order
297 from most encompassing to more specific entities to form a continuum. The
298 participants then identify those entities which can act as context for other
299 information sources. Any entities that are not necessary to maintain the
300 continuum and not serving as a context for a data source are candidates for
301 removal to simplify the context model. Once the remaining entities have
302 been listed, they are passed to the DDIM administrator.

303 The definition of the context is managed by a web-based interface that
304 is provided as part of the servers implementation. The context definition
305 interface is implemented as a software wizard that guides the administrator
306 through definition and population of the context model; this interface is
307 shown in Figure 3. Step 1 asks the administrator to select a set of contexts
308 that define each level of the hierarchy. In the work described here, that set
309 includes transmission node $\rightarrow$ small area (similar in population size to a
310 neighbourhood) $\rightarrow$ building. All other data are associated with one of these
311 levels.

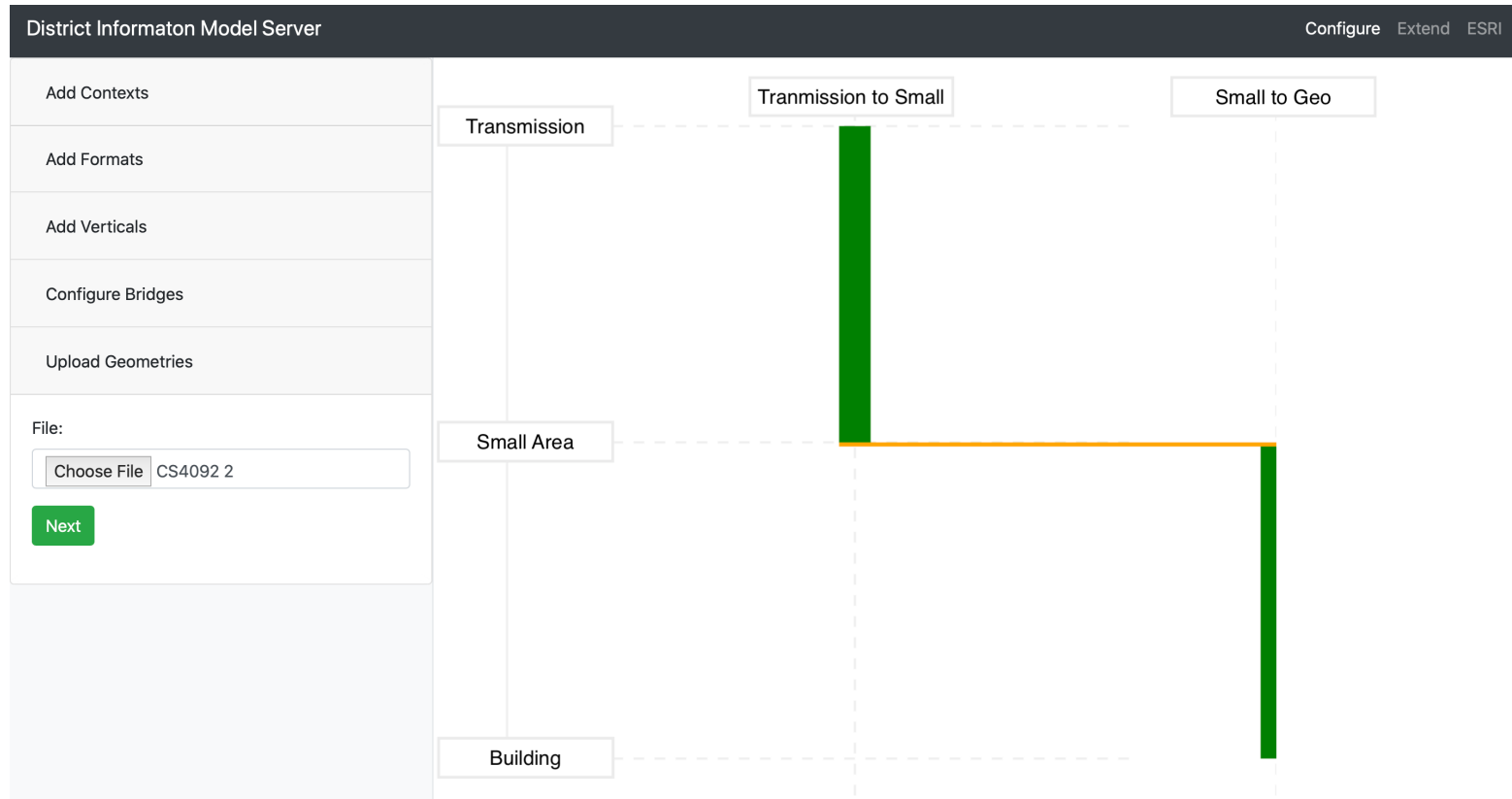


Figure 3: Screen shot of the context definition interface

312 In step 2, the administrator is asked to list the format sources, in this
313 case, three CSV files one listing transmission nodes and their associated
314 data: one describing small areas and their associated data (including the
315 transmission node that each small area is served by) and one containing
316 details of each building and their associated data including containing small
317 area and archetype. The data files associated with each format are uploaded
318 at this point.

319 The third step defines the entity types used to represent each context
320 level. A CSV parser extracts each column header from each format. For each
321 context level, a header is used to identify an entity type. In this example, the
322 transmission node identifier is associated with the transmission node context
323 level, a small area identifier is used to represent the small area level, and
324 a building identifier is used to represent the building level. These levels are
325 related using a **contains** (and inverse **contained by**) relationship between levels
326 and types.

327 In the fourth step, a rule or mapping is used to bridge between formats.
328 For example, an equivalency mapping (**equivalent-to** relationship) is created
329 between the transmission node identifiers in the transmission node and small
330 area data sets. A similar relationship is created between small areas and
331 buildings. When the context is wholly defined, the data captured by this
332 process is used to configure a Python script. This mines the data sources,
333 and based on the associations defined in the rules and produces an RDF
334 (turtle) formatted description of the project's context, a sample of which
335 is shown in Listing 1. The final context description generated 12.6 million
336 triples. This was automatically inserted into the DDIM server's GraphDB
337 triple store and added as the first entry to the data registry.

```
338 esippenergy:t1 idf:type esippenergy:entity;  
339     esippenergy:id 1;  
340     esippenergy:format "TRANSMISSION";  
341     esippenergy:type "NODE";  
342     esippenergy:context "TRANSMISSION";  
343     esippenergy:contains esippenergy:5076;
```

Listing 1: Sample of Turtle produced to Represent a Transmission Node

344 3.2. Energy Modelling of Domestic Stock

345 This project required the development of archetype models that repre-
346 sent the majority of building types found within the Irish residential building

stock. The classification of buildings plays a crucial role in energy modelling because energy consumption significantly varies according to building topology. Urban energy modellers using bottom-up modelling approaches often represent a building stock using archetypes in order to analyse a group of buildings with similar properties. Therefore, archetype-based modelling approaches are extensively used in the context of large-scale energy modelling studies that investigate the aggregated impact of energy efficiency strategies and new technologies [48].

Residential archetypes can be defined by geometry, layout and other physical features of buildings. Previous studies introduced archetypes for Irish buildings [49, 25]; based on these studies and considering the requirements of this research, this work developed six archetype dwellings that are representative of over 80 percent of the Irish building stock and are as follows: i) A two-storey detached dwelling, ii) A two-storey semi-detached dwelling, iii) A single-storey detached dwelling as bungalow archetype, iv) a terraced dwelling, v) a mid-floor apartment and 6) a top-floor apartment.

As the application of the DDIM server was, in this instance, to investigate the impact of large-scale residential heat-pump installations on the electricity consumption of the transmission grid, three separate scenarios are defined for each of the six archetypes, resulting in 18 models: i) the pre-retrofit, ii) post-retrofit and iii) new-built conditions. In the energy prediction of these 18 models, the dimensional characteristics for each archetype remained the same, but the physical characteristics of the constructions changed.

Since this study applies white-box modelling approach, the first step is to define the base case input parameter values, followed by generation of the energy models to represent Irish archetype dwellings. Energy models act as reference cases for load profiles of selected archetypes buildings. The energy modelling is performed in DesignBuilder 6.1.6 as a user interface for EnergyPlus modeling tool [50].

In the archetype development process, the values used for all input parameters are representative of the most probable values for Irish dwellings. The heating system for the pre-retrofit archetype models follows a common hot water radiator template powered by a standard boiler. The overall seasonal COP/efficiency of the heating system is 0.65. In post-retrofit and new-built scenarios, heat pumps with COP of 2.2 are used as the main heating system. Typical constructions for the external walls, roof and ground floor are extracted from Technical Guidance Document Part L [51] and used for developing the pre-retrofit, post-retrofit and new-built models. The U-

385 value for glazing systems are defined at three levels to form a high-efficient
 386 to low-efficient range corresponding to the three scenarios [51].

387 Once the modelling process is complete for all archetypes, the next step is
 388 to perform the energy simulations and record the annual energy consumption
 389 values for each of the 18 models. As it is shown in Fig. 4 , the total energy
 390 consumption which stands for heating, domestic hot water (DHW), lighting,
 391 and domestic equipment and appliances is drastically reduced in post-retrofit
 392 and new-built scenarios. Also, results show that the energy consumption of
 393 these archetypes are heavily dominated by space heating requirements (Fig.
 394 5). This point highlights the fact that retrofitting the heating systems of these
 395 buildings leads to a considerable contribution to their energy performance.

396 In this regards, post-retrofit and new-built scenarios use heat pumps
 397 instead of conventional gas boilers in addition to the other constructional
 398 retrofit measures (Fig. 5). To have a better understanding of the impacts
 399 of heat pumps on the energy performance of selected archetypes, Fig. 6 il-
 400 lustrates the monthly electricity consumption for post-retrofit and new-built
 401 scenarios.

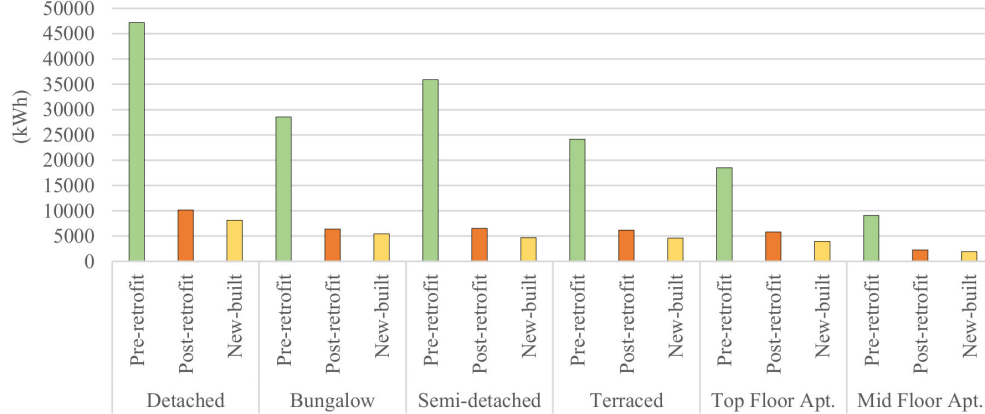


Figure 4: The total final energy consumption of selected archetypes for three scenarios of pre-retort, post-retrofit and new-built per annum

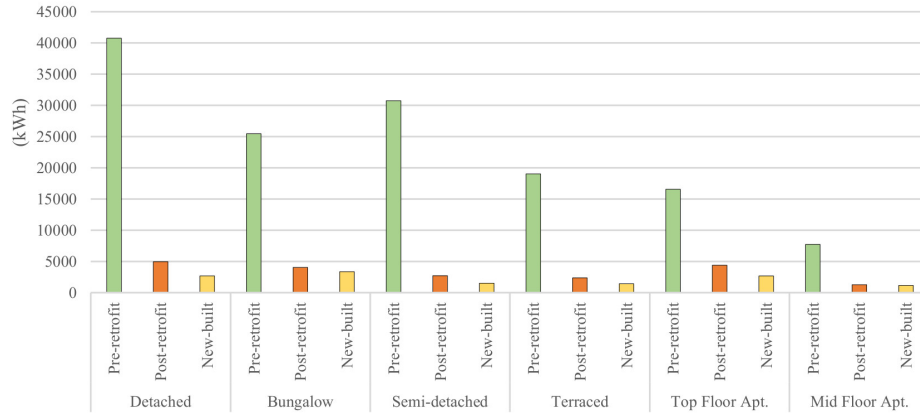


Figure 5: The annual final energy consumed for heating of the selected archetypes for three scenarios of pre-retrofit, post-retrofit and new-built

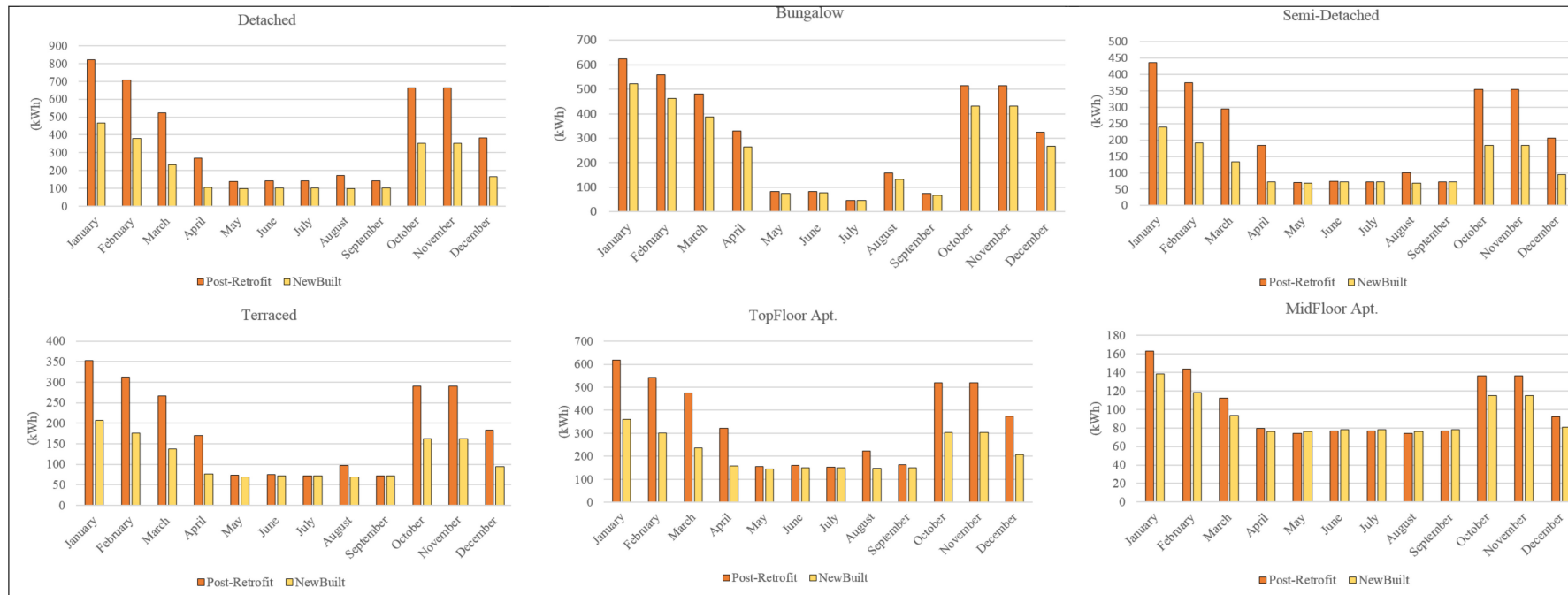


Figure 6: The monthly electricity consumption for heating of the selected archetypes in post-retrofit and new-built scenarios

3.3. Data Preparation

In this application of the DDIM server, data from several sources and formats were imported; as the project has evolved, other sources have been added. These include mapping data for transmission nodes in the Irish electricity grid, small areas and building information were received from the Irish Postal service; a description of each of these is provided in Table 2. These were combined with archetype data that has been described already.

Transmission node data was initially received as a CSV file, with details of the node’s name, an identifying number and geographical information. Geodirectory¹, a very complete information set developed by the Irish postal service describes major organisational information such as administrative boundaries such as counties and small areas, as well as information about every residential building in Ireland. Small area identities are mapped to other sources, such as Ordnance Survey Ireland’s small area GUIDs. Buildings include a unique identity for each building, geographical information such as address and longitude and latitude and also a building archetype. Finally, building archetypes (developed in Section 3.2) are imported from a CSV format; the archetype dataset includes hourly energy and heating energy consumption data for each archetype for a year.

Table 2: Summary description of context entities

Type	Purpose
Transmission	A transmission network is a high-voltage system for the bulk transfer of electric power from generation stations to points near population centres. Transmission nodes are interconnection points to the distribution network.
Small Area	A distribution network is a lower voltage network that transfers electricity from transmission nodes to buildings and other end points. A node on this network is called a transmission node. In this work these are represented using small areas.
Building	The domestic premises that consumes the electricity.

In order to create a unified data store that could be queried across various levels of detail, these information sources were imported to a number

¹<https://www.geodirectory.ie>

of information models. These are all associated with some level of detail from the context model described earlier. The models are shown (related to the core context model) in Figure 2. Note that because the context model contains a variety of physical entities, the Transmission and Building entities are conditionally associated with a `ddim:context` based on the context’s type; it should be noted that these type values are constant values agreed at project inception, and in this particular instance were defined as three strings, ‘Transmission’, ‘Small Area’ and ‘Building’. Building archetype observations are associated with Building entities through the type attribute of both. In reality, the observation data are treated as class level data and so are represented once to prevent duplication and is associated at query time.

The preparation of the core context was described earlier. The data was imported to this process from CSV files. For the other models, Transmission, Small Area and Building, the data was also imported from CSV files; this was possible as its structure was flat. Four data sources were created and integrated to create a DDIM for this project. These models were run on separate GraphDB instances. In the near future it is anticipated that some of these may run within the data providers’ own networks, allowing them to retain control of their data.

3.4. *The Data Registry and Federated Queries*

The second component of the DDIM is a data register where entries are described using W3Cs DCAT ontology. While multi-model approaches have been used in the past, these models’ triple stores have generally co-existed on a single server. While multiple models can be hosted by the server itself, the DDIM is more often configured to use distributed data stores and serves data through the use of federated queries. This configuration has two key advantages; first, maintenance of the service is simplified and so maintenance costs are reduced and distributed amongst data contributors, and second, it preserves data control. Maintenance is simplified because each data source is independent, thus ensuring that sources are discrete and self contained; most modifications to traditional APIs are driven by consumers of data, and changes can have effects that impact other users of the data. Where modifications to a user’s data requirements in the DDIM do not propagate to others, requiring only that queries are modified. Furthermore, since each project partner is responsible for serving their own data, the effort required to implement and maintain the servers data source is distributed across collaborators. Project partners can also remain in control of their data, publishing

460 only that subset that they wish, and so can remain as data controller for their
461 own data, preserving authority over commercial and regulatory sensitivities.

462 To achieve this, the DDIM implements a data register. This records a
463 catalogue of data sources, versions of data published and other metadata
464 such as the publishing agent. A catalogue record includes an endpoint for
465 data sources published (a network address where SPARQL queries can be
466 submitted to query the source) and the source's schema or data model.
467 Together, these models allow a project member to examine source models,
468 formulate federated queries across these and submit the query through the
469 DDIM server. A federated query is one that can be submitted to one triple
470 store but used to remotely access other stores to retrieve subsets of data.

471 The DDIM server implements the DCAT ontology through GraphDB ².
472 Every project implements a single DCAT catalogue. The server presents a
473 standard interface that admits management of registry entries, shown in Fig-
474 ure 7. The common context is added by default. Thereafter, administrators
475 can add other data sources that project collaborators have published. This
476 includes key details about the source, including endpoint URI, another URI
477 that describes the data, the source's context (instance or class) and other
478 metadata that can be used to describe the source's provenance. Each entry
479 is represented as a DCAT resource/record. The interface provides this infor-
480 mation to others that may wish to formulate queries and includes filters and
481 other search features to help discover relevant information sources.

²<https://www.ontotext.com/products/graphdb/>

District Informaton Model Server					
Project Data Catalogue Title: DDIM Context Model Description:model file Service Endpoint: http://127.0.0.1/DDIM.sparql					
filter value Submit New Record					
Title	Description	Modified	Context	Endpoint	Conforms To
Tranmission Data	Transmission Info	Nov. 10, 2019	TD1:Tranmission	http://127.0.0.1/transmission.sparql	TransOWL.owl
Small Data	Small Info	Nov. 10, 2019	SA1:Small	http://127.0.0.1/smallarea.sparql	SmallOWL.owl
Geo Data	Geodata Info	Nov. 10, 2019	GD1:Geo	http://127.0.0.1/geodata.sparql	GeoOWL.owl
Archetypes	Archetypes Info	Nov. 10, 2019	Arch:Archetypes	http://127.0.0.1/archetypesdata.sparql	ArchetypesOWL.owl

Figure 7: DDIM Data Registry Interface

482 In this work, there are currently five data stores, including the core con-
483 text, the archetype simulation results, transmission node related data, small
484 area related data and geo-directory (buildings); a summary of these sources
485 is provided in Table 3 organised by prefix. Record and resource entries were
486 made in the DDIM registry for each using the interface described above.
487 The organisation of each source was documented and a model diagram was
488 produced. This was uploaded as part of the entry along with the sources end-
489 points. A query interface was implemented to allow participants to generate
490 their own queries and view the results either as a table or as a downloadable
491 CSV file. Other queries were produced to inform the reporting interface that
492 is described next.

Table 3: Summary of namespace prefixes used for each data collection

Prefix	Purpose
essipenergy	Representing context hierarchy (transmission - district - building) for the ESIPP project.
essiparchtype	Representing building archetypes information generated during analysis process
dcat	WC3 DCAT Ontology.
transmissiondata	Data about transmission nodes.
smallareadata	small area data.

493 4. Results

494 The DDIM and its associated data was applied to inform exploration
495 of various scenarios that will impact Irish electricity transmission and dis-
496 tribution networks. An information exploration interface was developed to
497 facilitate query and analysis of the DDIM’s data; this allowed analysts to
498 create various scenarios that isolated relevant data that were, in turn, ex-
499 amined through the tool’s interface, or exported to CSV format for further
500 analysis in other tools. This section describes the tool and discusses some of
501 the findings of the analysis carried out.

502 4.1. Exploring Results

503 A scenario exploration tool was developed to assist users to explore and
504 compare the effects of various scenarios. The web-based interface allows

505 analysts to both define new scenarios and compare their effects at a national
506 or individual transmission node. Results can be viewed in graph or table
507 form, or can be downloaded as CSV files for further analysis in external
508 analysis tools. A screenshot from the interface is shown in Figure 8.

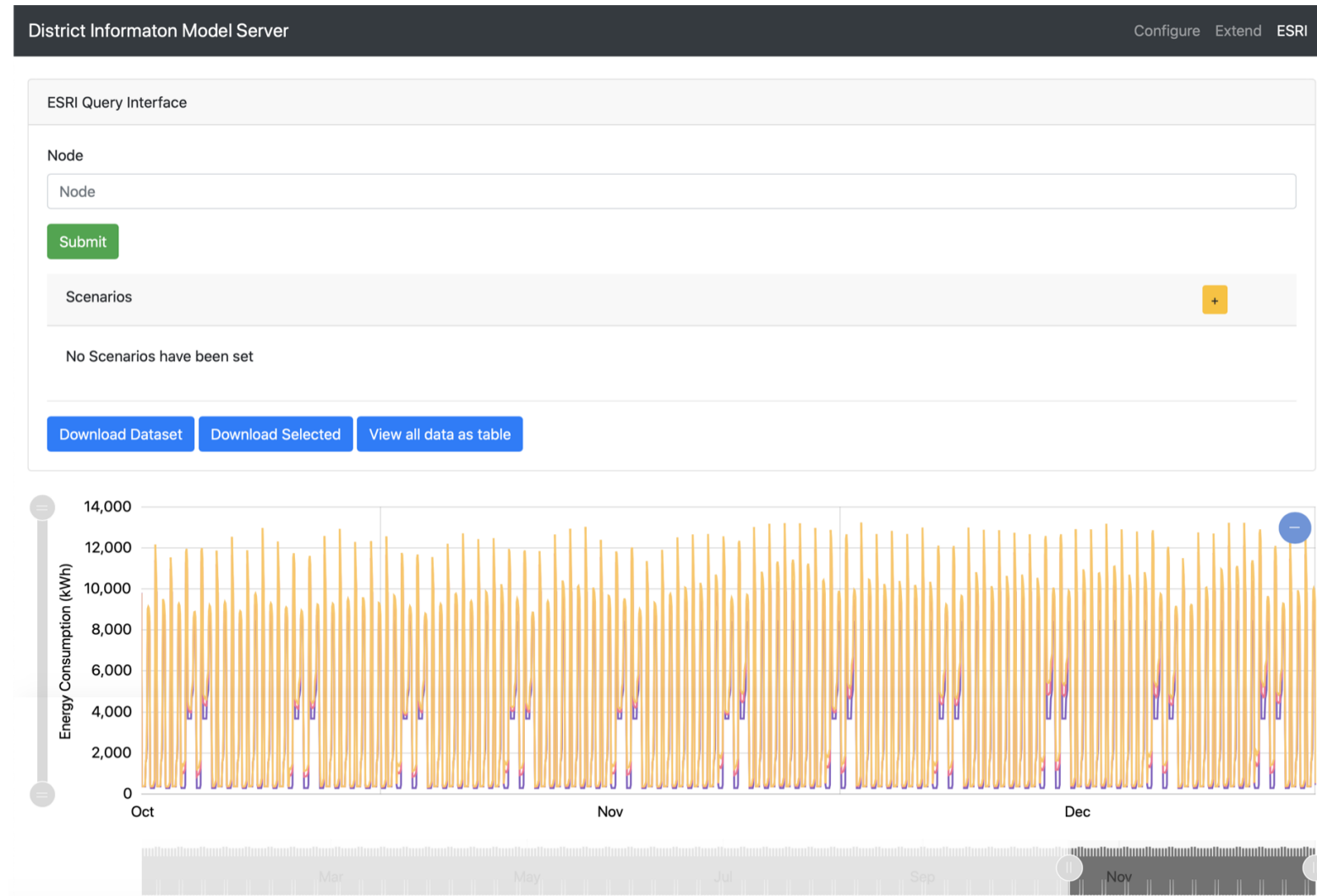


Figure 8: ESIPP Energy Consumption Reporting Interface

509 The interface assumes that a series of changes to housing stock archetypes
510 have been modelled, for example, the addition of electric vehicle charging
511 points or heat pumps. Users can define any number of scenarios. Each
512 scenario is given a name, before specifying what percentage of the housing
513 stock that has had some modelled change applied, for example, that 20% of
514 domestic housing stock are retrofitted with heat pumps. The user can select
515 several scenarios for comparison, and specify either all or specific transmission
516 nodes. These are submitted to a server that in turn submits several queries
517 to the DDIM to inform the scenarios.

518 The queries include one to retrieve a description of building archetype
519 numbers in the area served by a transmission node, the model information
520 for each archetype. A federated query to retrieve building archetypes is
521 shown in Listing 2. It first retrieves all buildings in all small areas served by
522 a transmission node, before retrieving an archetype for each of the buildings.
523 The system demonstrates its ability to use federated queries by retrieving
524 the archetype information from another SPARQL endpoint; this feature was
525 implemented to demonstrate this functionality rather than for any technical
526 reason. Archetype modelling information is retrieved for each archetype in a
527 separate query.

```

528 PREFIX esipparchetypes: <http://192.168.1.24/esipp/archetypes#>
529 PREFIX esippenergy: <http://192.168.1.32/esipp/MallowExample#>
530
531 select ?buildingarchetype (COUNT(?buildingarchetype) as
532     ?counttype) where {
533
534     #Get transmission node
535     ?t esippenergy:context "Transmission" .
536     ?t esippenergy:id 220 .
537
538     #Aggregate buildings under transmission node
539     ?t esippenergy:contains* ?childinstance .
540     ?childinstance esippenergy:type ?childtype
541         filter(?childtype = "SmallArea") .
542
543     #Get buildings at per small area
544     ?childinstance esippenergy:contains* ?buildinginstance .
545     ?buildinginstance esippenergy:type ?buildtype
546         filter(?buildtype = "Building") .
547     ?buildinginstance esippenergy:id ?buildingid .
548
549     #Access external data source for additional building level
550     data
551     SERVICE <repository:archetypes> {
552         ?archetypeinstance esipparchetypes:aid ?buildingid .
553         ?archetypeinstance esipparchetypes:oftype
554             ?buildingarchetype .
555     }
556 }
557 group by ?buildingarchetype

```

Listing 2: An example SPARQL federated query submitted to the project server to retrieve calculate buildings archetype count for a particular transmission node. Using a wildcard instead of a transmission node would retrieve a count of all archetypes in Ireland

These two result sets, along with the scenario descriptions are then used by the server to determine energy consumption. In the case of examining the effect of adding electric space heaters to 20% of housing stock, the modelled pre-retrofit consumption figures are applied to 80% of each housing stock

562 archetype, while the figures for post-retrofit are applied to 20% of each hous-
563 ing stock type. This calculation is applied to generate hourly figures over a
564 one year period. These results are processed and a JSON response to the
565 interface is displayed as a graph and a table. The interface can also provide
566 the results as a downloadable CSV.

567 4.2. Analysis

568 The decarbonisation of the heating and transport sectors are core parts of
569 Irish Government climate policy, and it is envisaged that this will be achieved
570 primarily through electrification of both [52]. The Irish Government has set
571 targets for six hundred thousand heat pumps to be installed by 2030, two
572 thirds of which are to be installed in existing dwellings. This will also re-
573 quire significant retrofitting to raise the energy efficiency of these dwellings.
574 Such large-scale electrification will have significant impacts on the electricity
575 generation and transmission network. In order to understand the long-run
576 power system impacts of this electrification, the ENGINE model was devel-
577 oped [53], which determines the least-cost development of the power system
578 under future policy scenarios.

579 The ENGINE model requires information on electricity demand at each
580 of the 676 nodes in the transmission system, and so the impact of the electri-
581 fication of heating at each node is required in order to determine the power
582 system impacts of future electrification. The Irish Transmission System Op-
583 erator EirGrid has developed projections of the electricity demand from heat
584 pumps at national level only [54]. In the absence of the DDIM, a “top-down”
585 approach of scaling this system data to transmission node level on the basis
586 of peak demand at each node was the only alternative method of projecting
587 electricity demand at each node. The DDIM, on the other hand, allowed the
588 determination of demand in a “bottom-up” manner, incorporating household
589 characteristics as well as the impact of retrofitting dwellings. Table 4 shows
590 the difference in the total number of dwellings by region that arises from
591 scaling EirGrid data versus using the data arrived at by the DDIM.

592 Because the DDIM server’s functionality facilitated integration and query-
593 ing of a greater number of data sources within the limitations of resources
594 available to the project, a more extensive data set could be used for analysis.
595 This admitted a more accurate analysis to determine the demand for heat at
596 each transmission node. Figure 9 compares the heat demand data from both
597 data sets. The DDIM input displays seasonal variation in the heat demand

Table 4: Comparison of total number of dwellings in each region in different data sets

Region ID	RegionName	Total number of dwellings	
		Scaled EirGrid Input	DDIM Input
1	Border	149196	204256
2	Midland	69748	92333
3	West	165517	260057
4	Dublin	341367	389366
5	Mideast	200099	221570
6	Midwest	125674	188018
7	Southeast	175100	237114
8	Southwest	233356	340417
9	Northern Ireland	9287	13956

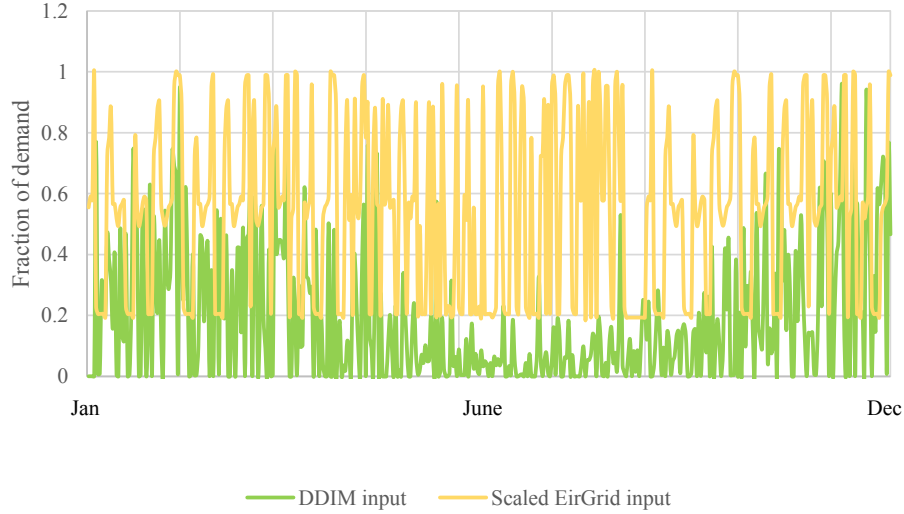


Figure 9: Heat demand from the DDIM and from scaling EirGrid’s national dataset

598 whereas the scaled EirGrid heat demand has no such variation, which may
599 lead to overestimation in the required infrastructure capacity.

600 The ENGINE model was used to test the impacts of electrifying 20%
601 (REF_20%) and 30% (REF_30%) of the building stock with heat pumps, as
602 well as increasing the energy efficiency of these dwellings to a B2 Building
603 Energy Rating. Some summary results are presented in Figure 10, which
604 shows the percentage change in various system metrics with respect to a

605 base case of no electrification.

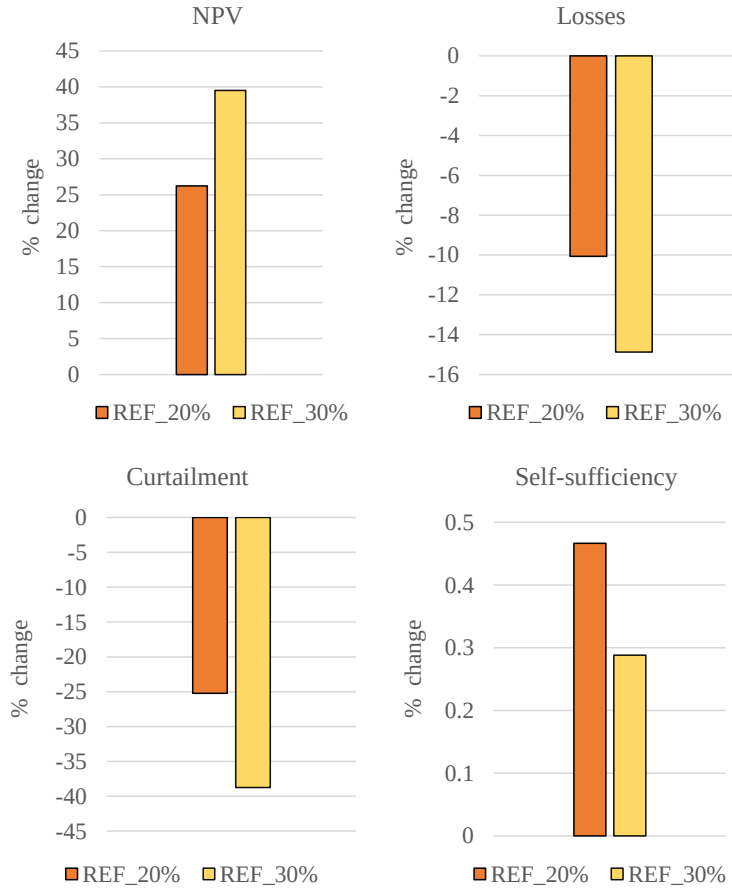


Figure 10: Percentage change in various system metrics as a result of electrification

606 The total costs measured as Net Present Value (NPV) of the system,
 607 comprising investment costs in generation and transmission assets as well as
 608 fuel and carbon costs of the operation of generation assets, increases with
 609 increasing levels of electrification due to the requirement to meet additional
 610 electricity demand. With a higher level of electrification, renewable curtail-
 611 ment decreases; thus, heat pumps increase the utilisation of renewable energy
 612 in the system. Overall system losses also decrease and the self-sufficiency of
 613 the system increases, indicating that heat pumps can lead to a more efficient

614 operation of the transmission system.

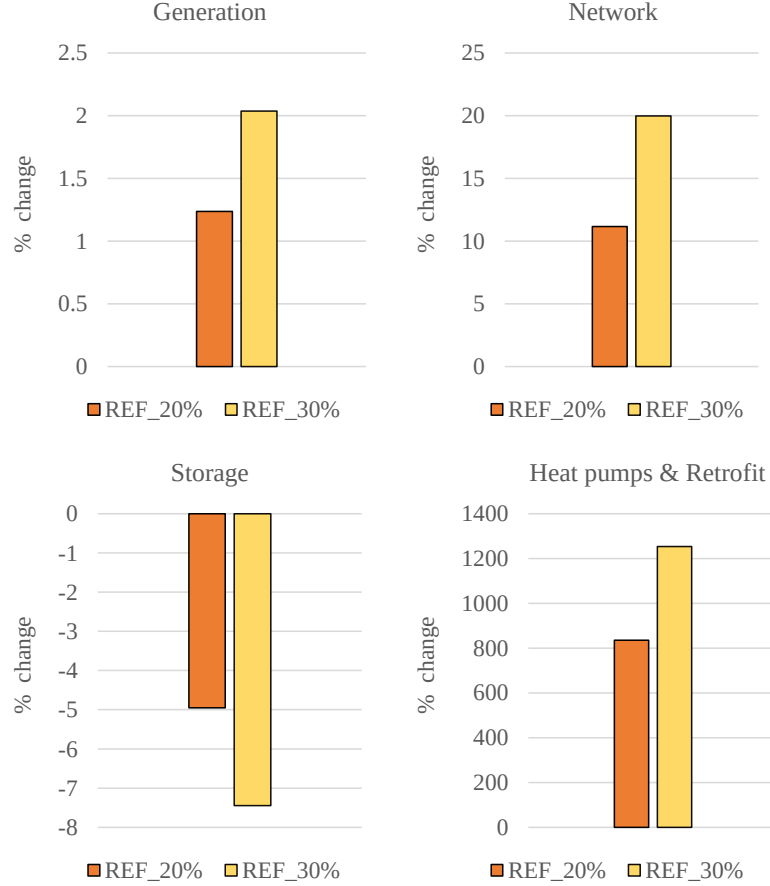


Figure 11: Percentage change in the components of total costs

615 The percentage change in various components of total costs with respect
 616 to the base case are presented in Figure 11. Higher levels of electrification
 617 lead to increased electricity demand from heat pumps, which in turn leads
 618 to higher generation and network costs. On the flip side, costs associated
 619 with storage decrease due to the potential to exploit flexibility from the
 620 operation of heat pumps. Heat pumps and retrofit costs obviously show a
 621 large increase with electrification as greater numbers of houses undergo deep
 622 retrofits to accommodate heat pumps.

623 Future work will consider the potential of using alternative fuels such as
624 hydrogen for the decarbonisation of heat, as well as the regional impacts of
625 various climate action policies.

626 5. Conclusions

627 This paper describes the use of linked data approaches to permit queries
628 across large, diverse information sources to allow reasoning about complex
629 questions at multiple scales. The methodology defines a central spine model
630 that provides a common hierarchical context; various heterogeneous external
631 information sources can be associated with these contexts, creating a homo-
632 geneous data network that can be queried using federated SPARQL queries.
633 These distributed sources are, in themselves, registered in a central catalogue
634 in order to facilitate data discovery; the actual data, however, remains under
635 the control of their source organisations. In this way a large, extensible, in-
636 terconnected network of distributed data describing a built environment can
637 be constructed; this network can be queried centrally to provide customised
638 views of subsets of the data.

639 The utility of the architecture was examined by establishing a data net-
640 work to support Ireland’s Economic and Social Research Institute (ESRI)
641 analysts to investigate the impact of various decarbonisation scenarios on
642 the Irish transmission and distribution networks. Large datasets were used
643 to support this investigation, including details of 220 transmission nodes,
644 c.16,500 small areas and approximately 2 million residential buildings. A
645 series of archetypes were developed, including as-is and energy consumption
646 per residence archetype for numerous scenarios, including the adoption of
647 heat pumps and electric vehicles. It was found that the DDIM approach
648 admitted finer grained data analysis, improving the accuracy of the analysis.
649 The analysis was supported development of a scenario exploration tool that
650 was also described.

651 6. Future Work

652 While the approach was successfully applied to a range of data, the pub-
653 lication of data as a SPARQL endpoint - as required by the DDIM’s archi-
654 tecture - can be simplified. While this project has the ability to cater for
655 CSV formatted data, other formats that might be encountered must be iden-
656 tified and catered for through the development of suitable parsers. These

657 must be integrated into a service that acts as a turn key wizard driven ap-
658 proach to ensure that the overhead encountered by project participants is
659 minimized. A second issue that was encountered, and will be addressed in
660 future iterations, was the impact of inefficient queries that are submitted as
661 part of a federated query; it was possible on occasion to create a backlog of
662 queries and so impact both the server servicing the remote data source and
663 the DDIM. Pre-processing of the query to estimate its complexity, along with
664 load balancing of queries at data endpoints will be investigated as potential
665 solutions.

666 7. Acknowledgments

667 This publication has emanated from research supported by US-Ireland
668 R&D Partnership Research Grant 2110171. The opinions, findings and con-
669 clusions or recommendations expressed in this material are those of the au-
670 thor(s) and do not necessarily reflect the views of the Science Foundation
671 Ireland or other funding agencies.

672 8. References

- 673 [1] M. Santamouris, Innovating to zero the building sector in europe: Min-
674 imising the energy consumption, eradication of the energy poverty and
675 mitigating the local climate change, *Solar Energy* 128 (2016) 61–94.
- 676 [2] I. Vorushylo, P. Keatley, N. Shah, R. Green, N. Hewitt, How heat pumps
677 and thermal energy storage can be used to manage wind power: a study
678 of ireland, *Energy* 157 (2018) 539–549.
- 679 [3] V. Bianco, F. Scarpa, L. A. Tagliafico, Estimation of primary energy
680 savings by using heat pumps for heating purposes in the residential
681 sector, *Applied Thermal Engineering* 114 (2017) 938–947.
- 682 [4] G. Hill, O. Heidrich, F. Creutzig, P. Blythe, The role of electric vehicles
683 in near-term mitigation pathways and achieving the uks carbon budget,
684 *Applied Energy* 251 (2019) 113111.
- 685 [5] J. Love, A. Z. Smith, S. Watson, E. Oikonomou, A. Summerfield,
686 C. Gleeson, P. Biddulph, L. F. Chiu, J. Wingfield, C. Martin, et al.,

- 687 The addition of heat pump electricity load profiles to gb electricity de-
 688 mand: Evidence from a heat pump field trial, *Applied Energy* 204 (2017)
 689 332–342.
- 690 [6] B. Bach, J. Werling, T. Ommen, M. Münster, J. M. Morales,
 691 B. Elmegaard, Integration of large-scale heat pumps in the district heat-
 692 ing systems of greater copenhagen, *Energy* 107 (2016) 321–334.
- 693 [7] S. E. Madnick, R. Y. Wang, Y. W. Lee, H. Zhu, Overview and frame-
 694 work for data and information quality research, *J. Data and Information*
 695 *Quality* 1 (1).
- 696 [8] J. Bonilla, J. Perez, I. Pinto-Seppa, A. Zarli, Artificial intelligence - re-
 697 search and innovation needs of manufacturing, energy intensive indus-
 698 tries, bio-based industries and construction, Tech. rep., The European
 699 Construction Technology Platform - and the Energy Efficient Building
 700 (07 2019).
- 701 [9] S. Boddy, Y. Rezgui, G. Cooper, M. Wetherill, Computer integrated
 702 construction: A review and proposals for future direction, *Adv. Eng.*
 703 *Softw.* 38 (10) (2007) 677687.
- 704 [10] L. G. Swan, V. I. Ugursal, Modeling of end-use energy consumption in
 705 the residential sector: A review of modeling techniques, *Renewable and*
 706 *sustainable energy reviews* 13 (8) (2009) 1819–1835.
- 707 [11] B. Talebi, P. A. Mirzaei, A. Bastani, F. Haghighat, A review of district
 708 heating systems: modeling and optimization, *Frontiers in Built Envi-*
 709 *ronment* 2 (2016) 22.
- 710 [12] R. Aghamolaei, M. H. Shamsi, M. Tahsildoost, J. ODonnell, Review
 711 of district-scale energy performance analysis: Outlook towards holistic
 712 urban frameworks, *Sustainable cities and society* 41 (2018) 252–264.
- 713 [13] E. Curry, J. ODonnell, E. Corry, S. Hasan, M. Keane, S. ORiain, Linking
 714 building data in the cloud: Integrating cross-domain building data using
 715 linked data, *Advanced Engineering Informatics* 27 (2) (2013) 206 – 219.
- 716 [14] C. F. Reinhart, C. C. Davila], Urban building energy modeling a review
 717 of a nascent field, *Building and Environment* 97 (2016) 196 – 202.

- 718 [15] N. Oza, J. Münch, J. Garbajosa, A. Yague, E. Gonzalez Ortega, Identifying potential risks and benefits of using cloud in distributed software
719 development, in: J. Heidrich, M. Oivo, A. Jedlitschka, M. T. Baldassarre (Eds.), Product-Focused Software Process Improvement, Springer
720 Berlin Heidelberg, Berlin, Heidelberg, 2013, pp. 229–239.
- 723 [16] Q. H. Cao, I. Khan, R. Farahbakhsh, G. Madhusudan, G. M. Lee, N. Crespi, A trust model for data sharing in smart cities, 2016 IEEE
724 International Conference on Communications (ICC) (2016) 1–7.
- 726 [17] Data privacy vocabularies and controls community group, <https://www.w3.org/community/dpvcg/>, accessed: 2020-04-01.
727
- 728 [18] Y. Li, R. Garca-Castro, N. Mihindukulasooriya, J. O'Donnell, S. Vega-Sanchez, Enhancing energy management at district and building levels
729 via an em-kpi ontology, Automation in Construction 99 (2019) 152 –
730 167.
- 732 [19] A.-C. N. Ngomo, S. Auer, J. Lehmann, A. Zaveri, Introduction to Linked
733 Data and Its Lifecycle on the Web, Springer International Publishing,
734 Cham, 2014, pp. 1–99.
- 735 [20] S. Consoli, V. Presutti, D. R. Recupero], A. G. Nuzzolese, S. Peroni,
736 M. Mongiovi', A. Gangemi, Producing linked data for smart cities: The
737 case of catania, Big Data Research 7 (2017) 1 – 15.
- 738 [21] Y. Rezgui, S. Boddy, M. Wetherill, G. Cooper, Past, present and future
739 of information and knowledge sharing in the construction industry: Towards semantic service-based e-construction?, Computer-Aided Design
740 43 (5) (2011) 502 – 515, emerging Industry Needs for Frameworks and
741 Technologies for Exchanging and Sharing Product Lifecycle Knowledge.
- 743 [22] J. Bonilla, J. Perez, I. Pinto-Seppa, A. Zarli, Artificial intelligence -
744 research and innovation needs, manufacturing, energy intensive industries, bio-based industries and construction, Tech. rep., The European
745 Construction Technology Platform (ECTP) - and the Energy-efficient Building (2019).
746
747 URL [http://www.ectp.org/fileadmin/user_upload/documents/ECTP/Miscellaneous_doc/ECTP-EeB_AI-in-Industrial_](http://www.ectp.org/fileadmin/user_upload/documents/ECTP/Miscellaneous_doc/ECTP-EeB_AI-in-Industrial_Technologies_FINAL.pdf)
748
749 [Technologies_FINAL.pdf](http://www.ectp.org/fileadmin/user_upload/documents/ECTP/Miscellaneous_doc/ECTP-EeB_AI-in-Industrial_Technologies_FINAL.pdf)
750

- 751 [23] S. Boddy, Y. Rezgui, G. Cooper, M. Wetherill, Computer in-
 752 tegrated construction: A review and proposals for future direc-
 753 tion, *Advances in Engineering Software* 38 (10) (2007) 677 – 687.
 754 doi:<https://doi.org/10.1016/j.advengsoft.2006.10.007>.
 755 URL [http://www.sciencedirect.com/science/article/pii/](http://www.sciencedirect.com/science/article/pii/S096599780600175X)
 756 [S096599780600175X](http://www.sciencedirect.com/science/article/pii/S096599780600175X)
- 757 [24] J. O. Yehong Li, R. Garca-Castro, S. Vega-Sanchez, Identifying
 758 stakeholders and key performance indicators for district and building
 759 energy performance analysis, *Energy and Buildings* 155 (2017) 1–15.
 760 doi:<https://doi.org/10.1016/j.enbuild.2017.09.003>.
 761 URL [http://www.sciencedirect.com/science/article/pii/](http://www.sciencedirect.com/science/article/pii/S0378778817304772)
 762 [S0378778817304772](http://www.sciencedirect.com/science/article/pii/S0378778817304772)
- 763 [25] J. Egan, D. Finn, P. H. D. Soares, V. A. R. Baumann, R. Aghamolaei,
 764 P. Beagon, O. Neu, F. Pallonetto, J. ODonnell, Definition of a useful
 765 minimal-set of accurately-specified input data for building energy per-
 766 formance simulation, *Energy and Buildings* 165 (2018) 172–183.
- 767 [26] D. Djenouri, R. Laidi, Y. Djenouri, I. Balasingham, Machine learning
 768 for smart building applications: Review and taxonomy, *ACM Comput.*
 769 *Surv.* 52 (2).
- 770 [27] D. Jato-Espino, E. Castillo-Lopez, J. Rodriguez-Hernandez, J. C.
 771 Canteras-Jordana, A review of application of multi-criteria decision
 772 making methods in construction, *Automation in Construction* 45 (2014)
 773 151 – 162.
- 774 [28] D. Arnott, G. Shijia, Behavioral economics for decision support systems
 775 researchers, *Decision Support Systems* 122 (7).
- 776 [29] D. E. Snediker, A. T. Murray, T. C. Matisziw, Decision support for
 777 network disruption mitigation, *Decision Support Systems* 44 (4) (2008)
 778 954 – 969.
- 779 [30] V. Gudivada, A. Apon, J. Ding, Data quality considerations for big data
 780 and machine learning: Going beyond data cleaning and transformations,
 781 *International Journal on Advances in Software* 10 (2017) 1–20.

- 782 [31] I. Stoica, D. Song, R. Popa, D. Patterson, M. Mahoney, R. Katz,
783 A. Joseph, M. Jordan, J. Hellerstein, J. Gonzalez, K. Goldberg, A. Gh-
784 odsi, D. Culler, P. Abbeel, A berkeley view of systems challenges for
785 ai.
- 786 [32] C. Bizer, T. Heath, T. Berners-Lee, Linked data - the story so far, Int.
787 J. Semantic Web Inf. Syst. 5 (2009) 1–22.
- 788 [33] Sparql 1.1 query language, [https://www.w3.org/TR/
789 sparql11-query/](https://www.w3.org/TR/sparql11-query/), accessed: 2020-04-01.
- 790 [34] Vocabularies, [https://www.w3.org/standards/semanticweb/
791 ontology](https://www.w3.org/standards/semanticweb/ontology), accessed: 2020-04-01.
- 792 [35] E. Curry, J. ODonnell, E. Corry, S. Hasan, M. Keane, S. ORiain,
793 Linking building data in the cloud: Integrating cross-domain building
794 data using linked data, Advanced Engineering Informatics 27 (2) (2013)
795 206 – 219. doi:<https://doi.org/10.1016/j.aei.2012.10.003>.
796 URL [http://www.sciencedirect.com/science/article/pii/
797 S1474034612000961](http://www.sciencedirect.com/science/article/pii/S1474034612000961)
- 798 [36] W. Terkaj, G. Schneider, P. Pauwels, Reusing domain ontologies in
799 linked building data: The case of building automation and control,
800 in: Proceedings of the 8th Workshop Formal Ontologies Meet Industry,
801 Joint Ontology Workshops 2017, CEUR Workshop Proceedings, 2017.
- 802 [37] K. Baumgrtel, M. Kadolsky, R. Scherer, An ontology framework for
803 improving building energy performance by utilizing energy saving regu-
804 lations, 2014. doi:10.1201/b17396-86.
- 805 [38] Data catalogue vocabulary (dcat), [https://www.w3.org/TR/
806 vocab-dcat/](https://www.w3.org/TR/vocab-dcat/), accessed: 2020-04-01.
- 807 [39] D. Mountain, The dimensions of context and its role in mobile informa-
808 tion retrieval, SIGSPATIAL Special 3 (2) (2011) 7177.
- 809 [40] R. J. Scherer, P. Katranuschkov, Context capturing of multi-information
810 resources for the data exchange in collaborative project environments,
811 in: Proceedings of the 2019 European Conference for Computing in
812 Construction, Vol. 1 of Computing in Construction, University College
813 Dublin, Chania, Crete, 2019, pp. 359–366. doi:10.35490/ec3.2019.

- 814 173.
815 URL <https://doi.org/10.35490/ec3.2019.173>
- 816 [41] R. Scherer, S.-E. Schapke, A distributed multi-model-based manage-
817 ment information system for simulation and decision-making on con-
818 struction projects, *Advanced Engineering Informatics* 25 (4) (2011) 582
819 – 599, special Section: Advances and Challenges in Computing in Civil
820 and Building Engineering.
- 821 [42] Linked building data community group, [https://www.w3.org/](https://www.w3.org/community/lbd/)
822 [community/lbd/](https://www.w3.org/community/lbd/), accessed: 2020-04-01.
- 823 [43] S. Simon, A. Degbelo, R. Lemmens, C. Elzakker, P. Zimmerhof, N. Kos-
824 tic, J. Jones, G. Banhatti, Exploratory querying of sparql endpoints
825 in space and time, *Semantic Web* 8 (2016) 65–86. doi:10.3233/
826 SW-150211.
- 827 [44] W. Terkaj, P. Pauwels, A method to generate a modular ifcowl ontology,
828 in: *Proceedings of the 8th Workshop Formal Ontologies Meet Industry,*
829 *Joint Ontology Workshops 2017, CEUR Workshop Proceedings.*, 2017.
- 830 [45] D. Allemang, S. Sakr, M. Wylot, R. Mutharaju, D. L. Phuoc,
831 C. Irini Fundulaki, Linked data: Storing, querying, and reasoning., *Journal of the Association for Information Science and Technology* 70 (8)
832 (2019) 905–907.
833
- 834 [46] A. A. Vaisman, K. Chentout, Mapping spatiotemporal data to rdf: A
835 sparql endpoint for brussels, *ISPRS Int. J. Geo-Information* 8 (2019)
836 353.
- 837 [47] C. Hoare, U. Ali, J. O'Donnell, Dynamic district information server:
838 On the use of w3c linked data standards to unify construction data,
839 in: *Proceedings of the 2019 European Conference for Computing in*
840 *Construction, Vol. 1 of Computing in Construction, University College*
841 *Dublin, Chania, Crete, 2019, pp. 265–273.*
- 842 [48] J. Allegrini, K. Orehounig, G. Mavromatidis, F. Ruesch, V. Dorer,
843 R. Evins, A review of modelling approaches and tools for the simula-
844 tion of district-scale energy systems, *Renewable and Sustainable Energy*
845 *Reviews* 52 (2015) 1391–1404.

- 846 [49] O. Neu, S. Oxizidis, D. Flynn, D. Finn, Utilising time of use surveys
847 to predict domestic hot water consumption and heat demand profiles of
848 residential building stocks, *British Journal of Environment and Climate*
849 *Change* 6 (2) (2016) 77–89.
- 850 [50] DesignBuilder, DesignBuilder Software Ltd - Home (2017).
851 URL <https://www.designbuilder.co.uk/>
- 852 [51] Building Regulations, Technical Guidance Document L- Conservation of
853 Fuel and Energy Dwellings, Tech. rep., Government of Ireland (2019).
- 854 [52] G. of Ireland, Climate action plan 2019 to tackle climate break-
855 down, available online: [https://www.dccae.gov.ie/en-ie/](https://www.dccae.gov.ie/en-ie/climate-action/publications/Documents/16/Climate_Action_Plan_2019.pdf)
856 [climate-action/publications/Documents/16/Climate_Action_](https://www.dccae.gov.ie/en-ie/climate-action/publications/Documents/16/Climate_Action_Plan_2019.pdf)
857 [Plan_2019.pdf](https://www.dccae.gov.ie/en-ie/climate-action/publications/Documents/16/Climate_Action_Plan_2019.pdf) (2019).
- 858 [53] D. Z. Fitiwi, M. Lynch, V. Bertsch, Enhanced network effects and
859 stochastic modelling in generation expansion planning: Insights from an
860 insular power system, *Socio-Economic Planning Sciences* (2020) 100859.
- 861 [54] EirGrid, Tomorrow's energy scenarios 2019, available online:
862 [http://www.eirgridgroup.com/site-files/library/EirGrid/](http://www.eirgridgroup.com/site-files/library/EirGrid/EirGrid-TES-2019-Report.pdf)
863 [EirGrid-TES-2019-Report.pdf](http://www.eirgridgroup.com/site-files/library/EirGrid/EirGrid-TES-2019-Report.pdf) (2019).